Suivi de la sécurité alimentaire en Afrique de l'Ouest : Quelles méthodes d'analyse de données pour traiter l'interdisciplinarité de la sécurité alimentaire

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

JOURNAL OF INTERDISCIPLINARY METHODOLOGIES AND ISSUES IN SCIENCE

Suivi de la sécurité alimentaire en Afrique de l'Ouest : Quelles méthodes d'analyse de données pour traiter l'interdisciplinarité de la sécurité alimentaire

H. DELEGLISE[*1], **A. BÉGUÉ**[1], **R. INTERDONATO**[1], **E. MAÎTRE d'HÔTEL**[2], **M. TEISSEIRE**[*3]

[1]CIRAD, UMR TETIS, F-34398 Montpellier, France
[2]MoISA, Univ Montpellier, CIHEAM-IAMM, CIRAD, INRAE, Institut Agro, IRD, Montpellier, France
[3]TETIS, Univ Montpellier, AgroParisTech, CIRAD, CNRS, INRAE, Montpellier, France

*Correspondance : hugo.deleglise@cirad.fr, maguelonne.teisseire@inrae.fr

DOI : 10.18713/JIMIS-120221-8-3
Soumis le 24 Juillet 2020 - Publié le 12 Février 2021
Volume : **8** - Année : **2021**
Titre du numéro : **Agriculture numérique en Afrique**
Éditeurs : *Mathieu Roche, Pascal Bonnet, Hélène Kirchner*

Résumé

La faim en Afrique augmente à nouveau après avoir reculé pendant plusieurs années. Elle représente un enjeu sociétal mondial auquel toutes les disciplines se préoccupant d'analyse de données sont confrontées. Dans le cadre d'une collaboration pluridisciplinaire entre informaticiens, économistes et télédétecteurs, nous nous intéressons dans cet article au cas du Burkina Faso qui est l'un des pays d'Afrique de l'Ouest les plus touchés par l'insécurité alimentaire. Nous détaillons les données disponibles qui permettront d'alimenter un système de surveillance et d'alerte précoce. Nous décrivons deux approches d'analyse de données hétérogènes (valeurs quantitatives, séries temporelles, images, etc.) qui sont les méthodes d'apprentissage et les méthodes d'extraction de motifs spatio-temporels. Puis, nous dressons les premiers bilans des expérimentations réalisées, les modèles utilisées ont permis d'obtenir une accuracy de 0.75 pour le score de consommation alimentaire et de 0.73 pour le score de diversité alimentaire. Nous listons enfin les travaux que nous poursuivons dans ce contexte.

Mots-Clés

Burkina faso, Sécurité Alimentaire ; Fouille de données ; Apprentissage ; Motifs spatio-temporels

I INTRODUCTION

La faim en Afrique augmente à nouveau après avoir reculé pendant plusieurs années. En Afrique de l'Ouest, la situation alimentaire s'est améliorée à un rythme régulier entre 2000 et 2014. Au cours de cette période, la prévalence de la sous-alimentation a progressivement diminué de 12,3

% à 10,7 % avant de remonter à près de 15 % en 2017 (FAO *et al.*, 2018). Le Burkina Faso est l'un des pays d'Afrique de l'Ouest les plus touchés par l'insécurité alimentaire, avec une prévalence de la sous-alimentation de 21,3 % entre 2015 et 2017 (FAO et ECA, 2018). Le pays est fortement touché par le "triple fardeau de la malnutrition", un concept qui met en évidence la complexité de la malnutrition et ses différentes expressions : sous-alimentation, carence en micronutriments et surpoids/obésité. En 2017, les prévalences de l'émaciation chez les enfants, du retard de croissance chez les enfants et de l'obésité chez les adultes étaient respectivement de 7,6 %, 27,3 % et 4,5 %, les deux premiers indicateurs étant parmi les plus élevés en Afrique de l'Ouest (FAO *et al.*, 2018). Les raisons de la détérioration de la situation alimentaire au Burkina Faso ces dernières années sont multiples et corrélées. Une première raison est que le changement climatique a entraîné une augmentation des phénomènes météorologiques extrêmes tels que les sécheresses et les inondations qui affectent la disponibilité alimentaire (Tapsoba *et al.*, 2019). Une deuxième raison est que les conflits au Sahel déplacent les populations et provoquent la chute de la production alimentaire et des perturbations des canaux de distribution (Lacher, 2012). Ces deux phénomènes provoquent un ralentissement économique aggravé par un contexte économique mondial déjà fragile.

Pour prévenir les crises alimentaires et concevoir des interventions appropriées, plusieurs systèmes d'alerte et de surveillance de la sécurité alimentaire comme le SMIAR (Système mondial d'information et d'alerte rapide) créé par l'Organisation des Nations unies pour l'alimentation et l'agriculture (FAO), et FEWSNET (Famine Early Warning Systems Network) fondé par l'Agence américaine pour le développement international ont été mis en place depuis la seconde moitié du siècle dernier par des ONG et des organisations étatiques, et sont aujourd'hui très actifs dans les pays touchés par l'insécurité alimentaire. Ils publient régulièrement des bulletins sur la situation alimentaire à l'échelle régionale et nationale. Ces systèmes visent à prévenir et à traiter les crises alimentaires par la mise en place de programmes d'aide alimentaire ciblés et appropriés.

Étant donné la difficulté liée à la prédiction des crises alimentaires, certains éléments de ces systèmes empêchent des prévisions plus précises et précoces. Tout d'abord, pour classer le niveau de Sécurité Alimentaire (SA) d'un territoire, les différentes sources d'information sont combinées et synthétisées manuellement selon des règles préétablies. Cette intervention exclusivement humaine est longue et repose sur un nombre limité de règles de décision. En outre, ces systèmes intègrent un nombre limité de données, l'intégration de données provenant d'autres domaines liés à la SA (prix des produits de base, événements violents, etc.) et d'autres types de données (séries temporelles, images à haute résolution) devrait permettre de la décrire de manière plus complète.

Dans cet article, nous explorons d'une part les données qui pourraient être utilisées pour traiter cet enjeu sociétal, puis nous décrivons quelques outils intégrant des méthodes de fouille de données pour obtenir une prédiction de l'état de la SA. Les données à considérer sont extrêmement hétérogènes : séries temporelles (précipitations, indice de végétation par différence normalisé (NDVI)), rasters (densités de population, occupation des sols, etc.), points GPS (hôpitaux, écoles, événements violents), variables quantitatives économiques (prix du maïs, variables de la Banque mondiale), et météorologiques.

Suivant le cycle de l'extraction de connaissances tel que proposé par (Fayyad *et al.*, 1996), un des enjeux importants consiste à réaliser les prétraitements adaptés à chaque type de données, puis de trouver les méthodes adaptées et la bonne échelle spatio-temporelle permettant de les combiner. Les méthodes classiques d'apprentissage automatique (régression logistique, forêt

aléatoire, machine à vecteurs de support, etc.) ne permettent pas à elles seules de traiter l'interdisciplinarité dans les données. L'objectif de cette étude est de présenter deux méthodes d'analyse de données qui ont peu été utilisées pour la prédiction d'indicateurs de SA : les méthodes d'apprentissage profond et les méthodes d'extraction de motifs spatio-temporels.

La suite de l'article est organisée de la façon suivante. Nous détaillons les données dans la section II puis décrivons les méthodes que nous souhaitons explorer dans la section III. Nous présentons les résultats obtenus avec les méthodes basées sur l'apprentissage en section IV. Enfin, nous concluons par les différentes perspectives associées à cette étude exploratoire.

II LES TYPES DE DONNÉES

2.1 Mesurer la sécurité alimentaire

La SA est un concept complexe et multifactoriel, résultant de facteurs multiples et interdépendants (par exemple, le climat, l'économie, les guerres). La SA est assurée "lorsque tous les individus ont, à tout moment, un accès physique et économique à une nourriture suffisante, saine et nutritive" (Shaw D.J., 2007). De cette définition, quatre éléments émergent : (i) la disponibilité en quantités suffisantes de denrées alimentaires de nature et de qualité appropriées ; (ii) l'accès de toutes les personnes aux ressources essentielles pour acquérir les denrées alimentaires nécessaires à un régime équilibré ; (iii) la stabilité de l'accès à la nourriture dans le temps malgré les chocs naturels ou économiques ; (iv) l'utilisation appropriée des denrées alimentaires (stockage, cuisson, hygiène, etc.). Ces composantes peuvent être appréciées à différents niveaux, par le biais de sources de données constituées au niveau national, régional, des ménages ou des individus. Il existe un grand nombre d'indicateurs de SA (Score de consommation alimentaire, score de diversité alimentaire, indices de stratégie de survie, etc.) et l'utilisation de plusieurs indicateurs est recommandée en raison de la complexité de la SA (Coates, 2013). Hoddinott (1999) a estimé le nombre d'indicateurs de SA à environ 450. Ces indicateurs sont longs et coûteux à obtenir avec les méthodologies classiques, c'est-à-dire avec les données collectées au niveau des ménages. Il existe également un grand nombre de proxies (indicateurs indirects) liés à une ou plusieurs composantes de la SA, tels que les indices de végétation, les précipitations, les prix des denrées alimentaires, la densité de population, le nombre d'événements violents, l'état des routes, le nombre d'écoles et d'hôpitaux, etc. Cet article a pour but de proposer des modèles d'apprentissage automatique capables d'utiliser des facteurs influençant la SA comme proxies (variables explicatives) pour prédire des indicateurs de SA (variables réponses).

2.2 Variables réponses

Les variables réponses, ou indicateurs de SA, sont issues de l'enquête permanente agricole, qui est menée chaque année en routine par le Ministère de l'agriculture depuis 1982 au Burkina Faso (Permanent Agricultural Survey, 2015). Le nombre de ménages a été choisi pour être représentatif dans chaque province (une unité administrative supérieure aux communes choisies comme échelle spatiale de référence), certaines communes sont donc possiblement légèrement sous-échantillonnées. L'enquête a été réalisée par échantillonnage stratifié à deux degrés (villages et ménages) renouvelé tous les 5ans. La base de sondage du premier degré est obtenue à partir du module agricole du recensement général de la population de 2006. Cette base a permis de disposer d'une liste de villages (7 871 villages et secteurs) contenant 1 219 241 ménages agricoles. En 2008, 1 424 909 ménages au total étaient agricoles, soit 81.5% des ménages (Bureau central du recensement général de l'agriculture, 2011), en 2019 80% des emplois étaient

encore liés à l'agriculture (World Bank, 2020). La base de sondage des ménages agricoles est créée dans chaque village échantillonné (sélectionnés avec une probabilité proportionnelle à leurs nombres de ménages agricoles) à partir d'une liste de ménages établie chaque année en dénombrant tous les ménages agricoles du village. Pour cette étude, nous prenons en compte les données qui sont disponibles de 2009 à 2018. L'ensemble de données qui en résulte contient les informations de 46400 ménages agricoles, soit une moyenne de 4640 ménages agricoles par an répartis dans 342 parmi les 351 communes représentées figure 1. Un ménage agricole est défini comme un ménage pratiquant l'une des activités suivantes : cultures temporaires (cultures pluviales et de contre-saison), fruticulture, élevage d'animaux. Dans ce travail, nous nous concentrons sur deux indicateurs calculés à partir des réponses aux enquêtes ménages : le score de consommation alimentaire (SCA) et le score de diversité alimentaire des ménages (SDA). Ils fournissent des informations sur la fréquence, la quantité et la qualité des aliments, et figurent parmi les indicateurs les plus populaires auprès des scientifiques et des organisations (Jones *et al.*, 2013; Maxwell *et al.*, 2014; Vhurumuku, 2014). Ces indicateurs sont agrégés par commune et considérés de 2009 à 2018, ce qui représente 3066 observations.

FIGURE 1 – Distribution spatiale des 351 communes du Burkina Faso (fond de carte : Google Maps).

Score de consommation alimentaire (SCA) : Cet indicateur est une mesure de la quantité de nutriments et de l'apport énergétique. Il s'agit d'une estimation de la fréquence cumulée de 9 groupes d'aliments consommés pendant 7 jours au sein de chaque ménage enquêté. La fréquence de consommation de chaque groupe d'aliments est pondérée par sa valeur nutritionnelle (Equation 1 ; Tableau 1). Plusieurs seuils permettant de différencier les ménages sont couramment utilisés. Nous choisissons les seuils fixés par le World Food Programme (WFP) :

acceptable (>42), limite (28-42), et faible (<28) (Wiesmann *et al.*, 2009)

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

$x_i \in \{$Fréquence de consommation de chaque groupe d'aliments $i\}$, $p_i \in \{$Poids des groupes d'aliments$\}$

Groupe d'aliments	**Poids**
Céréales et tubercules	2
Légumineuses	3
Légumes et feuilles	1
Fruits	1
Protéines animales	4
Produits laitiers	4
Sucreries	0.5
Huiles	0.5
Condiments	0

TABLE 1 – Groupes d'aliments et leur poids pour le calcul du Score de consommation alimentaire (SCA). *Source : (Wiesmann et al., 2009)*

Score de diversité alimentaire des ménages (SDA) : C'est un indicateur de la fréquence et de la diversité de la consommation alimentaire davantage axé sur la qualité nutritionnelle du régime alimentaire. Il s'agit d'une estimation du nombre de groupes d'aliments consommés au cours des dernières 24 heures. Il n'y a pas de consensus sur le nombre de groupes à utiliser et leurs limites. Par exemple, le WFP utilise les mêmes groupes que ceux utilisés pour les SCA, tandis que la FAO utilise une classification de 12 groupes alimentaires (Kennedy *et al.*, 2013). Le choix de la classification des aliments dépend du contexte (mettre davantage l'accent sur les produits riches en vitamines A, calories, etc.) et des données disponibles. Nous utilisons la méthodologie de la FAO pour calculer le SDA (Equation 2 ; Tableau 2).

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

$x_i \in \{0$: Aliment i non consommé, 1 : Aliment i consommé$\}$

Groupe d'aliments
Céréales
Racines et tubercules
Légumes
Fruits
Viandes
Oeufs
Poissons et fruits de mer
Légumineuses, noix et graines
Produits laitiers
Huiles et matières grasses
Sucreries
Condiments, épices et boissons

TABLE 2 – Groupes d'aliments pour le calcul du score de diversité alimentaire des ménages (SDA). *Source : (Kennedy et al., 2013)*

2.3 Variables explicatives

L'aspect multifactoriel de la SA implique l'utilisation de données hétérogènes pour en obtenir l'image la plus complète possible. Les variables proxies de la SA utilisées comme variables explicatives sont hétérogènes à trois niveaux : 1) Au niveau thématique : elles sont relatives à des domaines tels que la télédétection, la météorologie, l'économie, la démographie ou l'occupation du sol. Cela implique d'avoir une vision complète des facteurs de SA dans le lieu étudié ; 2) Au niveau de la structure des données, celles-ci sont de types variés : valeurs quantitatives, points GPS, séries temporelles, images. Cela nécessite l'utilisation d'outils et de méthodes adaptés au traitement de chaque donnée ; 3) Au niveau de l'échelle spatio-temporelle : les données peuvent être disponibles spatialement par région, commune, station ou pixel et temporellement par décennie, année, mois ou semaine. Cela implique de choisir l'échelle spatio-temporelle la plus appropriée. Tout d'abord, les proxies de la SA sont prétraités pour extraire des variables explicatives pertinentes à l'échelle de la commune, qui est la plus petite unité administrative pour laquelle les variables réponses sont spatialisées, cela permet de maximiser le nombre d'observations pour l'apprentissage des modèles. Certaines variables proxies de la SA ont une granularité plus fine et doivent être agrégées par commune par somme, moyenne ou par des agrégations plus complexe (coefficient de Gini, auto-corrélation, entropie différentielle) ; d'autres sont disponibles à une granularité plus grossière et doivent être interpolées sur chaque commune, nous avons choisi d'utiliser une interpolation par k plus proches voisins qui est une technique précise et éprouvée. Ensuite, pour chaque commune et chaque année, les variables explicatives obtenues sont sélectionnées en ne retenant que les variables explicatives significativement corrélées avec la variable réponse considérée (p-value inférieure à 0,05). Par exemple, nous avions au départ accès aux prix du maïs, du mil et du sorgho, les prix du mil et du sorgho n'amélioraient pas les modèles, ces deux variables n'ont donc pas été conservées. Enfin, chaque variable explicative est centrée réduite par rapport aux communes et aux années (consiste à soustraire la moyenne et à la diviser par l'écart-type). Les informations sur chaque ensemble de données sont disponibles dans le tableau 3. Les variables explicatives sélectionnées sont classées en 4 groupes selon leur granularité spatio-temporelle pour être traitées indépendamment par une méthode d'apprentissage automatique pour la prédiction de la variable réponse :

— **Séries temporelles** qui ont plusieurs valeurs par an et une valeur par commune. Elles sont agrégées en séries temporelles mensuelles (de mai à novembre de l'année au cours de laquelle l'indicateur de SA est collecté et de l'année précédente)
— **Données conjoncturelles** qui ont une valeur par an et une valeur par commune.
— **Données structurelles** qui ont une valeur par commune et sont invariantes par an.
— **Données à haute résolution spatiale** qui ont plusieurs valeurs par commune. Les valeurs sont des patchs de 10x10 pixels de 100m extraits de chaque source de données.

Variable	Résolution	Fréquence	Source	Mise à l'échelle
Séries temporelles [plusieurs valeurs par an ; une valeur par commune]				
Smoothed Brightness Temperature (SMT)	4km	7 jours	National Oceanic and Atmospheric Administration (Noaa)	Maximum
Précipitations	6km	10 jours	Mission de mesure des précipitations tropicales (Trmm)	Somme
Températures minimales et maximales moyennes	21km	1 mois	WorldClim	Moyenne
Prix du maïs	64 marchés	1 mois	Société Nationale de Gestion du Stock de Sécurité alimentaire (SONAGESS)	Interpolation K plus proches voisins
Données conjoncturelles [une valeur par an ; une valeur par commune]				
Données météorologiques	10 stations	1an	Plate-forme Knoema	Interpolation K plus proches voisins
Densité de population	100m	1an	Afripop	Autocorrélation spatiale à 2km et 5km, Entropie, Gini
Données économiques	Pays	1an	Banque Mondiale	Valeur du pays
Indice de végétation par différence normalisé	Commune	1an	MODIS	Moyenne
Événements violents	vecteurs de points	2018	Localisation des conflits armés & Event Data Project (ACLED)	Somme
Données structurelles [une valeur par commune]				
Hôpitaux, écoles	vecteurs de points	2018	Open Street Map	Somme
Cours d'eau	vecteurs de lignes	2008	Digital Chart of the World	Nombre, longueur
Altitude	1km	2018	Consultative Group on International Agricultural Research	Maximum, variance
Qualité des sols	1km	2008	Food and Agriculture Organization	Somme
Données à haute résolution spatiale [plusieurs valeurs par commune]				
Densité de population	100m	1an	Afripop	CNN
Occupation du sol (cultures, forêts, zones construites)	100m	2016	Agence spatiale européenne	CNN

TABLE 3 – Récapitulatif de l'ensemble des données

III LES MÉTHODES

Nous souhaitons explorer deux types d'approches. Tout d'abord les méthodes dîtes d'apprentissage qui nécessitent une partie des données labellisées et qui construisent un modèle à partir duquel il sera ensuite possible de prédire les nouvelles données à analyser. Puis, nous nous intéresserons aux approches basées sur les motifs fréquents, qui explorent sans a priori les données et construisent des patrons qui sont utilisés dans un deuxième temps comme des marqueurs de prédiction pour les nouvelles données si celles-ci vérifient le motif concerné. L'extraction de motifs a fait l'objet de nombreux travaux dans le domaine de la science des données

mais peu concernent la prédiction à partir de données fortement hétérogènes (hétérogénéité thématique, structurelle et spatio-temporelle détaillée dans la section 2.3).

3.1 Les méthodes basées sur l'apprentissage

L'apprentissage automatique désigne un ensemble de méthodes statistiques utilisées pour analyser un ensemble de données afin d'en déduire des règles qui constituent de nouvelles connaissances pour l'analyse de nouvelles situations (figure 2). Les méthodes d'apprentissage automatique sont de plus en plus utilisées pour extraire des informations pertinentes à partir de données complexes et hétérogènes liées à la SA, et plusieurs études ont tenté de détecter l'insécurité alimentaire et les crises en utilisant des techniques d'apprentissage automatique (Okori et Obua, 2011; Barbosa et Nelson, 2016; Lukyamuzi *et al.*, 2018) avec des résultats encourageants. Un groupe de méthodes d'apprentissage automatique appelé apprentissage profond est de plus en plus utilisé et est très efficace pour analyser des données complexes et hétérogènes (Valdés, 2018). L'apprentissage profond a été utilisé avec des résultats concluants pour l'analyse de sujets liés à la SA comme la pauvreté (Shailesh *et al.*, 2018), la sécheresse (Mumtaz *et al.*, 2018) ou les prix du marché (Min *et al.*, 2019).

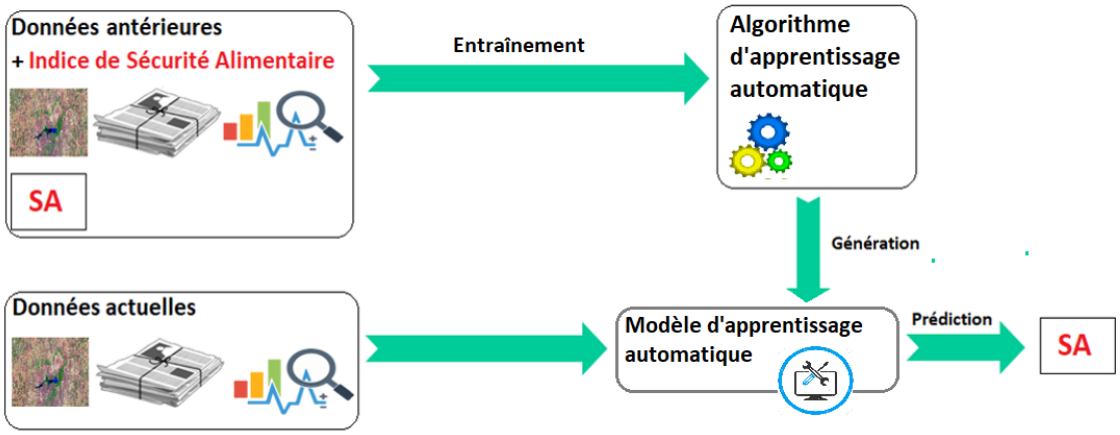

FIGURE 2 – Architecture d'un modèle d'apprentissage automatique

Nous avons choisi d'utiliser les méthodes d'apprentissage automatique suivantes :

— Une Forêt Aléatoire (FA) est une méthode d'apprentissage automatique qui fonctionne en générant une multitude d'arbres de décision (qui sont une méthode intuitive d'apprentissage automatique) chacun entraîné sur un sous-jeu de données et de variables explicatives. Les variables réponses sont agrégées par vote.
— Un réseau de neurones convolutionnel (CNN) est une méthode d'apprentissage profond utilisée dans la reconnaissance et le traitement des images et spécialement conçue pour l'analyse des pixels et la compréhension du contexte spatial.
— Un réseau de neurones récurrents à mémoire court-terme et long terme (LSTM) est une méthode d'apprentissage profond adaptée aux traitements de séries temporelles. Cette méthode conserve des informations en mémoire et peut prendre en compte à un instant t un certain nombre d'états passés.

Ces méthodes d'apprentissage automatique et profond sont intégrées dans des modèles qui ont pour but de combiner de manière appropriée chaque type de données (séries temporelles,

données variables par année, par commune, images satellites), afin de prédire l'indicateur de SA voulu en prenant le plus de facteurs possibles en compte (figure 3).

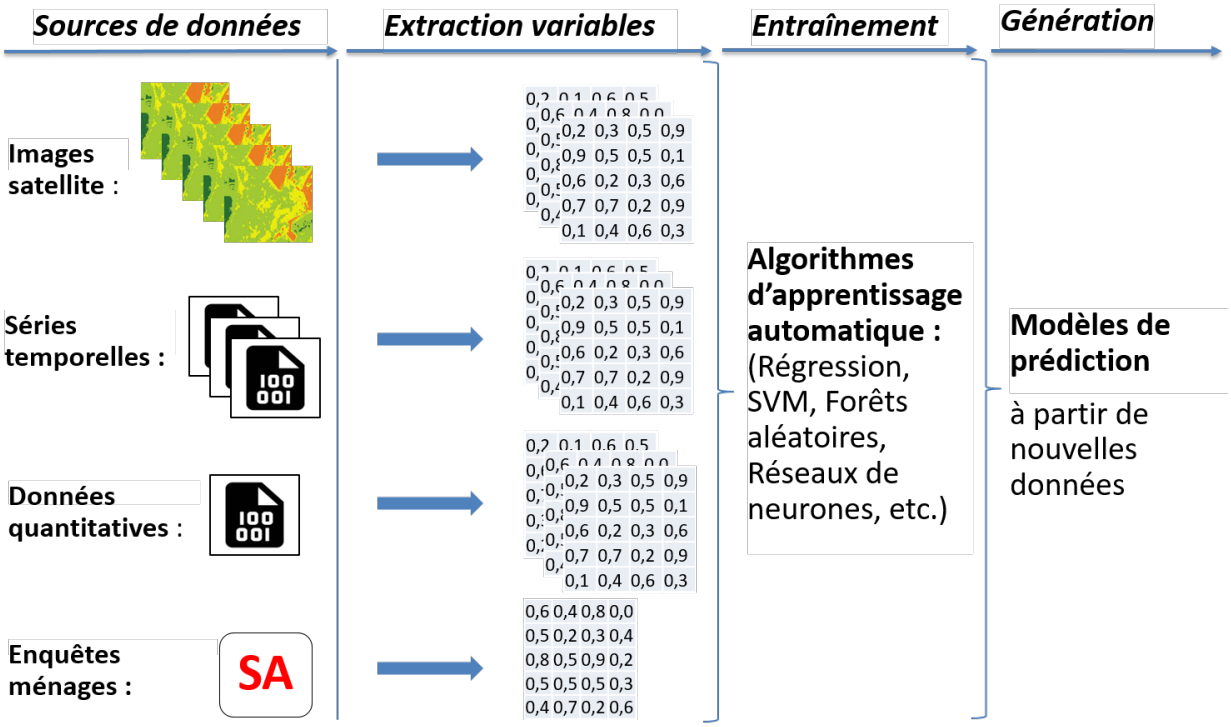

FIGURE 3 – Architecture type d'un modèle d'apprentissage automatique adapté au traitement de données hétérogènes.

Nous réalisons 4 types de modèles de classification qui prennent en entrée et combinent les variables explicatives présentées pour prédire le SCA et le SDA (figure 4). Nous réalisons des classifications binaires et en 4 classes, en utilisant la méthode du coude pour déterminer le nombre optimal de classes. Nous sélectionnons aléatoirement 85% de l'ensemble des données pour l'apprentissage des modèles et 15% pour le test, en répétant cette procédure 5 fois et en calculant les performances moyennes pour obtenir des résultats plus stables. Pour évaluer les performances de classification, nous utilisons l'accuracy qui est le rapport du nombre d'observations bien classées par le nombre total d'observations.

— **(a)** Nous appliquons une FA directement sur les variables initiales présentées en Section 2.3 : sur les séries temporelles uniquement, sur les variables Conjoncturelles & Structurelles (CS) uniquement et sur ces deux types de données conjointement.
— **(b)** Nous utilisons des modèles d'apprentissage profond adaptés au traitement de groupes de variables ayant une structure complexe : nous appliquons un LSTM adapté aux séries temporelles et un CNN adapté aux données à Haute Résolution Spatiale (HRS).
— **(c)** Nous agrégeons par vote les prédictions obtenues par 3 modèles : la FA sur les variables CS, le LSTM sur les séries temporelles et le CNN sur les données HRS.
— **(d)** Nous appliquons une FA sur les features extraits par les modèles d'apprentissage profond (64 features pour le LSTM et 128 features pour le CNN). Nous appliquons une FA sur les features du LSTM uniquement, sur les features du CNN uniquement, conjointement sur les features du CNN et les variables CS et conjointement sur les features du CNN et du LSTM et les variables CS.

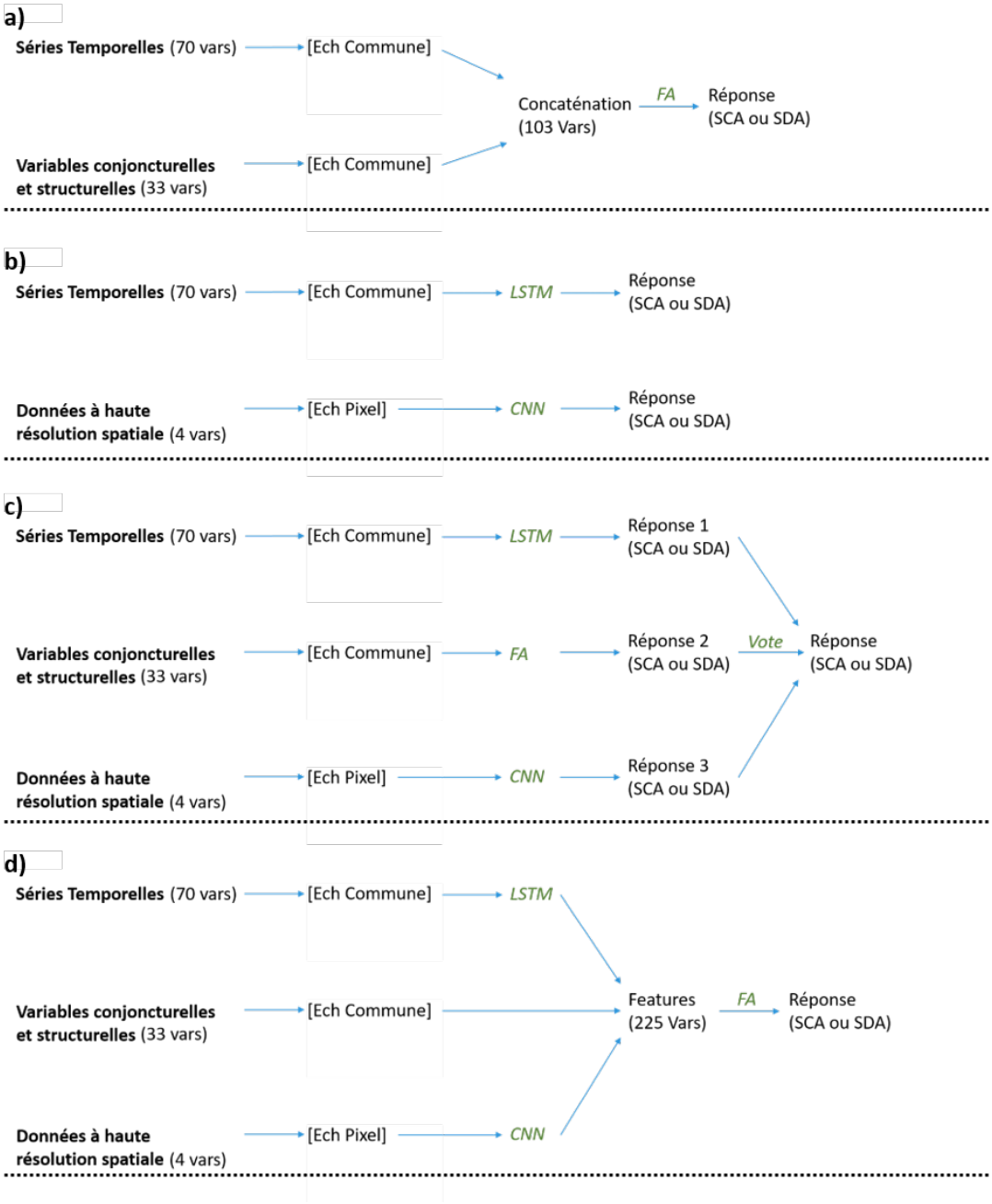

FIGURE 4 – Architecture des quatre types de modèles (a), (b), (c) et (d) utilisés.

3.2 Les motifs spatio-temporels

Adoptant une démarche différente à celle décrite précédemment qui est basée sur l'apprentissage, nous souhaitons mettre en place des algorithmes de recherche de motifs ou patrons. Ces motifs permettront d'expliquer les situations à risque et éventuellement, dans un second temps, de produire un modèle de prédiction de telles situations. La découverte de motifs repose sur l'identification des informations récurrentes dans les données correspondant à des comportements fréquents ou d'intérêt. Les méthodes à adopter doivent permettre de prendre en compte la particularité des données hétérogènes auxquelles nous sommes confrontés, en particulier selon les axes temporel et spatial. Il existe de nombreuses contributions dans la communauté fouille de données et nous en listons ci-dessous quelques-unes.

Le problème de la recherche de motifs séquentiels a été initiée par (Agrawal et Srikant, 1995) dans le contexte du panier de la ménagère. Les algorithmes développés ont été appliqués avec succès dans de nombreux domaines(Gupta et Han, 2013) comme la biologie (Wang *et al.*, 2004; Salle *et al.*, 2009), la fouille d'usage du Web (Srivastava *et al.*, 2000), la fouille de flux de données (Mendes *et al.*, 2008) ou la description des comportements au sein d'un groupe (Perera *et al.*, 2009), mais leur utilisation peut également être adaptée à des données moins classiques. Les motifs séquentiels ont été ainsi adoptés pour résumer les évolutions temporelles dans des séries d'images satellites (Julea *et al.*, 2010; Wu et Zhang, 2019) ou encore pour la classification de données textuelles (Yuan *et al.*, 2018).

Ce qui nous intéresse dans notre contexte, c'est la capacité à prendre en compte à la fois l'aspect spatial et temporel des données à analyser. Concernant la prise en compte de donnée spatio-temporelles, de nombreuses propositions ont été réalisées. Dans (Wang *et al.*, 2005), les données sont représentées sous la forme d'un ensemble de grilles dans lesquelles apparaissent des évènements (des items), chaque grille représentant l'état de la grille spatiale à un instant t. Pour chaque date et chaque position absolue, un itemset, ensemble d'évènements (d'items) est généré. Pour chaque position, une séquence d'itemsets est alors construite en considérant toutes les dates. Les motifs séquentiels sont alors extraits à partir de cet ensemble de séquences, en utilisant une position absolue comme point de référence. Un exemple de motif obtenu est $\langle$(Pluie(0,0))(Humidité(0,1))$\rangle$, qui signifie que l'on trouve fréquemment de la pluie aux coordonnées 0,0 et ensuite de l'humidité aux coordonnées 0,1. Ce type de motif a le désavantage d'être sensible au choix du point de référence et l'espace est limité à une représentation sous forme de grille.

Dans (Huang *et al.*, 2008), les auteurs proposent la notion d'événements proches dans le temps et l'espace. Une fenêtre temporelle et spatiale est définie par un intervalle de temps et par un intervalle de distance. Les motifs sont sous forme de règles d'associations telles que la règle $\langle$Pluie$\Rightarrow$ Humidité$\rangle$ qui signifie que dans des zones proches et à des dates proches, on trouve de la pluie suivie par de l'humidité. Ces motifs n'expriment pas le fait que les objets spatiaux soient liés par une relation, ni la présence de plusieurs échelles géographiques.

Une extraction de motifs spatio-temporels en utilisant une relation de voisinage ou de proximité entre séquences est proposée dans (Salas *et al.*, 2012). Les motifs proposés sont décrits selon la forme $\langle$(Humidité .[Pluie Vent])(Humidité Pluie)$\rangle$. La relation de voisinage est mise en évidence avec l'opérateur de voisinage **.** et l'opérateur de groupement **[]**. Prenons l'exemple d'une ville où nous trouvons ce motif. Cela signifie qu'il y a eu de l'humidité à une certaine date et au même moment de la pluie et du vent dans une ville proche (par une distance euclidienne ou définie par l'utilisateur). Plus tard, il y a eu de l'humidité et de la pluie dans la ville. Cette relation spatiale reste simple et il n'est pas possible de la spécialiser (i.e. une relation représentée par un domaine de valeurs), ni d'avoir plusieurs granularités (i.e. plusieurs échelles géographiques). Elle se limite à un seul type de relation : la proximité dans l'espace.

Une gestion de la granularité sur l'espace est proposée dans (Tsoukatos et Gunopulos, 2001). Comme dans (Wang *et al.*, 2005), la spatialité est représentée par une grille d'événements et la temporalité par un ensemble de grilles spatiales. L'utilisateur choisit un niveau de granularité qui va fusionner un ensemble de cases de la grille. Plus la valeur de granularité est élevée, plus les cases seront fusionnées, ce qui permet la généralisation des données d'un point de vue spatial. Pour extraire les motifs, il est donc nécessaire de donner une valeur de granularité et la

spatialité se limite à un seul niveau. De plus l'extraction revient à extraire des motifs séquentiels classiques du type $\langle$(Soleil)(Vent)(Soleil,Humidité=Faible)$\rangle$ qui se lit de la manière suivante : fréquemment l'événement Soleil est suivi de l'événement Vent, suivi des événements Soleil et Humidité=Faible pour un niveau de granularité précis.

Manipuler des objets géographiques liés entre eux et à différentes échelles reste une problématique complexe. Nous souhaitons adopter une méthode basée sur les motifs spatio-temporels avec différents types de relations spatiales telle que celle décrite dans (Fabrègue *et al.*, 2012). Ceci nous permettra d'avoir une gestion plus fine de la spatialité avec les notions de relations spatiales à différentes granularités. Le résultat attendu est l'obtention de motifs plus riches sémantiquement. Le problème associé à l'extraction de tels motifs est l'explosion de l'espace de recherche. Il est donc nécessaire de mettre en place des stratégies d'optimisation indispensables dans notre contexte. Cette approche a été appliquée sur des données hydrologiques (qualité des cours d'eau dans le bassin de la Saône). Les motifs ont été obtenus avec une navigation dans les hiérarchies spatiales. Ceci permet l'extraction de motifs plus spécifiques et plus expressifs qui n'auraient jamais pu être découverts avec une méthode classique. Par exemple, le motif [U2]$\langle(.Orient[ibgn\_11\text{-}15])(var\_taxo\_31\text{-}40)\rangle$ signifie que dans le bassin hydrographique U2, il y a fréquemment une valeur IBGN (Indice biologique global normalisé) comprise entre 11 et 15 dans une station voisine (i.e. en amont **ou** en aval) et plus tard dans le temps, une variété taxonomique comprise entre 31 et 40. Ce motif ne peut pas être obtenu avec les motifs séquentiels classiques ni avec la méthode des motifs spatio-temporels de (Salas *et al.*, 2012). Il est souvent difficile pour les experts de déterminer la meilleure échelle, c'est-à-dire celle qui permet d'obtenir les meilleures observations. Une démarche automatique de détection du motif selon le niveau de granularité spatiale la plus pertinente est réellement une plus value dans une démarche d'extraction de connaissances.

L'adoption de motifs spatio-temporels comme éléments clefs d'analyse des données hétérogènes dans le contexte de la sécurité alimentaire devrait ainsi nous permettre de : 1) considérer les dimensions temporelles et spatiales ; 2) capturer la sémantique et la structure des relations entre objets géographiques ; 3) et enfin extraire des motifs en parcourant les différentes granularités spatiales et temporelles possibles. La tâche à réaliser ensuite sera d'utiliser les motifs obtenus afin de mettre en œuvre un modèle de prédiction. Toute nouvelle séquence d'information (et/ou de données) sera considérée comme signal potentiel du risque dès lors qu'elle vérifiera les motifs identifiés comme représentatifs de l'insécurité alimentaire.

IV RÉSULTATS DES MÉTHODES D'APPRENTISSAGE

Actuellement, très peu d'études ont été menées pour prédire le SCA et le SDA avec des méthodes d'apprentissage automatique. Une équipe du PAM a travaillé dans plusieurs pays sur la prédiction par régression du SCA en utilisant des méthodes d'apprentissage automatique et profond. Au Burkina Faso, des résultats modestes ont été obtenus (R^2=0,34), cette équipe a également prédit le SCA dans d'autres pays en obtenant des résultats comparables (Sierra Leone : R^2=0,24 ; Sénégal : R^2=0,52 ; Kinshasa : R^2=0,18) [1]. Ces faibles résultats confirment que la prédiction de ces indicateurs de SA est une problématique complexe en raison de leur nature multifactorielle. Aucune étude n'a à ce jour fait ce type de traitement en utilisant des classifications.
Dans notre étude, les performances sont également modestes. Pour une classification en 2

1. https ://wfp-vam.github.io/HRM/

classes, les performances (accuracy) maximales sont obtenues à partir des modèles de type (d) et atteignent respectivement 0.75 et 0.73 pour le SCA et le SDA. Mais nous constatons un faible apport de performances issues des modèles d'apprentissage profond. Les facultés de ces modèles à reconnaître des patterns complexes s'expriment davantage pour la prédiction d'un plus grand nombre de classes, c'est pourquoi nous nous concentrons sur de la classification en 4 classes dans la suite de cette section. La prédiction de variables divisées en 4 classes plutôt que 2 est plus complexe et implique des performances globalement plus faibles, cela est donc moins pertinent au niveau opérationnel, mais permet de mieux distinguer les apports de chaque modèle, ce qui est plus intéressant aux niveaux conceptuel et technique. Nous observons que les modèles de type (a) utilisant une FA qui est une méthode classique d'apprentissage automatique donnent des performances déjà significatives et proches des modèles plus sophistiqués : en intégrant uniquement les variables CS nous obtenons pour le SCA et le SDA une accuracy de 0,467 et 0,446 respectivement. Cela valide la sélection des sources de données et les prétraitements appliqués aux données utilisées. Les modèles de type (b) sont conçus pour traiter des données présentant des structures complexes (séries temporelles et données HRS) à l'aide d'une méthode d'apprentissage profond appropriée. Le LSTM parvient à mettre en évidence l'existence dans les séries temporelles d'un aspect séquentiel en lien avec les indicateurs de SA étudiés, en offrant des performances légèrement plus élevées qu'en y appliquant seulement une FA : accuracy de 0.403 contre 0.387 pour le SCA et accuracy de 0.402 contre 0.357 pour le SDA. Le CNN sur les données HRS donne des performances intéressantes (accuracy de 0.473 pour le SCA et de 0.457 pour le SDA), ces performances sont améliorées en mettant les features extraites du CNN en entrée d'une FA (accuracy de 0.498 pour le SCA et de 0.478 pour le SDA). Les modèles de type (c) et (d) sont deux stratégies permettant de combiner différents types de données. Le modèle de type (c) qui consiste à agréger les réponses des autres modèles par vote permet d'obtenir des performances plus élevées qu'en considérant chaque modèle individuellement (accuracy de 0.48 pour le SCA et de 0.468 pour le SDA). Cela démontre que chaque modèle et type de données fournissent des informations complémentaire sur la SA. Les modèles de type (d) consistent à agréger les features extraits des autres modèles avec une FA. Les meilleures performances sont obtenues de cette manière, mais les features du CNN suffisent à maximiser les performances, qui ne sont pas significativement améliorées par l'ajout des features du LSTM et des variables CS.

Modèle	**SCA**	**SDA**
(a) FA(séries temporelles)	0.387	0.357
(a) FA(variables CS)	0.467	0.446
(a) FA(séries temporelles + variables CS)	0.407	0.383
(b) LSTM(séries temporelles)	0.403	0.402
(b) CNN(données HRS)	0.473	0.457
(c) Vote(FA, LSTM et CNN réponses)	0.48	0.468
(d) FA(features LSTM)	0.377	0.348
(d) FA(features CNN)	**0.498**	**0.479**
(d) FA(features CNN + variables CS)	0.497	0.478
(d) FA(features CNN et LSTM + variables CS)	0.493	0.475

TABLE 4 – Performances (accuracy) des 4 types de modèles - (a) :jaune ; (b) :vert ; (c) :bleu ; (d) :rouge - pour la prédiction du score de consommation alimentaire et du score de diversité alimentaire des ménages.

En raison de leur effet ”boite noire”, il est compliqué de décrire précisément les schémas complexes extraits par les modèles d'apprentissage profond utilisés. En nous basant sur les performances significatives du LSTM et du CNN, nous pouvons constater d'une part que les séries temporelles de températures, de précipitations et de prix des denrées mises en entrée du LSTM permettent d'obtenir des schémas temporels porteurs d'information sur la SA, et d'autre part que les données à haute précision spatiale que sont l'utilisation des sols et les dynamiques de population utilisées en entrée du CNN permettent d'obtenir des schémas spatiaux riches en informations sur la SA. Bien que ces schémas temporels et spatiaux complexes en lien avec la SA soient pour l'instant impossible à décrire précisément et d'un apport prédictif modéré, le fait même de mettre en évidence leur existence constitue un début de réponse encourageant à une question de recherche encore inexplorée. Pour les variables CS qui sont directement traitées par FA, la significativité des variables peut être approchée par l'importance de la fonction de permutation. L'importance de la fonction de permutation est définie comme étant la diminution du score d'un modèle lorsque les valeurs d'une variable sont mélangées de manière aléatoire dans l'ensemble de test. Les 10 premiers rangs des variables CS selon leur importance de fonction de permutation pour le SCA et le SDA sont présentés dans le tableau 5. Nous notons que des variables provenant de domaines multiples sont incluses dans ces deux top 10 : structure paysagère (3), dynamique des populations (3), qualité du sol (2), végétation (1), météorologie (1), sanitaire (1), insécurité (1) ou économique (1), ce qui confirme l'importance de l'utilisation combinée de sources de données provenant de multiples domaines. 9 variables sont incluses dans les top 10 du SCA et du SDA conjointement, elles semblent dans notre cas essentielles pour la prédiction de la SA. Parmi ces variables, nous trouvons 3 variables relatives aux structures paysagères : la longueur totale des cours d'eau, le maximum et la variance de l'altitude, nous trouvons également 3 variables exprimant les dynamiques de la population : les autocorrélations spatiales à 2km et 5km et l'entropie différentielle associée aux populations, ce qui confirme, avec les bonnes performances du CNN qui prend en compte les données d'utilisation du sol et de populations, l'importance de ces sources de données pour appréhender la SA du pays. Puis nous trouvons une variable de qualité des sols : la disponibilité de l'oxygène pour les racines qui est directement liée à la disponibilité des denrées issues de l'agriculture. Nous trouvons ensuite une variable d'insécurité : le nombre d'événements violents pour 1000 habitants qui est liée à la stabilité politique et économique. La dernière variable est une variable sanitaire : le nombre d'hôpitaux pour 1000 habitants qui exprime la capacité du pays à garder sa population en bonne santé. Nous constatons que certaines variables sont davantage spécifiques d'un indicateur de SA. Le top 10 du SCA contient le NDVI de l'année précédent l'enquête qui est absent du top 10 du SDA, cette variable semble être plus spécifique de la quantité de nutriments consommés. Il est intéressant de noter que cette variable est plus importante que le NDVI de l'année en cours. Inversement, la température maximale moyenne est présente dans le top 10 du SDA et absente du top 10 du SCA, cette variable semble donc être plus spécifiquement liée à la qualité et à la diversité de l'alimentation.

Rang	SCA	SDA
1	Altitude maximale	Altitude maximale
2	Qualité des sols (disponibilité de l'oxygène)	Entropie de la population
3	NDVI moyen de l'année précédente	Auto-corrélation de la population à 5km
4	Longueur totale des cours d'eau	Longueur totale des cours d'eau
5	Variance de l'altitude	Variance de l'altitude
6	Auto-corrélation de la population à 5km	Auto-corrélation de la population à 2km
7	Auto-corrélation de la population à 2km	Nombre d'hôpitaux
8	Entropie de la population	Qualité des sols (disponibilité de l'oxygène)
9	Nombre d'événements violents	Nombre d'événements violents
10	Nombre d'hôpitaux	Température maximum moyenne

TABLE 5 – Dix premiers rangs des variables selon leur importance de fonction de permutation pour le SCA et le SDA.

V DISCUSSION ET CONCLUSION

Nous avons détaillé les données que nous avons pu recueillir dans le cadre de collaborations avec des experts de la sécurité alimentaire du Burkina Faso. Nous avons présenté deux types de méthodes qui pourraient répondre en partie à la problématique d'un système d'alerte précoce dans le cadre de la surveillance de la sécurité alimentaire. Les premières expérimentations que nous avons réalisées concernent les méthodes d'apprentissage. Nous avons pu constater que les performances (accuracy) ne sont pas élevées, ne dépassant pas 0.75 et 0.73 d'accuracy pour la classification en 2 classes du SCA et du SDA respectivement et 0.498 et 0.479 d'accuracy pour de la classification en 4 classes du SCA et du SDA. Ces résultats indiquent que la prédiction de ces indicateurs de SA est une question complexe. Les travaux auxquels il serait intéressant de nous comparer sont pour l'instant rares, et même inexistants pour de la classification. Cette étude pose donc une première pierre à l'édifice de cette problématique. Nous avons également pu constater l'apport pour l'instant modeste, mais significatif des modèles d'apprentissage profond pour le traitement des données à structure complexe (séries temporelles et données HRS), l'utilisation de ces données dans le contexte de la SA constitue donc une piste de réflexion pertinente pour de futurs travaux. Enfin, nous avons observé que les variables pointées par les modèles comme étant les plus importantes pour la prédiction des indicateurs de SA sont issues de domaines multiples (structures paysagères, dynamiques des populations, qualité des sols, qualité de la végétation, météorologie, etc.), ce qui confirme la nécessité de mettre en lien la SA avec un grand spectre de domaines liés pour avoir une image la plus complète possible de ce concept complexe et multifactoriel.

En terme de perspectives, nous attendons également des collaborations étroites avec d'autres disciplines nous permettant d'obtenir un protocole d'évaluation qualitative. Les approches de fouille de données ne trouvent leur raison d'être que dans un partenariat étroit avec les différents acteurs, que cela soit au niveau de la production des données, de leur prétraitement, mais également au niveau de la validation des connaissances qui ont pu être extraites. Un chantier d'une telle envergure ne peut trouver sa pleine réalisation que dans un étroit partenariat disciplinaire, où scientifiques, chercheurs et utilisateurs collaborent dans une écoute mutuelle permanente.

VI REMERCIEMENTS

Ces travaux ont été soutenus par la Région Occitanie et par les Fonds Européens de Développement Régional (FEDER) dans le cadre du projet SONGES - Science des dONnées hétéroGènES (http ://textmining.biz/Projects/Songes) et par l'Agence nationale française de la recherche dans le cadre du programme Investissements d'avenir #DigitAg, référencé ANR-16-CONV-0004.

Références

Agrawal R., Srikant R. (1995). Mining sequential patterns. In P. S. Yu et A. L. P. Chen (Eds.), *Proceedings of the Eleventh International Conference on Data Engineering, March 6-10, 1995, Taipei, Taiwan*, pp. 3–14. IEEE Computer Society.

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

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

Coates J. (2013). Build it back better : Deconstructing food security for improved measurement and action. *Global Food Security 2*, 188 – 194.

Fabrègue M., Braud A., Bringay S., Ber F. L., Teisseire M. (2012). Extraction de motifs spatio-temporels à différentes échelles avec gestion de relations spatiales qualitatives. In *Actes du XXXème Congrès INFORSID, Montpellier, France, 29 - 31 mai 2012*, pp. 123–140.

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

FAO, FIDA, OMS, WFP, UNICEF (2018). L'état de la sécurité alimentaire et de la nutrition dans le monde en 2018 : renforcer la Résilience face aux changements climatiques pour La sécurité alimentaire et la nutrition. Technical report.

Fayyad U., Piatetsky-Shapiro G., Smyth P. (1996). From data mining to knowledge discovery in databases. *AI magazine 17*(3), 37–37.

Gupta M., Han J. (2013). Approaches for pattern discovery using sequential data mining. In *Data Mining : Concepts, Methodologies, Tools, and Applications*, pp. 1835–1851. IGI Global.

Hoddinott J. (1999). *Choosing Outcome Indicators Of Household Food Security, Vol. Technical Guide No 7*. International Food Policy Research Institute.

Huang Y., Zhang L., Zhang P. (2008, April). A Framework for Mining Sequential Patterns from Spatio-Temporal Event Data Sets. *IEEE Transactions on Knowledge and Data Engineering 20*(4), 433–448. doi:10.1109/TKDE.2007.190712.

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

Julea A., Méger N., Bolon P., Rigotti C., Doin M.-P., Lasserre C., Trouvé E., Lăzărescu V. N. (2010). Unsupervised spatiotemporal mining of satellite image time series using grouped frequent sequential patterns. *IEEE Transactions on Geoscience and Remote Sensing 49*(4), 1417–1430.

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

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

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

Maxwell D., Vaitla B., Coates J. (2014). How do indicators of household food insecurity measure up ? An empirical comparison from Ethiopia. *Food Policy 47*, 107–116.

Mendes L. F., Ding B., Han J. (2008). Stream sequential pattern mining with precise error bounds. In *2008 Eighth IEEE International Conference on Data Mining*, pp. 941–946. IEEE.

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

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

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

Perera D., Kay J., Koprinska I., Yacef K., Zaïane O. R. (2009). Clustering and sequential pattern mining of online collaborative learning data. *IEEE Transactions on Knowledge and Data Engineering 21*(6), 759–772.

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

Salas H. A., Bringay S., Flouvat F., Selmaoui-Folcher N., Teisseire M. (2012). Vers une approche efficace d'extraction de motifs spatio-séquentiels. In *Extraction et gestion des connaissances (EGC'2012), Actes, janvier 31 - février 2012, Bordeaux, France*, Volume RNTI-E-23 of *Revue des Nouvelles Technologies de l'Information*, pp. 201–212. Hermann-Éditions.

Salle P., Bringay S., Teisseire M. (2009). Mining discriminant sequential patterns for aging brain. In C. Combi, Y. Shahar, et A. Abu-Hanna (Eds.), *Artificial Intelligence in Medicine, 12th Conference on Artificial Intelligence in Medicine, AIME 2009, Verona, Italy, July 18-22, 2009. Proceedings*, Lecture Notes in Computer Science, pp. 365–369.

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

Shaw D.J. (2007). *World Food Security : A History Since 1945*, Volume 1. Palgrave MacMillan.

Srivastava J., Cooley R., Deshpande M., Tan P.-N. (2000, January). Web usage mining : Discovery and applications of usage patterns from web data. *SIGKDD Explor. Newsl. 1*(2), 12–23. doi:10.1145/846183.846188.

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

Tsoukatos I., Gunopulos D. (2001). Efficient mining of spatiotemporal patterns. In *Proceedings of the 7th International Symposium on Advances in Spatial and Temporal Databases*, SSTD '01, London, UK, UK, pp. 425–442. Springer-Verlag.

Valdés J. J. (2018). Extreme learning machines with heterogeneous data types. *Neurocomputing 277*, 38—-52.

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

Wang J., Hsu W., Lee M. (2005). Mining generalized spatio-temporal patterns. In L. Zhou, B. C. Ooi, et X. Meng (Eds.), *Database Systems for Advanced Applications, 10th International Conference, DASFAA 2005, Beijing, China, April 17-20, 2005, Proceedings*, Volume 3453 of *Lecture Notes in Computer Science*, pp. 649–661. Springer. doi:10.1007/11408079_60.

Wang K., Xu Y., Yu J. X. (2004). Scalable sequential pattern mining for biological sequences. In *CIKM '04 : Proceedings of the thirteenth ACM International Conference on Information and Knowledge Management*, New York, NY, USA, pp. 178–187. ACM. doi:1031171.1031209.

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

World Bank (2020). *World Bank statistics in Burkina Faso*. https://www.worldbank.org/en/country/burkinafaso. Accessed : 2020-04-01.

Wu X., Zhang X. (2019). An efficient pixel clustering-based method for mining spatial sequential patterns from serial remote sensing images. *Computers & Geosciences 124*, 128–139.

Yuan X., Chang W., Zhou S., Cheng Y. (2018). Sequential pattern mining algorithm based on text data : taking the fault text records as an example. *Sustainability 10*(11), 4330.