# Examining the role of agroecology in food security prediction: evidence from Burkina Faso using machine learning methods

Agroecology and food security prediction using machine learning

Theodore Nikiema[1], Pamela Giselle Katic[2*], Sylvain Kpenavoun Chogou[3], Eugene C. Ezin[4],

[1] Institute of Mathematics and Physics Sciences, University of Abomey Calavi, Cotonou, Benin.
Email: theodore.nikiema@imsp-uac.org ORCID: https://orcid.org/0000-0002-3602-7606
[2] Natural Resources Institute, University of Greenwich, Chatham, UK.
Email: P.G.Katic@gre.ac.uk ORCID: https://orcid.org/0000-0001-7594-1081
[3] Faculty of Agronomic Sciences, University of Abomey Calavi, Cotonou, Benin.
Email: sylvain.kpenavoun@uac.bj ORCID: https://orcid.org/0000-0003-0717-3599
[4] Institute of Training and Research in Computer Science, University of Abomey Calavi, Cotonou, Benin.
Email: eugene.ezin@uac.bj ORCID: https://orcid.org/0000-0002-7850-8600
[*] **Corresponding author**: Pamela Giselle Katic, email: P.G.Katic@gre.ac.uk

**Abstract**

**Background**: Food insecurity affects a large proportion of the world's population, with enormous disparities to the disadvantage of low-income countries, especially those in sub-Saharan Africa. In this context, public decision-makers, funders, and NGOs urgently need streamlined indicators on households in precarious situations to better orient their policies and aid. However, existing food security monitoring tools focus on isolated dimensions of the problem, and some remain static. We aim to fill these gaps by focusing on a composite metric of food security to explore associations and identify variables with predictive relevance for food security outcomes using machine learning techniques.

**Methodology**: This paper applies machine learning methods to a panel dataset from Burkina Faso to explore which agroecological practices are associated with improvements in food security and have an impact on the prediction of food security outcomes.

**Results**: Overall, the trained models recorded accuracies varying from 64% for the Multi-Layer Perceptron model to 72% for the assembly models, with learning times generally <1 minute, except for the SVM model, which recorded around 30 minutes, suggesting good potential for these models to identify food-insecure households. Our results show that agroecological intensification through a combination of multiple agroecological practices, such as livestock and crop species diversification, plays a key role in predicting households' food security status. Indeed, households practicing agroecology with good species diversification recorded a food insecurity rate of 6.6%, compared with 15% for conventional households, and pairwise differences in the prevalence of food security are significant (p-values most often <0.01). The partial dependence plots show that the probability of belonging to the food-secure class increases by 3% as species richness increases or as livestock are integrated.

**Conclusion**: The performance of the algorithms using a simplified measure of food security suggests that machine learning model-driven approaches could significantly improve food insecurity crisis response.

**Keywords**: Food security; Agroecology; farm household; Prediction; Machine learning; Burkina Faso; Sub-Saharan Africa

## 1. Introduction

In 2023, the prevalence of moderate or severe food insecurity in Sub-Saharan Africa (SSA) (63.3%) was more than double the global average (28.9%) (FAO et al., 2024). Food insecurity in SSA has its roots in various sources. Indeed, SSA remains one of the regions most affected by armed conflict and displacement, precarious household living standards, widespread poverty, political instability, and climate change. Agricultural intensification is often cited as an effective strategy to reduce food insecurity. However, its sustainability and effect on biodiversity are being questioned (Ickowitz et al., 2019; Kopittke et al., 2019; Tscharntke et al., 2012; Waha et al., 2018). In this context, agroecology is increasingly recognized as a production system that can sustainably feed SSA's population (Bezner Kerr et al., 2021; HLPE, 2019; Wezel et al., 2020).

Agroecology is defined as "a set of agricultural practices aiming to produce significant amounts of food while valuing ecological processes and ecosystem services" (Wezel et al., 2020). It is theorized to improve food security by diversifying production, enhancing ecosystem services, and strengthening farmers' capacity to adapt to climate change (Altieri et al., 2015). In contrast to conventional farming, agroecological practices emphasize sustainability over short-term farm objectives, addressing not only production but also the socio-economic and environmental dimensions of food security. This study is grounded in the agroecological framework proposed by Dalgaard et al. (2003), which conceptualizes agroecology as the study of the interactions between plants, animals, humans, and the environment within agricultural systems. This theoretical basis aligns with (Wezel et al., 2020) agroecological framework, where food security is presented as an ultimate goal of the defined practices while preserving the environmental and ecosystem services.

Agroecological practices focus on optimizing ecological processes, enhancing environmental and public health, and reducing social-ecological costs associated with agriculture, such as soil degradation, water contamination, greenhouse gas emissions, depletion of non-renewable resources, and social inequities (HLPE, 2019). Rather than adhering to standardized methods, agroecological farming systems are guided by principles that adapt to specific agroecosystems and socio-cultural contexts. Critics have pointed out that agroecology can make farming even more labor-intensive, and that its reliance on traditional methods may reinforce subsistence production and low yields, trapping farmers in a cycle of poverty with limited opportunities for economic advancement or meaningful contribution to national food security (Bernard & Lux, 2017; Mugwanya, 2019; Muhumuza, 2023). While these critics have questioned the impact of agroecology on food security and poverty, a review of recent evidence showed that the vast majority of studies found positive outcomes linked to the use of agroecological practices on food security in households in low- and middle-income countries (Bezner Kerr et al., 2021). However, despite the growing importance of agroecology (Liu et al., 2019; Nikiema et al., 2023), there are still gaps in understanding its links to farm household food security. Recent works designed to show the evidence of agroecology on food security have relied on small data sets, reducing their ability to use predictive modelling techniques capable of identifying food-insecure households in real-time, and limiting the scope of their conclusions (Lucantoni, 2020; Palomo-Campesino et al., 2021; Pronti & Coccia, 2021). More empirical evidence is needed to demonstrate agroecology's benefits, especially its predictive relevance and its ability to ensure food security.

In Burkina Faso (BF), the adoption of agroecological practices varies across the country's different regions (Nikiema, Ezin, & Chogou, 2024), with growing interest at the national level (MARAH, 2023). Evidence of its impact on food security is key, as the number of people facing acute food insecurity in BF during the annual lean seasons rose from nearly 700,000 in 2019 to 2.7 million in 2024 (WFP, 2024b). Understanding the role of farming practices on household food insecurity status can help to target assistance more effectively and create more effective long-term strategies to tackle food insecurity. To this end, national and humanitarian institutes need indicators available in a timely manner.

Recently, machine learning methods have introduced a transformative approach to the provision of food security estimates (Deléglise et al., 2023), while using easily accessible public data. These methods are gaining increasing recognition and are starting to be integrated into the methodological approaches of major multilateral organizations such as the HungerMapLIVE of the World Food Program (WFP) (Martini et al., 2022) and the World Food Security Outlook (WFSO) database of the World Bank (Andree et al., 2020). However, previous studies using machine learning methods face several limitations. First, they focus on individual dimensions of food insecurity, such as the Food Consumption Score (FCS), the Integrated Food Security Phase Classification (IPC), and calorie-based food security indicators. Second, they do not explore the role of households' agricultural activities, thus ignoring the effect of sustainable farming practices, such as agroecology, on food security (Hossain et al., 2019; Lentz et al., 2019; Martini et al., 2022). In this study, we focus on the composite food security indicator resulting from the combination of three dimensions of food security: Food Consumption Score (FCS), Livelihood Coping Strategies (LCS), and household Food Expenditure Shares (FES), following the Consolidated Approach for Reporting Indicators of Food Security (CARI) (WFP, 2021). This indicator was developed to standardize the food security assessment process across the world, and it allows us to present a multi-classification problem given by four food security statuses.

The main objective of this study is to develop and demonstrate a method to evaluate which agroecological practices are associated with household food security and have predictive relevance for food security outcomes. As an illustration, we use data from Burkina Faso's seven-month Permanent Agricultural Survey (PAS) covering a large-scale dataset of 5,308 farms over the period June to December 2020. We calculate and compare the suitability of alternative supervised learning algorithms to predict food security status. In addition to agroecological practices, we also assess the role of socio-economic, demographic, and farming characteristics as potential food insecurity risk factors.

This study offers both methodological and empirical contributions. Methodologically, it integrates agroecological variables into machine learning-based food security prediction models, showcasing the relevance of advanced predictive algorithms for agricultural sustainability research. It further demonstrates how predictive analytics can support more effective policies and decision-making. Empirically, it applies the methods to a large panel dataset from Burkina Faso, generating new evidence on the relationship between agroecological practices and food security outcomes in rural areas of SSA. We demonstrate that the combination of certain agroecological practices is associated with an improvement of household food security, thereby advancing the limited quantitative literature on the impacts of agroecology in low- and middle-income countries.

The rest of the paper is organized as follows. Section 2 provides key information on the case study area, the data, and the methods used. Section 3 presents the findings of our study, highlights the performance of alternative machine learning methods in inferring predictors of food security, and examines the implications of agroecological practices when predicting food insecurity. Section 4 discusses our findings and their policy implications, section 5 presents the limitations, and section 6 concludes and provides recommendations for future research efforts.

## 2. Materials and methods

### Case study area: Burkina Faso

Burkina Faso is a landlocked West African country with a surface area of around 274,000 $km^2$. It borders six countries: Mali to the northwest, Niger to the northeast, Benin and Togo to the southeast, Ghana to the south, and Ivory Coast to the southwest. According to the latest general population and housing census, it has a population of around 21 million, with 40% living below the national poverty line (INSD, 2022). Agriculture is the country's main economic activity (employing 80% of the population), with cereals (sorghum, millet, maize, and rice) and cotton dominating agricultural production. Arboriculture and market gardening also play an important role. Nearly all arable land is rain-fed, increasing the sector’s vulnerability to climate change. Smallholder farmers often lack access to markets, quality agricultural inputs, and advanced farming techniques.

The country is part of the Sudanian zone, benefitting from a dry tropical climate with two seasons: a dry season from November to June (with temperatures reaching 42°C in the shade from March to April in the Sahel region) and a rainy season from July to October. However, between November and February, a transitional season occurs between the main seasons, with temperatures dropping to between 20° and 25°. Four climatic zones can be identified: the Southern Sudanian zone with more than 900 mm of rainfall per year, the North Sudanian zone with rainfall between 700 and 900 mm per year; the Sub-Sahelian zone, with rainfall between 400 and 700 mm per year; and the Sahelian zone, with less than 400 mm of rainfall per year (Fig. 1). As a result, rainfall is generally unevenly distributed across regions, which has a major impact on food availability and, consequently, on the population’s nutritional status.

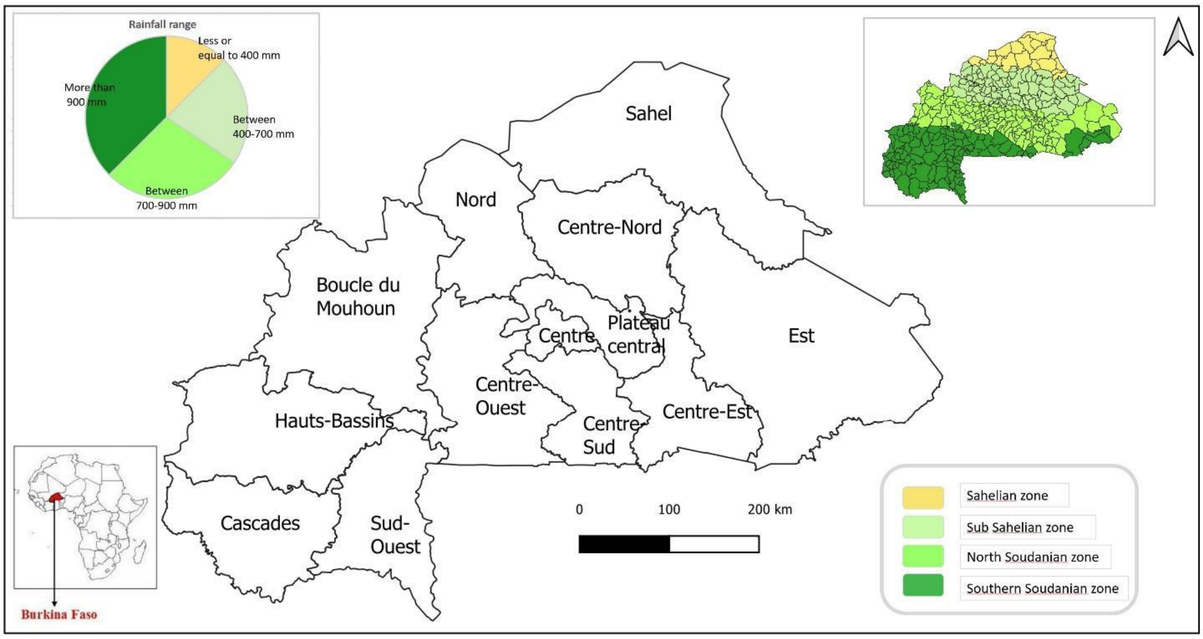


Fig. 1: Map of Burkina Faso, showing the country's delimitation into four main climatic zones in the upper quadrants and its location within Africa in the lower left panel. In some national reports, the North Sudanian and South Sudanian zones are presented as a single Sudanian zone (Kabore et al., 2017; Kabré et al., 2020).

**Data and measures**

To explore the role of agroecology as a predictor of food insecurity, we use data from Burkina Faso's seven-month Permanent Agricultural Survey (PAS), covering 5,308 farms over the period June to December 2020. The survey is a two-stage stratified sample survey with induced stratification at the second stage. The primary units (PUs) of the PAS survey consist of all the administrative villages and sectors of semi-rural localities in the national territory. The secondary units (SUs) represent all the agricultural households in the PUs. Therefore, the unit of observation is the household, in which one or more members farm plots or raise animals on behalf of the household. The geographical area covered is the national territory through its 45 provinces, with the province (3rd administrative level of the country) as the survey's representative level. Households are stratified according to their agricultural potential, their production capacity, and the villages to which they belong. Thus, a representative number of households with high agricultural production potential is included, which can reach 100 times the potential of smallholders, reflecting the structural heterogeneity of farming systems in the study area.

To calculate food security, we use the composite food security indicator resulting from the combination of three dimensions of food security: Food Consumption Score (FCS), Livelihood Coping Strategies (LCS), and household Food Expenditure Shares (FES). FCS captured household dietary diversity through a dietary recall of the last 7 days before the survey. For this purpose, 8 food groups were defined: cereals and tubers, legumes, vegetables, fruit, meat and fish, milk, sugar, and oils, in line with WFP recommendations (WFP, 2024a). Each food group was assigned a weight according to its energy, protein, and micronutrient density (WFP, 2024a). With a weight of 4, the meat and fish food groups had the highest weighting, while the sugars and oils groups, with weights of 0.5, had the lowest. The FCS was obtained by multiplying the weight of each food group by the number of days it was consumed during the week, as described by the formula (1), with $k$ running through the list of food groups, $z_k$ designates the $k$-th food groups, and $p_k$ their weights:

$$FCS = \sum_{k=1}^{8} z_k * p_k \qquad (1)$$

FES captured the share of household expenditure devoted to food. It was obtained by dividing total food expenditure by total household expenditure, while including the value of food consumed but not purchased. LCS was used to capture the dietary behaviour adopted by households to deal with various economic or food shocks and long-term changes in food security. It was derived from a series of questions related to households' experiences with livelihood strategies due to food scarcity during the 30 days before the survey (WFP, 2021).

The three indicators (FCS, LCS, and FES) were aggregated by considering their equivalent values within a 4-point standard classification scale. The integer part of the mean resulting from this aggregation represents the level of household food security (WFP, 2021). This resulted in the composite food security indicator with four classes: 1 for 'food security', 2 for 'limited food security', 3 for 'moderate food insecurity', and 4 for 'severe food insecurity'. Food-secure households are those that can meet their food needs without engaging in reduced coping and livelihood strategies. Limited food-secure households have minimally inadequate food consumption, rely on reduced coping, and apply stress coping strategies to secure food needs. Moderately food-insecure households have food consumption gaps and are unable to meet required food needs without applying crisis coping strategies. Severely food-insecure households have extreme food consumption gaps or extreme loss of livelihood assets that may lead to food consumption gaps or worse.

The different classes obtained were used to label data and train the various supervised learning models. In developing our machine learning models, we excluded the Food Consumption Score (FCS) from the predictor set despite its influence on predictive ability. In preliminary predictions, we found that when this variable was combined with the agroecological variables without FES and LCS, around 80% accuracy was achieved. Instead, we retained only Food Expenditure Share (FES) and Livelihood Coping Strategy (LCS) among the composite indicator components, as together they contribute less than the FCS. This decision was made to avoid circularity and ensure that model predictions are not directly driven by the primary food consumption measure, but rather reflect broader household economic and adaptive characteristics relevant to food security outcomes.

**Algorithms and materials**

Five classification algorithms were selected for training: Adaptive Boosting (AdaBoost), eXtreme Gradient Boosting (XGBoost), Support Vector Machine Linear (SVM_L), Support Vector Machine Radial (SVM_R), and Multilayer Perceptron (MLP). They were respectively implemented using the R software (version 4.5) packages "adabag", "xgboost", "e1071", and the packages tensorflow and keras under Python software (version 3.11) for the MLP model. The relationship between food status and agroecological practices was analysed using a chi-square test, and the differences between the proportions of food security categories were tested using a two-sided test with Benjamini-Hochberg specification, considering a p-value threshold of 0.05 for statistical significance. The database was divided into three subsets:70% for the training, 15% for validation of the hyperparameters, and 15% to test the models' ability to reproduce the automated task on a new dataset. Finally, the data cleaning was conducted using SPSS version 25.

In supervised learning, which is used in this study, the model is first trained on a dataset containing input features (household characteristics, farm practices) and known output labels (food security status in our case). The general process included the model training step, the model evaluation, and the prediction step. The model training step is where the algorithm identifies patterns in the data by minimizing a loss function, which measures the difference between the predicted outputs and the actual known outputs. The algorithms used in this study do this in different ways. XGBoost builds a sequence of decision trees, each correcting the errors of the previous one. AdaBoost combines several weak learners by focusing more on data points misclassified by previous learners. SVM finds the optimal boundary (hyperplane) that best separates the classes in the data space. MLP processes data through interconnected layers of nodes, learning complex nonlinear relationships. The final step consists of testing the models on new data to assess how accurately they can predict outcomes they have not seen before (classifying households into food security categories based on their features). Overall, the algorithms used in this study enable predictions that support decision-making and targeted interventions. The mathematical models and the equations are described in the appendix (A.1 to A.4).

Regarding model settings, for the radial SVM model, we specified a list of values for the "*cost*" and "*gamma*" regularization parameters: list("*cost"* = c(0.001, 10, 50, 100), "*gamma"* = c(1e-04, 0.001, 0.01, 1)) and a value of "cross"=4 for cross-validation. For the Adaboost model, we used several 200 iterations for "*mfinal*" and a bootstrapping approach by specifying the parameter "*boos"='True'*. For the XGBoost model, we used a "*max_depth*" = 10 for the tree depth, an "*eta*" = 0.01 for the learning rate, and a regularization parameter "*gamma*"=1. For the MLP model, we used a decreasing number of neurons from the input layer, the hidden layer to the output layer (210, 70, 4), and a *relu* activation function for the first two layers. Since we were dealing with a multiclassification problem, we used the *softmax* activator for the output layer, a loss parameter *loss='categorical_crossentropy*', and an *optimizer='adam'* for loss-dependent adjustment of model weights. Finally, we used a *batch_size* of 10 and an epoch of 15 for cross-validation.

## 3. Results

### Descriptive summary statistics

Tables 1 and 2 provide descriptive statistics of our sample (further details can be found in Table B.1 and Table B.2 of the Appendix). Burkina Faso's farming households are characterized by a large average number of 10 members, with only 8% of farms being managed by women. Farmers have low membership in farmers' organizations (with only 26% belonging to a farmers' organization) and have limited access to extension activities (around 20%). Most farmers combine agriculture with livestock, with 70% of farms practicing extensive sedentary livestock rearing. Taking into account the farm's various plots, which are often distant from each other, we note a good diversification of crops (more than half of farmers grow more than 5 different crops). Various agro-sylvo-pastoral activities other than rainfed cultivation, such as fishing, gathering, and harvesting wood products (firewood and service wood) and charcoal, are simultaneously practiced. However, within the same farm plot, we observe a low level of crop association (25%). Average yields for the main cereal crops (mainly maize, millet, or rice) are 1,372 kg/ha, with an average production of 2.3 tonnes. Agricultural equipment remains essentially basic, with less than 10% of farms having motorized equipment and a water source linked to their agricultural activities, and 17% of households having access to agricultural credit. Regarding food consumption and expenditure, households consume an average of 68% of their harvests, with around 40% of households allocating more than 50% of their income to food expenditure.  Dietary diversity remains limited (with only 54% of households having good dietary diversity), and the use of livelihood coping strategies remains limited (less than 30%). Some variables of the study display a standard deviation larger than their mean, reflecting the inherent heterogeneity of our sample due to the presence of large-scale producers who account for a disproportionately high share of total production.

However, we chose not to remove these large producers because they are valid and important observations that reflect the real diversity of production systems in the study area.

**Table 1.** Socio-economic characteristics of the farms.

| Variables | Mean | Median | Std. Deviation | Minimum | Maximum |
|---|---|---|---|---|---|
| Number of household members | 10 | 8 | 7 | 1 | 78 |
| Age of the index woman | 40 | 38 | 14 | 13 | 99 |
| Total agricultural area (Ha) | 4,6 | 3,1 | 5,0 | 0,1 | 53,3 |
| Number of agroforestry trees held | 44,6 | 25,0 | 60,4 | 0 | 879 |
| Number of agro-sylvo-pastoral species held | 7,2 | 6,0 | 5,1 | 0 | 34 |
| Number of edible crop species | 5,7 | 6,0 | 5,0 | 0 | 18 |
| Number of livestock species held | 3,9 | 4,0 | 1,8 | 0 | 10 |
| Number of livestock that are self-consumed | 5 | 2 | 9 | 0 | 199 |
| Proportion of crops self-consumed % | 68% | 70% | 22% | 0% | 100% |
| Proportion of agro-sylvo-pastoral products self-consumed % | 64% | 72% | 30% | 0% | 100% |
| Value of livestock sales (FCFA) | 139129 | 35250 | 317677 | 0 | 5773000 |
| Value of crop sales (FCFA) | 184076 | 15000 | 1370125 | 0 | 43984160 |
| Value of agro-sylvo-pastoral products sales (FCFA) | 107343 | 1406 | 2504495 | 0 | 94618203 |
| Labor (man days) | 5 | 0 | 18 | 0 | 376 |
| Labor costs (FCFA) | 5523 | 0 | 20295 | 0 | 357500 |
| Total production of the farm's main cereal crop (tons) | 2,3 | 1,5 | 3,1 | 0,0 | 74,6 |
| Average yield of the farm main cereal (Kg/ha) | 1372,2 | 1240,0 | 629,6 | 76,5 | 5925,9 |
| Value of total inputs purchased with own funds | 110307 | 56250 | 201081 | 0 | 9213000 |
| Value of total inputs purchased with credit | 40152 | 0 | 167533 | 0 | 6143000 |
| Total number of livestock | 50 | 31 | 106 | 0 | 4067 |

**Table 2.** Socio-economic characteristics of farms (categorical variables).

| Variables | Proportion % | * (Mode) |
|---|---|---|
| **Gender of the head of the household** | | |
| Man | 92,1 | * |
| Woman | 7,9 | |
| **Responsibility in Farmer organisations** | | |
| Executive member | 7,8 | |

| | | |
|---|---|---|
| Honorary member | 0,9 | |
| Ordinary member | 17,8 | |
| Not a member | 73,4 | * |
| **Access to agricultural extension activities** | | |
| No | 80,6 | * |
| Yes | 19,4 | |
| **Off-season home garden cultivation** | | |
| No | 86,6 | * |
| Yes | 13,4 | |
| **Food expenditure share** | | |
| Less than 50% | 63,4 | * |
| Between 50% et 65% | 18,8 | |
| Between 65% et 75% | 8,3 | |
| Great than 75% et plus | 9,5 | |
| **Livelihood coping strategies** | | |
| No | 71 | * |
| Stress | 27,5 | |
| Crisis | 0,9 | |
| Urgency | 0,6 | |
| **Dietary diversity** | | |
| Poor | 16,7 | |
| Limit | 29,5 | |
| Acceptable | 53,8 | * |

**Relationship between agroecology and food security**

To analyse the relationship between agroecology and food security, we first defined the intensity of agroecological practices of the different farms. Preliminary work using a clustering-based multidimensional analysis approach (Nikiema, Ezin, & Chogou, 2024) enabled us to identify three main farm groups. Group 1 comprised farms with traditional practices close to agroecology. These farms combine or rotate crops, use organic inputs, employ traditional equipment, integrate trees, and use soil and water conservation techniques (CES-DRS). Group 2 comprised farms using more chemicals (urea, pesticides, NPK), employing motorized machinery, having large numbers of trees, and a weak crop combination. The last group comprised farms with mixed practices, employing both organic and chemical inputs, often abandoning degraded soils and using local and traditional seedings. Following FAO's TAPE approach (FAO, 2019), we subdivided these groups into eight subgroups according to the level of integration of trees and the level of diversification of animal and crop species on the farms, as follows:

- Highly diversified agroecological farms with more than 8 crop species, more than 4 livestock species, and at least 20 trees integrated into their farms. From the point of view of CES-DRS techniques, these farms mainly use stone cordons, half-moons, filtering dikes, and Zai.
- Well-diversified agroecological farms with at least 6 crop species, more than 4 livestock species, and more than 20 trees integrated into their farm. In contrast to the first group, these farms use other CES-DRS techniques such as soil beads, grass strips, and hedges to conserve their soils.

- Moderately diversified agroecological farms with 2 to 4 crop species, less than 4 livestock species, and a range of 6 to 50 trees integrated into their farms.
- Sustainable farms with poor livestock and tree integration. These farms have practices common to agroecological farms, such as using CES-DRS, organic fertilizers, local seeds, and crop rotation. However, they do not integrate livestock and trees into their farms, with no more than 1 animal species (less than 5 heads) and fewer than 5 trees at best.
- Conventional farms with good integration of diversified animal species. These farms share common conventional farming practices, such as using chemical fertilizers, not combining crops, and not using techniques to conserve water and soil. However, they have no motorized equipment and integrate livestock farming.
- Conventional farms with good crop diversification and good use of soil and water conservation techniques. Like the previous group, these farms have practices in common with conventional farms. However, they do not own motorized equipment or animals but have at least 2 crop species and use CES-DRS techniques.
- Conventional farms that do not integrate livestock and do not diversify crops.
- Mixed farms without a clear, fixed production model.

Figure 2 summarizes the distribution of farms within these sub-groups.

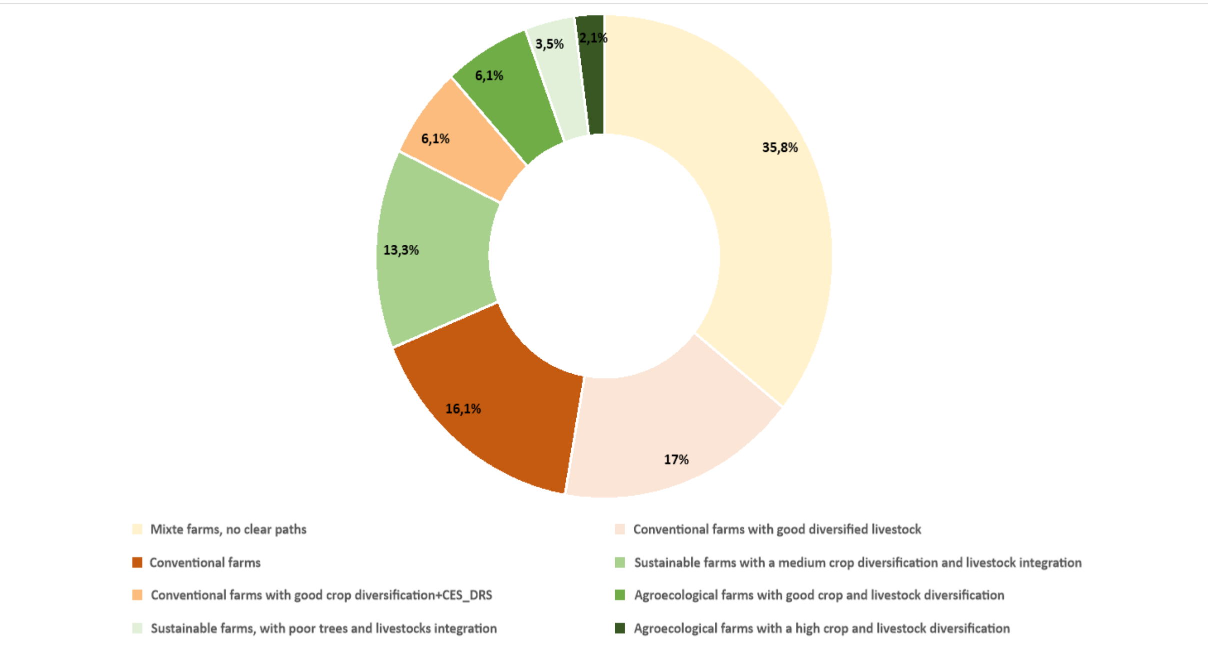


*Fig. 2: Distribution of farms according to their intensity of agroecological practices.*

Once we had established the agroecological intensity of the farms, we carried out a chi-square test of agroecological intensity with food security status. The test revealed a statistically significant association between agroecological intensity and food security, with a p-value of 0.000. We found a food insecurity prevalence of less than 7% in highly diversified agroecological farms, compared with 17% for non-diversified sustainable farms with no livestock and trees integration, and 15% for non-diversified conventional farms with no livestock and trees integration. Conventional farms that integrate livestock and diversified crops are also less affected by food insecurity (10%) (Table 3). A comparison between the various farm groups using the Benjamini-Hochberg test confirmed that agroecological farms with good crop and animal species diversification are more food-secure than all other farm groups, with a significance of less than 0.006 (Table 4). Taken individually, most farming practices were found to be significantly associated with households' food security status (Table B.3 of the appendix). Specifically, farms with at least 4 animal species are much less affected by food insecurity than those without animals, with 25% compared to 10% respectively. In terms of crop diversification, farms that grow at least 5 different crops have a food insecurity rate of 10%, compared with 34% for those that do not diversify their crops (Tables B.4 to B.6 of the appendix).

**Table 3.** Distribution of food insecurity according to the intensity of agroecological practices.

| | **Proportion of food-insecure households** |
|---|---|
| **Agricultural classes** | |
| Conventional farms | 15,1% |
| Conventional farms with good crop diversification & CES_DRS | 10,2% |
| Conventional farms with diversified livestock | 10,7% |
| Mixed farms, no clear paths | 12,8% |
| Sustainable farms, with no or poor integrated (trees and livestock) | 17,2% |
| Sustainable farms with medium crop diversification and livestock integration | 12,5% |
| Agroecology farms with good crop diversification and livestock integration | 7,0% |
| Agroecology farms with high crop diversification and livestock integration | 6,6% |

**Table 4.** Significance of the difference in food security prevalence across agroecological practice categories.

| | | Agricultural classes | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Conventional farms | Conventional farms with good crop diversification & CES_DRS | Conventional farms with diversified livestocks | Mixed farms, no clear paths | Sustainable farms, with no or poor integrated (trees and livestocks) | Sustainable farms with a medium crop diversification and livestock integration | Agroecology farms with good crops diversification and livestock integration | Agroecology farms with a high crop diversification and livestock integration |
| | | (A) | (B) | (C) | (D) | (E) | (F) | (G) | (H) |
| Food security status | In food security | E( ,006)<br>F( ,000) | E( ,000)<br>F( ,000) | E( ,004)<br>F( ,000) | E( ,001)<br>F( ,000) | | | A( ,002)<br>C( ,004)<br>D( ,005)<br>E( ,000)<br>F( ,000) | E( ,005)<br>F( ,002) |
| | In limited food security | | | | | | A( ,004)<br>B( ,003)<br>C( ,025)<br>D( ,001)<br>G( ,001) | | |
| | Food insecure | G( ,006) | | | G( ,036) | A( ,036)<br>B( ,006)<br>C( ,002)<br>D( ,002)<br>G( ,000)<br>H( ,005) | C( ,036)<br>D( ,036)<br>G( ,002)<br>H( ,039) | | |

Results are based on two-sided tests. For each significant pair, the key of the category with the smaller column proportion appears in the category with the larger column proportion.
Significance level for upper case letters (A, B, C, D, E, F, G, H.): ,05

*Tests are adjusted for all pairwise comparisons within a row of each innermost subtable using the Benjamini-Hochberg correction.

**Performance of prediction models**

Overall, the trained models performed well. On the test data, the models recorded accuracies varying from 64% for the MLP model to 72% for the XGboost model for the global performances (Table 5). In the absence of food consumption and diversity scores, the models trained in the present study can predict the food security status of farms in at least 70% of cases without error for the assembly models, and in at least 63% of cases for the other models. Macro average F1_score and Recall values are around 50% for most models, except XGBoost, which has a Recall of 72% and an F1_score around 60%, making this model a better candidate for future automation. Figure 3 shows the confusion matrices of selected models, highlighting the predictions for the different classes on the test data. Although model convergence times were relatively low (less than a minute) for most algorithms, it is worth noting that SVM models recorded the longest times (over 30 minutes), which can be explained by the training mechanism of the SVM model. Given that we were unable to imagine in advance the values of the hyperparameters that optimize these algorithms, we compiled a list of values so that, during the training stage, several combinations of values were made before identifying the most appropriate ones. The parameter values for the other models were set at levels used by previous studies and were found to be optimal. The average recall and F1_score values obtained demonstrate the natural imbalance between food security statuses. Naturally, under desirable conditions, there are fewer severely food-insecure households than food-secure, and the opposite should not be the case; otherwise, we would be facing an exceptionally undesirable humanitarian catastrophe.

**Table 5.** Overall model performance on the test set (macro average)

| | Accuracy | Recall | Precision | F1_score |
|---|---|---|---|---|
| SVM_L | 66,5% | 49,3% | 62,0% | 49,6% |
| SVM_R | 67,3% | 43,6% | 69,4% | 47,4% |
| MLP | 64,1% | 53,0% | 49,0% | 51,0% |
| Adaboost | 70,6% | 41,4% | 75,2% | 43,4% |
| XGBoost | 72,2% | 72,0% | 50,7% | 56,4% |

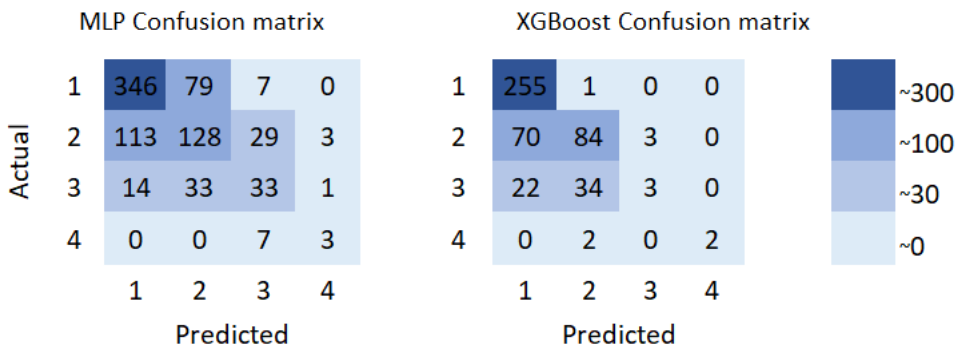


*Fig. 3: MLP and XGBoost confusion matrices highlighting predicted class levels based on the test data.*

When examining the top predictors of food security, we found that certain practices associated with agroecology are particularly important (in bold in Table 6). These are the total number and type of livestock, use of local seeds, rotation practices, thinning practices, in-line seeding, number of plots, agro-sylvo-pastoral species richness, and production diversification. In general, the amount and the source of funds used to acquire inputs, self-consumption of own production, the value of agricultural products, the share of food in total expenditure, and the use of survival strategies, such as the sale of certain movable assets during periods of financial instability, were commonly identified as the top predictors for all models. In the XGBoost and Adaboost models, cereal yield, number of animals owned, number of agricultural plots, product diversity, and number of household members were among the variables that contributed most to identifying food security status. For the SVM models, the number of income sources, the diversity of agricultural products, the agroecological intensity, the use of local seeds, the number of animals, the use of paid labor, and the source of input financing contributed the most to identifying the food security status of farms. Finally, the type of input, the number of plots, the type of livestock, crop loss factors, access to extension services, the number of sources of revenues, and the type of ploughing were among the variables with the highest contributions in the MLP model.

**Table 6.** Top 15 predictors contributing to the constitution of food security classes for the models.

| XGB | ADABOOST | MLP | SVM_L | SVM_R |
|---|---|---|---|---|
| FES | FES | FES | FES | FES |
| LCS | LCS | LCS | LCS | LCS |
| Amount of input purchase from own funds | Amount of inputs purchased from own funds | **Type of inputs used** | **Fallow duration** | Number of revenue sources |
| Proportion of agro-sylvo-pastoral products self-consumed | Value of rainfed crop product sales | The impact of a loss factor | **Total number of livestock** | **Agro-sylvo-pastoral species richness** |
| Self-consumption of rainfed products | Proportion of agro-sylvo-pastoral products self-consumed | Self-consumption of rainfed products | **Number of livestock sold** | Amount of inputs purchased from own funds |
| Production | Value of agro-sylvo-pastoral products sales | Amount of inputs purchased from own funds | Number of labor (men day) | **Use of local seed** |
| Value of rainfed crop products sales | **Value of the livestock sales** | **Types of livestock** | The source of funds used to purchase the input | Proportion of agro-sylvo-pastoral products self-consumed |
| Yield of farm main cereals | **Agro-sylvo-pastoral species richness** | **Number of plots** | Value of agro-sylvo-pastoral product sales | Value of rainfed crop product sales |
| Value of agro-sylvo-pastoral products sales | **Number of plots** | **Thinning** | **Rotation** | **Fallow duration** |
| **Total number of livestock** | **Total number of livestock** | Yield of farm main cereals | Access to extension services | Self-consumption of rainfed products |
| **Value of the livestock sale** | Self-consumption of rainfed products | Number of revenue sources | **Value of livestock sale** | **Number of livestock that are self-consumed** |
| Age of index woman | Yield of farm main cereals | **Rotation** | Type of roof | Proportion of agro-sylvo-pastoral products sales |
| **Number of plots** | Number of livestock that are self-consumed | Age of index woman | **Agroecology intensification class** | Main source of revenue |
| Number of household members | **In-line seeding** | Access to extension services | Amount of input purchased from own funds | Proportion of rainfed products sold |

| | | | | |
|---|---|---|---|---|
| Proportion of rainfed products sold | Number of revenue sources | Type of ploughing | **Use of local seed** | Type of wall |

The Partial Dependence Plots (PDP) below show how the probability of being in a specific food security class changes with increasing each predictor individually while keeping the rest constant. Overall, the probability of belonging to class 1 (around 32%) is higher than that of class 2 (around 27%), and this remains constant up to class 4 (around 15%), reflecting how important each class is in the database. The probability of belonging to the food-secure class gradually increases from 0.32 to 0.35 as species richness increases from 0 to 4 (Fig.4). This means that higher agricultural product diversification is positively associated with increased food security, and the more diversified the crops, the more likely the household is to be food secure. Moreover, households relying on multiple activities (crops + livestock + trees + off-farm work) are more likely to be food secure. Indeed, results suggest that the integration of livestock and the ability to sell up to 300,000 FCFA worth of livestock are associated with a 3% improvement in household food security (Fig.4), while the contribution of agroforestry is very marginal but still positive (Annex Fig. C1).

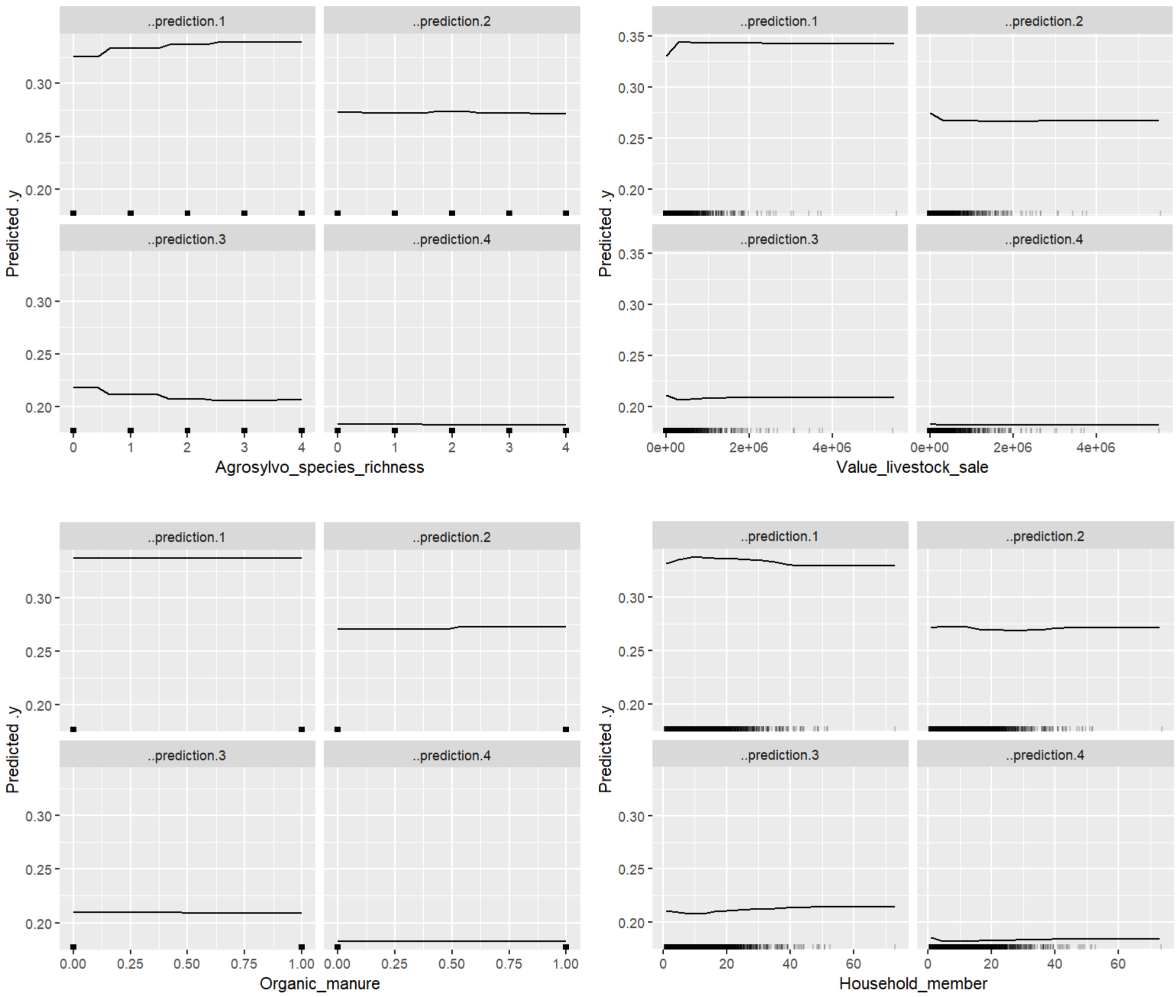


*Fig. 4: Partial dependence plots for selected predictors highlighting their marginal effects on food security categories.*

Off-farm work diversifies household income sources, increasing the probability of food security (Fig. C1). Although having a marginal positive effect on the probability of being in the limited food secure class, organic manure alone is not sufficient to improve household food security and may require a combination of other practices (Fig.4). Among the socio-demographic variables, we found that the age of the index woman negatively affects households' food security. In fact, for young and middle-aged women (around 25-50 years), there is a progressive increase in the probability of food security, while the probability decreases after the age of 50, reflecting changes in socio-economic status (Fig. C1). Beyond a certain level (8 members), an increase in household size increases the risk of food insecurity by 2% (Fig.4), as household members initially constitute a labor force that can help to produce more. We also found that when the amount allocated to input acquisition increases up to 700,000 FCFA, the probability of being food-secure increases by around 5% (Fig. C1). This suggests that investment in inputs initially has a positive impact on food security, but beyond a certain level, it can affect household savings in the absence of credit.

Although we have the possibility of further improving the performance of the various models by integrating food consumption and diversity scores as predictors (Nikiema, Ezin, et al., 2024), we opted for a configuration where such information is not available to analyse their ability to predict the composite food security indicator, essentially based on households' agricultural activities. In such a context, the accuracy obtained remains acceptable. National agricultural surveys are very costly due to the length of the questionnaires, the collection times, and the number of staff required for data collection. Thus, initiatives aimed at shortening the length and complexity of these operations while maintaining accuracy are critical.

## 4. Discussion

In this study, we propose an approach to explore the role of agroecological practices in predicting food security, without relying on food consumption scores, exploring one of the recommendations made by Deléglise et al. (2023). This represents a critical contribution to the targeting of priority areas of intervention by policymakers, especially in contexts where there is a growing need to reduce the length of agricultural surveys (UNSD, 2018, 2022). In general, the algorithms presented in this study perform better than other studies in the existing literature (Hossain et al., 2019; Martini et al., 2022). However, it is important to note that our modelling approach does not aim to replace the composite food security indicator calculated by WFP, which remains the gold standard for assessing household food security status fully. Instead, our results demonstrate that even in situations where key components such as FCS are unavailable, models leveraging other collected indicators and household agricultural practices can still provide meaningful predictions with reasonable accuracy (72%). This can support decision-makers with indicative assessments, especially in a context of budget constraints.

Our results suggest that, in general, farms practicing agroecology are less affected by food insecurity than those practicing conventional agriculture. The importance of certain demographic characteristics, such as age, household size, and education level, agrees with previous literature (Abdullah et al., 2019; Maziya et al., 2017). Less than 20% of farms have access to agricultural credit and extension services. They are forced to reduce their purchasing power to acquire agricultural inputs. It is therefore necessary to intensify agricultural advisory activities in line with the National Agroecology Development Strategy (MARAH, 2023) and revitalize the agricultural credit system. In the same vein, Osei Mensah et al. (2013) and Tesfaye & Nayak (2022) show that the possession of an additional source of income and access to agricultural credit for the acquisition of agricultural inputs improve food security.

Households that practice integrated livestock farming are less affected by food insecurity than households that do not (11% versus 28% respectively). These positive effects confirm the findings of Puech & Stark (2023) and Vall et al. (2017). Diversification of crop species showed a significant difference in food security compared with monoculture (10% for farms with at least 5 crop species compared with 33% for those with 1 crop species). Thus, our findings go beyond those obtained by Ayenew et al. (2018) and Ng'endo et al. (2015) to confirm a positive association between crop diversity and food security. Taken individually, certain agroecological practices known for their environmental benefits, such as crop rotation, organic manure, or stone barriers for soil and water conservation, are not significantly associated with food security. However, the combination of several practices simultaneously characterized in high agroecological intensity proved to be an important predictor of food security. Kerr et al (2021) obtained a similar result from a literature review including 56 articles on low and middle-income countries.

While the study primarily focused on assessing the contribution of agroecological practices, it is important to note that some non-agroecological variables were among the top predictors identified by our models. For instance, the value of product sales, the number of household members, and the amount used to purchase the inputs reflect the economic capacity of households. Additionally, the source of funds used to purchase inputs and the access to extension services emerged as top predictors, indicating the role of institutional and financial access in shaping production decisions and food security. The number of revenue sources also underlines the importance of income diversification strategies in enhancing food security resilience through off-farm activities.

Although the eight farm sub-groups we constructed reflect a continuum, we believe they offer a practical framework to design and prioritise targeted policies for improving food security and advancing agroecological sustainability. Initially, we grouped farms into only three broad agroecology categories. However, this approach failed to capture meaningful differences in food security outcomes because these groups were internally heterogeneous. By developing the current eight sub-groups, we were able to achieve greater homogeneity within each class. For instance, for the group of farms with poor crop diversification, poor tree, and poor livestock integration, policies can incentivise agroforestry adoption, crop diversification, and improved livestock integration to strengthen ecological functions. Finally, there are some differences in the top predictors identified by the different models. This variation reflects the complementary strengths of different algorithms in capturing the complex and non-linear relationships inherent in farming systems (the learning mechanisms of the models, as explained in the appendix A1 to A2, are not the same). Importantly, differences in variable rankings suggest that there is no single pathway to improving food security. This finding is valuable for policymakers, as it indicates the need for context-specific, tailored interventions that leverage the most relevant agroecological practices to enhance food security outcomes.

This paper provides a step change in the quantitative modelling of the role that agroecology can play in achieving food security worldwide, and particularly in the context of low-income countries in sub-Saharan Africa. In Burkina Faso, previous analyses show that regions exposed to adverse climatic and environmental conditions (low rainfall, erosion, high evapotranspiration and temperature) are those with a higher rate of adoption of agroecological practices compared to the national adoption rate, which is below 10% (Debray et al., 2015; Nikiema, Ezin, & Chogou, 2024). The conclusions of this study provide further evidence of the role of different agroecological practices and inform further adoption.

## 5. Limitations

One limitation of this study is that models were validated using a hold-out sample from the same dataset, without temporal or geographic external validation. However, it is important to note that models were trained following a standard machine learning approach, where the test sets used to test model performance are treated as unseen data during the modelling process. For application to other geographic contexts or countries, it will also be essential to ensure that comparable variables or appropriate proxies are collected. These models can then be applied to data from other years and geographic areas, provided that the required variables are available in the same format, with names and codes consistent with those used in this study.

Another limitation of this study is that, while class imbalance was acknowledged, we did not apply data resampling techniques such as SMOTE or class weighting during model training. Instead, we prioritized reporting robust class-specific performance metrics, including macro average (F1-score, recall, and precision), to provide a transparent and balanced evaluation of the models' accuracy across the four classes of food security, without artificially altering the distribution of the data with resampling methods. This approach allows us to assess how the models perform on each class on the original data distribution. Future research could explore integrating resampling or weighting strategies to potentially improve model performance on imbalanced classes and further validate the findings.

Finally, this study does not establish a direct causality between agroecological practices and household food security outcomes. While our analyses demonstrate significant associations between the variables, the findings should not be interpreted as evidence of a direct causal effect. Future research employing experimental or quasi-experimental designs, such as randomized controlled trials or natural experiments, would be needed to rigorously assess the causal effects of agroecological intensification on household food security. Nonetheless, our findings provide valuable empirical evidence suggesting that agroecological practices are associated with improved food security, highlighting their potential as an important area of intervention.

## 6. Conclusion

In this paper, two complementary analytical approaches were employed to explore the role of agroecology in enhancing food security. First, a descriptive analysis was conducted to explore the associations between agroecological practices and household food security outcomes. This involved examining the distribution of key agroecological variables across food security categories and identifying patterns of association in the data. Second, a predictive analysis was implemented using machine learning algorithms to model and predict household food security status based on agroecological and socio-demographic variables, illustrating the methods with a nationally representative dataset from Burkina Faso. This approach aimed to assess the predictive power of agroecological practices for food security outcomes and to identify the importance of the predictors in determining household food insecurity status.

Our results indicate that the overall prediction accuracy of the composite food security status rises from 63% to 72%. We also find that even without some primary information on households' food consumption patterns, such as their food consumption score and dietary diversity, the models trained can predict the composite state of food security based on households' agricultural variables, which are easy to collect and quantify. Yield, integration of crops with livestock and trees, self-consumption, food expenditure share, the income from the sale of agricultural products, and the source of funds used to acquire agricultural inputs are among the top variables contributing to the model learning process.

Additionally, our findings also indicate that households with good agroecological intensity are less affected by food insecurity compared to others, suggesting that agroecological practices are associated with improved food security status. However, it is important to note that this study does not establish a causal relationship between agroecological practices and food security improvements. The analytical methods employed are correlational and predictive, and causal inference would require experimental or quasi-experimental designs, which are beyond the scope of this study. Although the majority of socio-demographic variables included in the study did not contribute significantly to the model learning process, this analysis could be enriched by additional factors such as household access to water and electricity, and information on the effect of distance on farmers' product commercialization and profitability.

The algorithms used in this study not only make it possible to provide timely indicators of household food security status but could also generate real gains in a context of budgetary constraints, as is the case with most national agricultural surveys in low-income countries, which are very often forced to restrict questionnaire length and collection costs.

The implications of these findings are multiple. First, they highlight the potential for agricultural surveys and food security monitoring systems in low-income countries to leverage easily collectable agricultural data for the assessments of food security, especially in a context where we are seeking to reduce the length of data collection tools, which entail high financial or logistical costs. Second, our results support policies promoting agroecological intensification as a pathway to improve household food security, indicating that investments in training and incentivizing farmers to combine multiple agroecological practices could have a positive impact on food security outcomes. Third, by demonstrating the feasibility of predicting food security status using primarily agricultural variables, this research opens avenues for the integration of machine learning-based predictions into food security assessment.

Future food security monitoring initiatives using machine learning algorithms could explore the use of open-access data sources, such as market price data and remote sensing-derived indicators, in their predictive modelling frameworks. Leveraging these available data can enable timely, cost-effective, and spatially extensive early warning systems, providing preliminary estimates of household food security status between survey rounds.

## References


Abdullah, Zhou, D., Shah, T., Ali, S., Ahmad, W., Din, I. U., & Ilyas, A. (2019). Factors affecting household food security in rural northern hinterland of Pakistan. *Journal of the Saudi Society of Agricultural Sciences*, *18*(2), 201-210. https://doi.org/10.1016/j.jssas.2017.05.003

Altieri, M. A., Nicholls, C. I., Henao, A., & Lana, M. A. (2015). Agroecology and the design of climate change-resilient farming systems. *Agronomy for Sustainable Development*, *35*(3), 869-890. https://doi.org/10.1007/s13593-015-0285-2

Andree, B., Chamorro, A., Kraay, A., Spencer, P., & Wang, D. (2020). *Predicting Food Crises*. https://doi.org/10.1596/1813-9450-9412

Ayenew, H. Y., Biadigilign, S., Schickramm, L., Abate-Kassa, G., & Sauer, J. (2018). Production diversification, dietary diversity and consumption seasonality : Panel data evidence from Nigeria. *BMC Public Health*, *18*, 988. https://doi.org/10.1186/s12889-018-5887-6

Bernard, B., & Lux, A. (2017). How to feed the world sustainably : An overview of the discourse on agroecology and sustainable intensification. *Regional Environmental Change, 17*(5), 1279-1290. https://doi.org/10.1007/s10113-016-1027-y

Bezner Kerr, R., Madsen, S., Stüber, M., Liebert, J., Enloe, S., Borghino, N., Parros, P., Mutyambai, D. M., Prudhon, M., & Wezel, A. (2021). Can agroecology improve food security and nutrition? A review. *Global Food Security*, *29*, 100540. https://doi.org/10.1016/j.gfs.2021.100540

Dalgaard, T., Hutchings, N. J., & Porter, J. R. (2003). Agroecology, scaling and interdisciplinarity. *Agriculture, ecosystems & environment*, *100*(1), 39-51.

Debray, V., Derkimba, A., & Roesch, K. (2015). *Des innovations agroécologiques dans un contexte climatique changeant en Afrique*. https://www.fao.org/sustainable-food-value-chains/library/details/en/c/427589/

Deléglise, H., Bazié, Y., Bégué, A., Interdonato, R., Roche, M., Teisseire, M., & Maitre d'hotel, E. (2023). Validity of household survey indicators to monitor food security in time and space : Burkina Faso case study. *Agriculture & Food Security*, *11*. https://doi.org/10.1186/s40066-022-00402-4

FAO. (2019). *Tool for Agroecology Performance Evaluation (TAPE)—Test version*. FAO. https://openknowledge.fao.org/handle/20.500.14283/ca7407en

FAO, IFAD, UNICEF, WFP, & WHO. (2024). *The State of Food Security and Nutrition in the World 2024*. https://openknowledge.fao.org/handle/20.500.14283/cd1254en

HLPE. (2019). Agroecological and other Innovative Approaches for Sustainable Agriculture and Food Systems that Enhance Food Security and Nutrition. *High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security: Rome, Italy*.

Hossain, M., Mullally, C., & Asadullah, M. N. (2019). Alternatives to calorie-based indicators of food security : An application of machine learning methods. *Food Policy, 84*, 77-91. https://doi.org/10.1016/j.foodpol.2019.03.001

Ickowitz, A., Powell, B., Rowland, D., Jones, A., & Sunderland, T. (2019). Agricultural intensification, dietary diversity, and markets in the global food security narrative. *Global Food Security*, *20*, 9-16. https://doi.org/10.1016/j.gfs.2018.11.002

INSD. (2022). *Rapport du 5e Recensement Général de la Population et de l'Habitat : Caractéristiques des ménages et de la population*.

Kabore, P. N., Ouedraogo, A., Sanon, M., Yaka, P., & Some, L. (2017). Caractérisation de la variabilité climatique dans la region du centre-nord du burkina faso entre 1961 et 2015. *Climatologie*, *14*, 82-95. https://doi.org/10.4267/climatologie.1268

Kabré, B., Belem/Ouédraogo, M., Lankoandé, B., & Ouédraogo, A. (2020). Variabilité démographique de Saba senegalensis (A. DC.) Pichon suivant le gradient climatique au Burkina Faso. *BOIS & FORETS DES TROPIQUES*, *345*, 75-85. https://doi.org/10.19182/bft2020.345.a31930

Kopittke, P. M., Menzies, N. W., Wang, P., McKenna, B. A., & Lombi, E. (2019). Soil and the intensification of agriculture for global food security. *Environment International*, *132*, 105078. https://doi.org/10.1016/j.envint.2019.105078

Lentz, E. C., Michelson, H., Baylis, K., & Zhou, Y. (2019). A data-driven approach improves food insecurity crisis prediction. *World Development*, *122*, 399-409. https://doi.org/10.1016/j.worlddev.2019.06.008

Liu, W., Wang, J., Li, C., Chen, B., & Sun, Y. (2019). Using Bibliometric Analysis to Understand the Recent Progress in Agroecosystem Services Research. *Ecological Economics*, *156*, 293-305. https://doi.org/10.1016/j.ecolecon.2018.09.001

Lucantoni, D. (2020). Transition to agroecology for improved food security and better living conditions : Case study from a family farm in Pinar del Río, Cuba. *Agroecology and Sustainable Food Systems*, *44*(9), 1124-1161. https://doi.org/10.1080/21683565.2020.1766635

MARAH, M. de l’Agriculture, des Ressources Animales et Halieutiques. (2023). *Stratégie Nationale de Développement de l’Agroécologie (SND-AE) 2023-2027* (FAO). https://www.fao.org/faolex/results/details/en/c/LEX-FAOC217823/

Martini, G., Bracci, A., Riches, L., Jaiswal, S., Corea, M., Rivers, J., Husain, A., & Omodei, E. (2022). Machine learning can guide food security efforts when primary data are not available. *Nature Food*, *3*(9), 716-728.

Maziya, M., Mudhara, M., & Chitja, J. (2017). What Factors Determine Household Food Security Among Smallholder Farmers? Insights from Msinga, KwaZulu-Natal, South Africa. *Agrekon*, *56*, 1-13. https://doi.org/10.1080/03031853.2017.1283240

Mugwanya, N. (2019). Why agroecology is a dead end for Africa. *Outlook on Agriculture*, *48*, 113-116. https://doi.org/10.1177/0030727019854761

Muhumuza, J. R. (2023). Why current agroecology rhetoric stands to protract farmer poverty in the developing world. *Outlook on Agriculture*, *52*(3), 303-310. https://doi.org/10.1177/00307270231195381

Ng'endo, M., Keding, G. B., Bhagwat, S., & Kehlenbeck, K. (2015). *Variability of on-farm food plant diversity and its contribution to food security : A case study of smallholder farming households in Western Kenya*. https://hdl.handle.net/10568/72451

Nikiema, T., Ezin, E. C., & Chogou, S. K. (2024). Categorizing Farms to Promote Agroecology : A Supervised Learning Approach. *IEEE*, 326-331. https://doi.org/10.1109/SMARTCOMP61445.2024.00076

Nikiema, T., Ezin, E. C., & Kpenavoun Chogou, S. (2023). Bibliometric Analysis of the State of Research on Agroecology Adoption and Methods Used for Its Assessment. *Sustainability*, *15*(21), Article 21. https://doi.org/10.3390/su152115616

Nikiema, T., Ezin, E., Kpenavoun Chogou, S., & Katic, P. (2024). *Farm Households Food Security Status Automation Through Supervised Learning Approach : A Look at Agroecological Farms* (p. 325). https://doi.org/10.1109/SMARTCOMP61445.2024.00075

Osei Mensah, J., Aidoo, J., & Tuffour, R. (2013). *Determinants of Household Food Security in the Sekyere-Afram Plains District of Ghana*.

Palomo-Campesino, S., García-Llorente, M., & González, J. A. (2021). Characterizing agroecological and conventional farmers : Uncovering their motivations, practices, and perspectives toward agriculture. *Agroecology and Sustainable Food Systems*, *45*(9), 1399-1428. https://doi.org/10.1080/21683565.2021.1933671

Pronti, A., & Coccia, M. (2021). Multicriteria analysis of the sustainability performance between agroecological and conventional coffee farms in the East Region of Minas Gerais (Brazil). *Renewable Agriculture and Food Systems*, *36*(3), 299-306. https://doi.org/10.1017/S1742170520000332

Puech, T., & Stark, F. (2023). Diversification of an integrated crop-livestock system : Agroecological and food production assessment at farm scale. *Agriculture, Ecosystems & Environment*, *344*, 108300. https://doi.org/10.1016/j.agee.2022.108300

Tesfaye, T., & Nayak, D. (2022). Does participation in non-farm activities provide food security? Evidence from rural Ethiopia. *Cogent Social Sciences*, *8*(1), 2108230. https://doi.org/10.1080/23311886.2022.2108230

Tscharntke, T., Clough, Y., Wanger, T. C., Jackson, L., Motzke, I., Perfecto, I., Vandermeer, J., & Whitbread, A. (2012). Global food security, biodiversity conservation and the future of agricultural intensification. *Biological Conservation*, *151*(1), 53-59. https://doi.org/10.1016/j.biocon.2012.01.068

UNSD. (2018). *Guidelines on the use of electronic data collection technologies in population and housing censuses*. https://eeca.unfpa.org/sites/default/files/pubpdf/guidelines_on_the_use_of_electronic_data_collection_technologies_in_population_and_housing_censuses.pdf

UNSD. (2022). *Modernization of the Production of Time-use Statistics*. https://unstats.un.org/unsd/statcom/53rd-session/documents/BG-3h-Modernization_UN_EG_TUS2021_FINAL_SENT_rev-E.pdf

Vall, E., Marre-Cast, L., & Kamgang, H. J. (2017). Chemins d'intensification et durabilité des exploitations de polyculture-élevage en Afrique subsaharienne : Contribution de l'association agriculture-élevage. *Cahiers Agricultures*, *26*(2), Article 2. https://doi.org/10.1051/cagri/2017011

Waha, K., van Wijk, M. T., Fritz, S., See, L., Thornton, P. K., Wichern, J., & Herrero, M. (2018). Agricultural diversification as an important strategy for achieving food security in Africa. *Global Change Biology*, *24*(8), 3390-3400. https://doi.org/10.1111/gcb.14158

Wezel, A., Herren, B. G., Kerr, R. B., Barrios, E., Gonçalves, A. L. R., & Sinclair, F. (2020). Agroecological principles and elements and their implications for transitioning to sustainable food systems. A review. *Agronomy for Sustainable Development*, *40*(6), 40. https://doi.org/10.1007/s13593-020-00646-z

WFP. (2021). *Technical Guidance For WFP's Consolidated Approach for Reporting Indicators of Food Security (CARI)* (Third). WFP.

WFP. (2024a). *Food Consumption Score & Food Consumption Score Nutritional Analysis : Guidance Note*. World Food Programme. https://docs.wfp.org/api/documents/WFP-0000158062/download/

WFP. (2024b). *WFP Burkina Faso Country Brief, September 2024*. https://reliefweb.int/report/burkina-faso/wfp-burkina-faso-country-brief-september-2024

**Conflict of interest statement**

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.

**Funding**

This work was supported by the Partnership for Skills in Applied Sciences, Engineering, and Technology (PASET) Regional Scholarship Innovation Fund (RSIF) awarded to Mr Theodore Nikiema.

**Author contributions**

Theodore Nikiema led the methodological advancement, analysis of the data and writing. Pamela Katic was responsible for the development of the research questions and conceptualisation of the paper and contributed to the writing. Sylvain Kpenavoun Chogou and Eugene C. Ezin supported the foundations of earlier work that this paper builds upon and provided valuable feedback into various drafts of the paper.

**Data availability**

The data used in this study were obtained from the Department of Statistics of Burkina Faso's Ministry of Agriculture, and interested parties are invited to contact the department.

## APPENDIX

## Appendix A: Overview of Machine Learning Algorithms

### A.1 Extreme Gradient Boosting (XGBoost)

**Description**

The principle of the XGBoost model is to train decision trees sequentially so that with each new iteration, a new additionnal tree is trained based on the previous tree. In this way, the tree of an iteration *n+1* is trained to correct the imperfections and errors of the tree of step *n*.

**Mathematical Formulation**:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i) \tag{2}$$

$$obj^{(t)} = \sum_i^n l(y_i, \hat{y}_i{}^{k-1} + f_t(x_i)) + \Omega(f_t)\ , \tag{3}$$

*K* is the number of trees, $f$ belongs to the set of decision trees space, $x_i$ denotes the input features, and $\hat{y}_i$ denotes the prediction. The first term of the objective function is the loss function, which measures the difference between the prediction $\hat{y}_i$ and the target $y_i$ while the second term $\Omega(f_t)$ is the regularization parameter. The goal is to optimize the objective function $obj^{(t)}$.

### A.2 Adaptive Boosting (Adaboost)

**Description**

Adaboost is a boosting technique among ensemble methods like XGBoost. It is called Adaptive Boosting as the weights are re-assigned to each instance, with higher weights assigned to incorrectly classified instances. At the beginning, we assign equal weights $\omega_i = \frac{1}{N}$ to all data points, where *N* is the total number of data points. For *t* in range 1, …, *T*, where *T* is the number of iterations, we train a weak learner (for example, a decision stump) on the training data with the current weights.

**Mathematical Formulation**:

$$\epsilon_t = \sum_{i=1}^{N} \omega_i * 1(y_i \neq h_t(x_i))\ , \tag{4}$$

$$\alpha_t = \frac{1}{2} ln\ (\frac{1-\epsilon_t}{\epsilon_t}) \tag{5}$$

$$\omega_{i,t+1} = \sum_{i=1}^{N} \omega_{i,t}\ .\ exp(-\alpha_t\ .\ y_i\ .\ h_t(x_i))\ , \tag{6}$$

$$\omega_{i,t+1} = \frac{\omega_{i,t+1}}{\sum_j^N \omega_{j,t+1}} \tag{7}$$

$$H(x) = sign(\sum_{t=1}^{T} \alpha_t\ * h_t(x_i)) \tag{8}$$

where $y_i$ is the true label, $\alpha_t$ is the calculated weighted error derived from the weak learner. $h_t(x_i)$ is the prediction of the weak learner, and 1(.) is the indicator function. $\omega_{i,t+1}$ are the updated weights of the data points for the next iterations, which are normalized through (7). $H(x)$ is the classifier resulting from the selection process when the algorithm stops at iteration T. It is a weighted sum of weak learners.

### A.3 Multi Layer Perceptron (MLP)

**Description**

A Multilayer Perceptron (MLP) consists of a combination of one input layer, one or more hidden layers, and one output layer. In this study, we trained a model with one hidden layer. The algorithm of the model can be explained as follows. Initially, the input vector x is assigned to the input layer. Then, each node in the hidden layer computes a weighted sum of the inputs and applies an activation function. Let $\omega_{i,j}{}^{(1)}$ be the weight connecting the i-th input to the j-th hidden node, and $b_j{}^{(1)}$ be the bias of the j-th hidden node and $\omega_{j,k}{}^{(2)}$ be the weight connecting the *j*-th hidden node to the *k*-th output and $b_k{}^{(2)}$ be the bias of the *k*-th output.

**Mathematical Formulation**:

$$\mathcal{Z}_j{}^{(1)} = \sum_i \omega_{ij}{}^{(1)}.\, x_i + b_j{}^{(1)} \qquad (9)$$

$$\mathcal{Z}_k{}^{(2)} = \sum_j \omega_{jk}{}^{(2)} * h_j + b_k{}^{(2)} \qquad (10)$$

Where $\mathcal{Z}_j{}^{(1)}$ is the weighted sum for the j-th hidden node and $\mathcal{Z}_k{}^{(2)}$ is the weighted sum for the k-th output node. The output of the hidden node is $h_j{}^{(1)} = f(Z_j{}^{(1)})$ and the output of the second node (output node) is $h_k{}^{(2)} = f(Z_k{}^{(2)})$, where *f* is the activation function (commonly sigmoid, tanh, or ReLU).

**A.4 Support Vector Machine**

The SVM model consists in identifying the optimal boundary (hyperplane) separating the classes in the feature space. It is a reliable way of modeling circumstances in which classes cannot be linearly separated. Given a training set of instance-label pairs $(x_i, y_i)$, $i$ = 1, ..., l; where $x_i \in R^n$ and $y_i \epsilon \{class_1, \dots, class_P\}^l$, the SVM model requires the solution of the following optimization problem.

**Mathematical Formulation**:

$$\underset{w,b,\xi}{\mathrm{Min}}\left(\frac{1}{2} w^T w + c \sum_{i=1}^{l} \xi_i\right)$$

$$\text{subject to } \; y_i(w^T \phi(x_i) + \mathrm{b}) \geq 1 - \xi_i \qquad (11)$$

*Linear kernel:*

$$\mathrm{K}(x_i, x_j) = x_i{}^T x_j \qquad (12)$$

*Radial Kernel:*

$$\mathrm{K}(x_i, x_j) = \exp\left(-\gamma \parallel x_i - x_j \parallel^2\right), \gamma \geq 0 \qquad (13)$$

where $\xi_i \geq 0$, C is a regularization parameter which allows us to minimize the training error. Training vectors $x_i$ are mapped into a higher dimensional space by the function $\phi$. $\mathrm{K}(x_i, x_j) = \phi(x_i) T \phi(x_j)$ is called the kernel function. As food insecurity is multidimensional, we have chosen to use the two kernel functions below, which allow us to compare the non-linear separator with the linear separator.

**Appendix B: Supplementary information**

**Table B.1.** Descriptive statistics.

| Variables | Proportion % | * (Mode) |
|---|---|---|
| **Land and water conservation practices** | | |
| No | 62,1 | * |
| Stony ridge | 15,0 | |

| | | |
|---|---|---|
| Filter dike | 1,0 | |
| Half moon | 0,4 | |
| Zaï | 4,9 | |
| Soil bead | 2,5 | |
| Grass strip | 4,1 | |
| Dead/live hedge | 8,7 | |
| Others | 1,3 | |
| **Main equipment** | | |
| Basics (ploughs, seeders, hoes) | 64,6 | * |
| Draught animals | 25,5 | |
| Motorized equipment and personal source of water | 9,9 | |
| **Organic manure** | | |
| No | 30,6 | |
| Yes | 69,4 | * |
| **Type of input** | | |
| Biological | 15,9 | |
| Non-biological | 84,1 | * |
| **Range of agroforestry trees held** | | |
| Between 0-5 trees | 17,6 | |
| Between 6-19 trees | 27,2 | * |
| Between 20-50 trees | 26,8 | |
| Between 50 and 100 | 15,3 | |
| More than 100 trees | 13,1 | |
| **Range of crop species richness held** | | |
| 0 or 1 specie | 4,1 | |
| 2-4 species | 33,9 | * |
| 5-7 species | 41,6 | |
| 8-10 species | 16,6 | |
| more than 10 species | 3,9 | |
| **Range of agro-sylvo-pastoral species richness held** | | |
| 0 or 1 specie | 9,7 | |
| 2-5 species | 39,4 | * |
| 6-10 species | 31,1 | |
| 11-15 species | 14,3 | |
| more than 15 species | 5,5 | |
| **Range of livestock species richness held** | | |
| 0 or 1 specie | 10,7 | |
| 2-3 species | 34,2 | * |
| 4-6 species with less than 50 animals | 26,4 | |
| 4-6 species with more than 50 animals | 24,4 | |
| more than 6 species | 4,4 | |
| **Use of local seeds** | | |
| Non | 25,8 | |
| Oui | 74,2 | * |

**Table B.2.** Descriptive statistics (continued)

| Variables | Proportion % | * (Mode) |
|---|---|---|
| **Geographical region** | | |
| Boucle Du Mouhoun | 12,1 | * |
| Cascades | 2,5 | |
| Centre | 3,7 | |
| Centre - Est | 8,7 | |
| Centre-Nord | 10,9 | |
| Centre-Ouest | 8,8 | |

| | | |
|---|---|---|
| Centre-Sud | 6,0 | |
| Est | 11,8 | |
| Hauts-Bassins | 7,4 | |
| Nord | 7,4 | |
| Plateau Central | 6,3 | |
| Sahel | 8,9 | |
| Sud-Ouest | 5,5 | |
| **Crop rotation practices** | | |
| No | 87,7 | * |
| Yes | 12,3 | |
| **Crop Association practices** | | |
| No | 75,3 | * |
| Yes | 24,7 | |
| **Microdose practices** | | |
| No | 89,4 | * |
| Yes | 10,6 | |
| **Sarclo-binage practice or Thinning** | | |
| No | 36,5 | |
| Yes | 63,5 | * |
| **Type of livestock** | | |
| No livestocks | 5,8 | |
| Extensive sedentary livestock farming | 70,1 | * |
| Semi-intensive livestock farming | 16,0 | |
| Intensive sedentary livestock farming | 6,2 | |
| Transhumance | 1,9 | |
| **Matrimonial status of the index woman** | | |
| Single | 9,9 | |
| Married monogamous | 53,1 | * |
| Polygamous married | 35,9 | |
| Widow | 0,5 | |
| Divorced/Separated | 0,2 | |
| Free union | 0,4 | |
| **Instruction level of the index woman** | | |
| Non-literate | 69,5 | * |
| alphabetized or rural studies | 9,2 | |
| Primary level | 13,8 | |
| Medersa | 1,7 | |
| High-school or + | 5,8 | |
| **Main source of inputs funds** | | |
| Own fund | 82,8 | * |
| Credit | 17,2 | |
| **Type of roofs** | | |
| Sheet metal | 80,3 | * |
| Banco | 9,6 | |
| Straw | 10 | |
| Others | 0,1 | |
| **Type of walls** | | |
| Breeze blocks | 17,9 | |
| Improved banco | 33,5 | |
| Banco | 46,6 | * |
| Straw/shrubs | 1,4 | |
| Others | 0,6 | |

**Table B.3.** Correlation between household farming practices and food security status.

| Variables | Pearson chi-square value | df | Phi value | Cramer's V value | P-value |
|---|---|---|---|---|---|
| Livelihoods coping strategies | 441284,824[a] | 9 | 0,591 | 0,341 | 0,000* |
| Food expenditures shares | 563201,016[a] | 9 | 0,524 | 0,302 | 0,000* |
| Agrosylvopastoral species richness | 87503,882[a] | 12 | 0,233 | 0,135 | 0,000* |
| Livestock species richness | 75867,393[a] | 12 | 0,217 | 0,125 | 0,000* |
| Crops species richness | 75704,525[a] | 12 | 0,217 | 0,125 | 0,000* |
| Farm total agricultural area | 53475,238[a] | 12 | 0,182 | 0,105 | 0,000* |
| Possession of a non-agricultural source of income | 49630,451[a] | 6 | 0,176 | 0,124 | 0,000* |
| Agroforestry species richness | 38842,861[a] | 12 | 0,155 | 0,090 | 0,000* |
| Micro-dose practice | 4131,415[a] | 3 | 0,120 | 0,120 | 0,000* |
| Land and water conservation practices | 21731,407[a] | 24 | 0,116 | 0,067 | 0,000* |
| Type of input | 20446,106[a] | 3 | 0,113 | 0,113 | 0,000* |
| Access to agricultural extension activities | 19348,631[a] | 3 | 0,110 | 0,110 | 0,000* |
| Practice of in-line seeding technique | 16630,711[a] | 3 | 0,102 | 0,102 | 0,000* |
| Sarclo-binage or thinning practices | 9407,058[a] | 3 | 0,076 | 0,076 | 0,000* |
| Organic manure | 7389,622[a] | 3 | 0,068 | 0,068 | 0,000* |
| Crop rotation practices | 5873,873[a] | 3 | 0,060 | 0,060 | 0,000* |
| improved seed | 23341,932[a] | 3 | 0,051 | 0,051 | 0,000* |
| Home gardening cultivation | 3384,208[a] | 3 | 0,046 | 0,046 | 0,000* |
| Agricultural classes | 20807,763[a] | 21 | 0,114 | 0,066 | 0,000* |

a. 0 cells (0,0%) have expected count less than 5.

*. The p-value is significant at the 95% level.

**Table B.4.** Distribution of food insecurity prevalence according to household farming practices.

| | Proportion of food-insecure households |
|---|---|
| **Type of livestock** | |
| No livestock | 27,7% |
| Extensive sedentary livestock farming | 10,9% |
| Semi-intensive livestock farming | 11,8% |

| | |
|---|---|
| Intensive sedentary livestock farming | 12,5% |
| Transhumance | 19,9% |
| **Practice of crop rotation** | |
| Does not practice rotation | 12,0% |
| Practices rotation | 14,4% |
| **Type of association** | |
| Pure | 11,5% |
| Association | 14,5% |
| **Practice of inline seeding technique** | |
| No | 17,0% |
| Yes | 10,2% |
| **Practice of the sarclo-binage technique** | |
| No | 15,4% |
| Yes | 10,4% |
| **Type of input** | |
| Biological input | 16,2% |
| Non-biological input | 11,5% |
| **Use of organic manure** | 0,0% |
| No | 15,4% |
| Yes | 10,9% |
| **Possession of a self-help workforce** | |
| No | 15,4% |
| Yes | 9,4% |
| **Range of agroforestry trees held** | |
| Between 0-5 trees | 18,9% |
| Between 6-19 trees | 14,1% |
| Between 20 -50 trees | 9,7% |
| Between 50 and 100 | 8,3% |
| More than 100 trees | 9,3% |
| **Ranges of livestock species** | |
| 0 or 1 species | 24,7% |
| 2-3 species | 14,1% |
| 4-6 species with less than 50 animals | 10,4% |
| 4-6 species with more than 50 animals | 7,4% |
| More than 6 species | 5,3% |
| **Range of crop species** | |
| 0 or 1 specie | 33,7% |
| 2-4 species | 15,0% |
| 5-7 species | 10,3% |
| 8-10 species | 7,3% |
| more than 10 species | 7,8% |
| **Range of agrosylvopastoral species** | |
| 0 or 1 species | 24,3% |
| 2-5 species | 16,6% |
| 6-10 species | 7,4% |
| 11-15 species | 5,3% |
| more than 15 species | 5,6% |

**Table B.5.** Significance of food security level differences across livestock species ranges

| | Range of livestock species | | | | |
|---|---|---|---|---|---|
| | 0 or 1 species | 2-3 species | 4-6 species with less than 50 animals | 4-6 species with more than 50 animals | More than 6 species |
| | (A) | (B) | (C) | (D) | (E) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Food security status | In food security | | A( ,003) | A( ,000)<br>B( ,007) | A( ,000)<br>B( ,000)<br>C( ,000) | A( ,000)<br>B( ,000)<br>C( ,000) |
| | In limited food security | D( ,002)<br>E( ,004) | D( ,000)<br>E( ,000) | D( ,000)<br>E( ,000) | | |
| | Food insecure | B( ,000)<br>C( ,000)<br>D( ,000)<br>E( ,000) | C( ,001)<br>D( ,000)<br>E( ,000) | D( ,000)<br>E( ,018) | | |

Results are based on two-sided tests. For each significant pair, the key of the category with the smaller column proportion appears in the category with the larger column proportion.
Significance level for upper case letters (A, B, C): ,05

a. Tests are adjusted for all pairwise comparisons within a row of each innermost subtable using the Benjamini-Hochberg correction.

**Table B.6.** Significance of food security level differences between crop species ranges.

| | | Range of crop species | | | | |
|---|---|---|---|---|---|---|
| | | 0 or 1 species | 2-4 species | 5-7 species | 8-10 species | More than 10 species |
| | | (A) | (B) | (C) | (D) | (E) |
| Food security status | In food security | | A( ,003) | A( ,000)<br>B( ,000) | A( ,000)<br>B( ,000) | A( ,000)<br>B( ,000)<br>C( ,000)<br>D( ,000) |
| | In limited food safety | E( ,001) | C( ,005)<br>E( ,000) | E( ,001) | E( ,000) | |
| | Food insecure | B( ,000)<br>C( ,000)<br>D( ,000)<br>E( ,000) | C( ,000)<br>D( ,000)<br>E( ,000) | D( ,035)<br>E( ,038) | | |

Results are based on two-sided tests. For each significant pair, the key of the category with the smaller column proportion appears in the category with the larger column proportion.
Significance level for upper case letters (A, B, C): ,05

a. Tests are adjusted for all pairwise comparisons within a row of each innermost subtable using the Benjamini-Hochberg correction.

## Appendix C: Supplementary Partial Dependence Plots

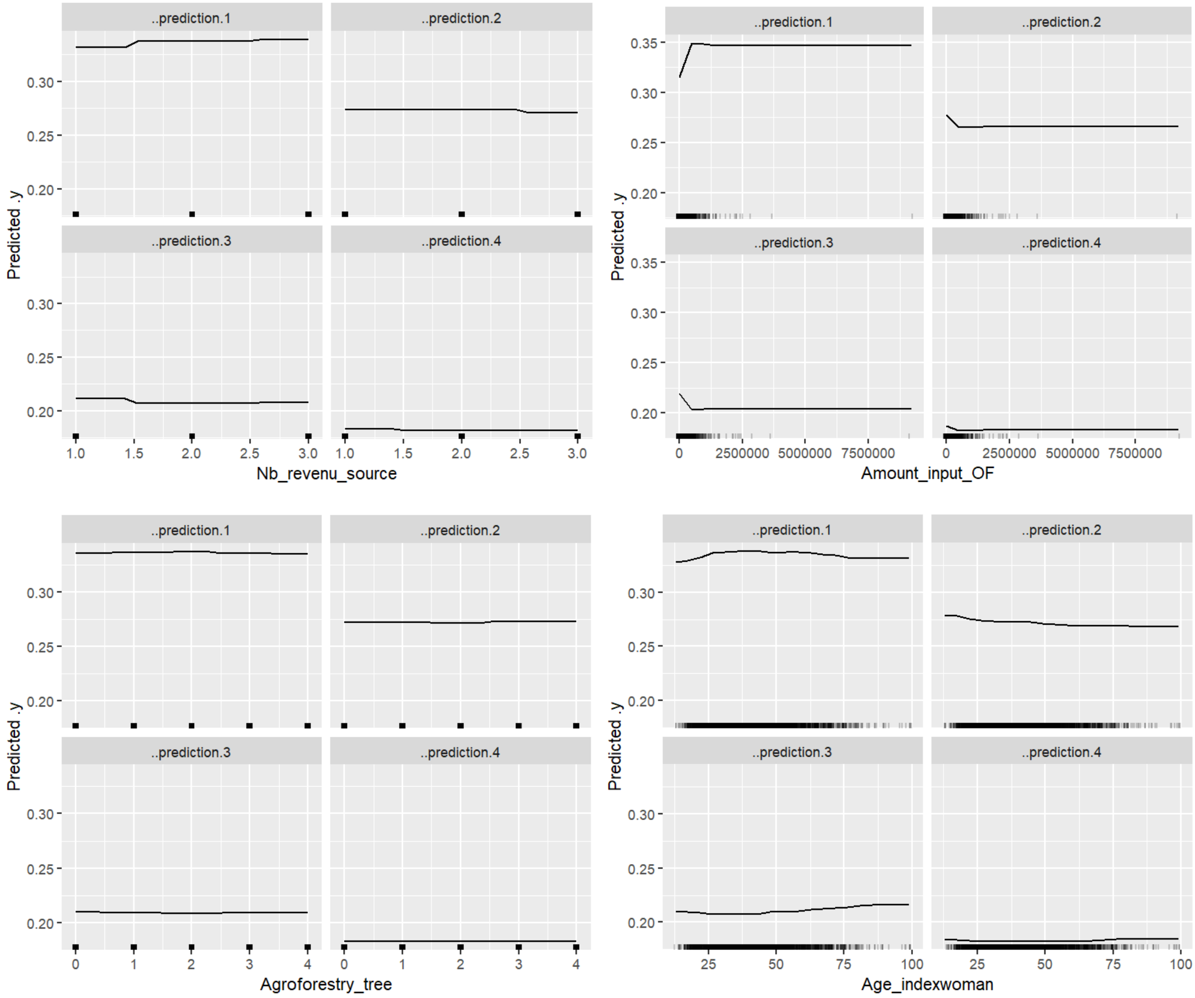


Fig. C1: Additional partial dependence plots highlighting marginal effects of the predictors on food security categories