



HAL
open science

Détection de situations à risques pour des personnes fragiles sur des données multimodales

Thinhinane Yebda

► **To cite this version:**

Thinhinane Yebda. Détection de situations à risques pour des personnes fragiles sur des données multimodales. Réseau de neurones [cs.NE]. Université de Bordeaux, 2022. Français. NNT : 2022BORD0003 . tel-03956044

HAL Id: tel-03956044

<https://theses.hal.science/tel-03956044>

Submitted on 25 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

présentée à

L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE DE MATHÉMATIQUES ET D'INFORMATIQUE

par

THINHINANE YEBDA

POUR OBTENIR LE GRADE DE

DOCTEUR

SPÉCIALITÉ : INFORMATIQUE

Détection de situations à risques pour des personnes fragiles sur des données multimodales

Date de soutenance : 21/01/2022

Devant la commission d'examen composée de :

Pascal DESBARATS	Professeur	Université de Bordeaux	Président
Monique THONNAT	Directrice de Recherches	Inria Sophia Antipolis	Rapporteure
Mihai MITREA	MdC, HDR	Télécom SudParis	Rapporteur
Cathal GURRIN	Professeur	Dublin City University	Examinateur
Jenny BENOIS-PINEAU	Professeure	Université de Bordeaux	Directrice de thèse
Hélène AMIEVA	Professeure	Université de Bordeaux	Codirectrice de thèse
Olivier BLANC	PDG	InflexSys	Invité
Rafael IGLESIA	CTO	InflexSys	Invité

Titre

Détection de situations à risques pour des personnes fragiles sur des données multimodales

Résumé

Les prévisions actuelles montrent que la population mondiale vieillit. Cette population âgée est sujette à la fragilité avec des risques accrus dans leur quotidien. Ainsi, vieillir dans les meilleures conditions possibles et maintenir à domicile des sujets fragiles est devenu un défi, presque une nécessité. De nombreux projets ont alors été menés dans le monde, avec des architectures incluant des objets connectés avec différents types de capteurs et différentes techniques de détection pour assister les personnes âgées dans leur quotidien. La fragilité est un syndrome gériatrique caractérisé par la faiblesse, l'émaciation et une faible activité physique. L'objectif de cette thèse est d'identifier les risques au quotidien liés à la fragilité et d'y remédier. Nous proposons un dispositif portable pour prévenir les situations à risques des personnes fragiles à leur domicile. La prolifération des wearables a permis de collecter une énorme quantité de données physiologiques, motrices et environnementales pour évaluer l'état et les risques encourus par les humains dans différents environnements. Ainsi, instrumenter une personne fragile pour pouvoir l'accompagner dans sa vie quotidienne devient une tâche de plus en plus abordable. Différentes situations à risques sémantiques ont été rapportées telles que le risque de fraude et le risque d'accident domestique. Les situations à risques dites "sémantiques" sont à la fois complexes et complémentaires, ce qui rend difficile les études dans le monde réel avec des sujets fragiles. Pour nos recherches, nous avons mis en place un dispositif porté par des sujets volontaires sains enregistrant des données à l'aide d'un kit portable incluant des dispositifs IoT. À l'aide de ce dispositif, nous avons enregistré un dataset avec des situations à risques simulées dans un environnement réel. Après des tests concluants, effectués sur un dataset Lifelog pour la détection de situations à risques sémantiques, nous nous sommes intéressés à la détection de situations à risques sur séries temporelles. Nous nous sommes dans un premier temps essayé aux réseaux de neurone LSTM comme algorithme de base avant de tester des réseaux de neurones avec un ajout de l'attention.

Mots-clés

Réseaux neuronaux, IoT, CNN, RNN, LSTM, Attention, séries temporelles, eHealthcare.

Title

Detection of Risky Situations for Fragile People using Multimodal Data

Abstract

Current projections show that the world's population is aging. This elderly population is prone to frailty with increased risks in their daily lives. Thus, aging in the best possible conditions and maintaining frail subjects at home has become a challenge, almost a necessity. Numerous projects have been carried out around the world, with architectures including connected objects with different types of sensors and different detection techniques to assist the elderly in their daily lives. Frailty is a geriatric syndrome characterized by weakness, emaciation and low physical activity. The objective of this thesis is to identify the daily risks related to frailty and to remedy them. We propose a wearable device to prevent at-risk situations for frail individuals in their homes. The proliferation of wearables has enabled the collection of a huge amount of physiological, motor and environmental data to assess the condition and risks faced by humans in different environments. Thus, instrumenting a frail person to be able to accompany them in their daily life is becoming an increasingly affordable task. Different semantic risky situations have been reported such as fraud risk and domestic accident risk. The difficulty of real world studies with frail subjects is that semantic risky situations are both complex and "complementary". For our research we set up a device worn by healthy volunteers recording data using a wearable kit including IoT devices. Using this device we recorded a dataset with simulated risk situations in a real environment. After conclusive tests, performed on a Lifelog dataset for the detection of semantic risk situations, we were interested in the detection of risk situations on time series. First, we tried LSTM neural networks as a basic algorithm and then we tested neural networks with added attention.

Keywords

Neural Networks, IoT, CNN, RNN, LSTM, Soft and Hard Attention, Time series, eHealthcare.

Table des figures

2.1	Évolution des personnes âgées de 60 ans et plus en France de 1960 à 2060. Blanpain et Chardon (2010a)	15
2.2	Un cadre de santé publique pour un vieillissement sain Organization (2015)	17
2.3	Différents types de capteurs et le principe de fusion de capteurs.	25
2.4	Répartition des approches de détection de l'absorption de liquides et de nourriture en fonction de la technologie utilisée dans l'état de l'art. L'orange représente les dispositifs portables, le violet la fusion, le vert les conteneurs intelligents, le bleu les surfaces avec capteurs intégrés et le gris les approches basées sur la vision et l'environnement Cohen et al. (2021)	26
2.5	Système de collecte et de traitement de données de détection d'une tasse intelligente, où sont annotés les épisodes de consommation (s'asseoir et boire, se tenir debout et boire et marcher et boire) et les épisodes de non-consommation (verser de l'eau dans la tasse, tenir la tasse et marcher et tenir la tasse et discuter avec d'autres personnes). Liu et al. (2020)	28
2.6	(a) une jeune participante qui boit de l'eau tout en déjeunant ; (b) une jeune participante qui boit du jus d'orange tout en lisant le journal ; (c) un participant plus âgé qui boit du vin blanc froid tout en lisant un message au téléphone et en déjeunant ; et (d) un participant plus âgé qui se fait resservir du jus d'orange tout en dégustant son déjeuner Jayatilaka et Ranasinghe (2017)	29
2.7	Dispositif et schéma du positionnement du dispositif portable Borzì et al. (2021)	31
2.8	La relation entre la démarche et les chutes chez les personnes âgées atteintes de démence. Taylor et al. (2014)	32
2.9	Illustrations graphiques de la Gated Recurrent Unit (GRU) originale (en haut à gauche), de la Gated Recurrent Unit with Decay (GRU-D) proposée (en bas à gauche) et de l'architecture complète du réseau (à droite) Che et al. (2018)	38
2.10	(à droite) l'architecture du réseau présenté par Friedrich et al. (2021) , (à gauche) Détails d'architecture des blocs constituant le réseau de neurone.	38

2.11	Architecture d'un réseau Convolutional Neural Networks (CNN) pour la reconnaissance de lettres manuscrites (Figure extraite du Lecun et al. (1998)).	39
2.12	L'architecture CNN 3D pour la reconnaissance d'actions humaines, présentée par Ji et al. (2012).	40
2.13	Architecture présentée par Feichtenhofer et al. (2016) pour la fusion des informations spatio-temporelles pour la reconnaissance d'action sur les données vidéo.	41
3.1	Résumé des éléments de l'examen physique lors de l'évaluation du risque de chute chez une personne âgée Cuevas-Trisan (2017).	47
3.2	Architecture du système et topologie du réseau présentée Sokullu et al. (2020).	48
3.3	Illustration du risque d'intrusion sur le dataset NTCIR Yebda et al. (2020).	49
3.4	À droite : Signaux bruts, mesurés par un accéléromètre placé sur le poignet. À gauche : le schéma d'un geste de prise de liquide du début à la fin.Chun et al. (2019).	51
3.5	Relevés par accéléromètre et gyroscope des activités de prise de médicaments Fozoonmayeh et al. (2020).	53
3.6	Réponse électrodermale sur 3 plages temporelles (Figure extraite de Drover et al. (2017).	61
3.7	Accélération verticale gauche et l'accélération verticale droite d'un corps humain en mouvement (Figure extraite de Garbarino et al. (2014)).	61
3.8	Aperçu des portées et des débits possibles avec chaque technologie de transmission.(Figure extraite de René et Alexandre).	64
3.9	Architecture du système pour le projet BIRDS.	66
3.10	Dispositif d'acquisition des données capteurs.	67
3.11	Microsoft Band 2.	68
3.12	MetaWearC Pro.	69
3.13	Présentation de la caméra retenue pour la solution finale.	71
3.14	Présentation du bracelet E4 Empatica (2021).	74
3.15	Présentation du dispositif connecté MetamotionR.	77
3.16	Présentation de la solution BIRDS retenue.	78
3.17	À gauche la figure représente le signal timestamp de référence et le signal timestamp filtré. À droite l'erreur quadratique entre le signal timestamp de référence et le timestamp output avant et après filtrage. Taille du filtre = 1, nombre d'échantillons= 1000, pas de descente = 0.0001.	85

TABLE DES FIGURES

3.18	À gauche la figure représente le signal timestamp de référence et le signal timestamp filtré. À droite l'erreur quadratique entre le signal timestamp de référence et le timestamp output avant et après filtrage. Taille du filtre = 2, nombre d'échantillons= 1000, pas de descente = 0.0001.	85
3.19	À gauche la figure représente le signal timestamp de référence et le signal timestamp filtré. À droite l'erreur quadratique entre le signal timestamp de référence et le timestamp output avant et après filtrage. Taille du filtre = 1, nombre d'échantillons= 1000, pas de descente = 0.0001.	86
3.20	À gauche la figure représente le signal timestamp de référence et le signal timestamp filtré. À droite l'erreur quadratique entre le signal timestamp de référence et le timestamp output avant et après filtrage. Taille du filtre = 1, nombre d'échantillons= 1000, pas de descente = 0.00001.	86
3.21	À gauche la figure représente le signal timestamp de référence et le signal timestamp filtré. À droite l'erreur quadratique entre le signal timestamp de référence et le timestamp output avant et après filtrage. Taille du filtre = 1, nombre d'échantillons= 1000, pas de descente = 0.0001.	87
3.22	À gauche la figure représente le signal timestamp de référence et le signal timestamp filtré. À droite l'erreur quadratique entre le signal timestamp de référence et le timestamp output avant et après filtrage. Taille du filtre = 1, nombre d'échantillons= 2000, pas de descente = 0.0001.	87
3.23	À gauche la figure représente le signal timestamp de référence et le signal timestamp filtré. À droite l'erreur quadratique entre le signal timestamp de référence et le timestamp output avant et après filtrage. Taille du filtre = 1, nombre d'échantillons= 5000, pas de descente = 0.0001.	88
4.1	Exemple du modèle de caméra porté pour l'enregistrement du dataset LSC NTCIR 13 Gurrin et al. (2014)	94
4.2	Exemple d'images enregistrées par le lifelogueur dans le cadre du corpus LSC NTCIR 13 Gurrin et al. (2014)	94
4.3	Présentation de l'architecture du dataset NTCIR 13 Gurrin et al. (2014)	96
4.4	Mise-en-correspondance des concepts du NTCIR 13 et des risques sémantiques.	97
4.5	Architecture AlexNet Krizhevsky et al. (2012a)	100
4.6	Méthodes d'augmentation des données, rotation et retournement. Rangée du haut : remplissage flou. Rangée du bas : Remplissage de bruit gaussien.	102

4.7	Courbes d'AlexNet pour le risque d'accident domestique. La précision maximale (83,33 %) est atteinte à l'époque 14.	104
4.8	Courbes d'AlexNet pour le risque de fraude. La précision maximale (96,88 %) est atteinte à l'époque 22.	105
4.9	Distribution des données LSC NTCIR 13 selon la taxonomie BIRDS.	106
4.10	Courbes LSTM obtenues sur le jeu de données LSC.	107
5.1	Diagramme de la durée d'enregistrement du corpus BIRDS.	112
5.2	Exemples d'images provenant du jeu de données LSC Gurrin et al. (2019b).	113
5.3	Données du Corpus BIRDS.	115
5.4	Interface d'annotation BIRDS.	116
5.5	Données du Corpus BIRDS.	116
5.6	Exemples du floutage.	119
6.1	Architecture d'une cellule de mémoire à long et court terme (LSTM).	129
6.2	Architecture du réseau de neurones LSTM entrées multiples et sorties multiples.	130
6.3	Architecture du réseau de neurones LSTM entrées multiples et sorties multiples intégrant un auto-encodeur avec un bloc d'attention.	135
6.4	Matrices de confusion obtenues avec les méthodes d'analyse de données. Dans l'ordre LSTM auto-encodeur avec attention molle et LSTM auto-encodeur sans attention, LSTM auto-encodeur avec attention dure proposées sur le dataset BIRDS avec la classification multi-classes. Les classes sont représentées par Sans risque - cl(0), Risque de chute - cl(1), Risque d'accident domestique - cl(2), Hydratation par apport d'eau - cl(3).	142

Liste des tableaux

3.1	Tableau résumant la taxonomie envisagée dans le cadre du projet BIRDS.	52
3.2	Les différentes caractéristiques des signes vitaux.	55
3.3	Les différentes caractéristiques des signes vitaux.	56
3.4	Les capteurs contextuels, inertiels et physiologiques les plus couramment utilisés dans les objets connectés.	60
3.5	Technologies de communication pour les maisons intelligentes (Tableau extrait de Majumder et al. (2017)).	65
3.6	Caractéristiques techniques du Microsoft Band 2.	68
3.7	Caractéristiques techniques du MetaMotionC Pro.	70
3.8	Caractéristiques techniques d’Empatica E4.	75
3.9	Caractéristiques techniques du dispositif MetaMotionR.	77
3.10	Tableau récapitulatif des capteurs utilisés par le dispositif BIRDS.	79
3.11	Caractéristiques de la dispositif IoT MetaMotionR avec les capteurs retenus.	80
3.12	Résultats obtenues par l’algorithme de synchronisation LMS en temps réel au sens de la MSE (mean square error) entre le timestamp du dispositif IoT et le timestamp lié à l’horloge interne du contrôleur principal au moment de la réception des données.	88
4.1	Concepts pour les situations à risques à partir du dataset Lifelog.	97
4.2	Données visuelles annotées suivant la taxonomie BIRDS pour la détection de situations à risques sémantiques	99
4.3	Ensembles de données utilisés pour la formation pour chaque situation de risque.	102
4.4	Distribution des données LSC NTCIR 13 obtenue avec une annotation correspondant à notre taxonomie BIRDS.	106
5.1	Description du corpus BIRDS.	113
5.2	Situations à risques enregistrées dans le corpus BIRDS.	114
5.3	Différence de timestamp entre le début de collecte des données.	118
6.1	Distribution des situations à risques dans le dataset sélectionné.	138
6.2	Tableau d’optimisation du taux d’apprentissage.	139
6.3	Tableau d’optimisation de la taille des lots.	139

6.4 Comparaison des résultats obtenus avec les méthodes proposées sur le dataset BIRDS sélectionné avec les scores de précision et les F-Scores pour la classification multi-classes. Les classes sont représentées par Sans risque - cl(0), Risque de chute - cl(1), Risque d'accident domestique - cl(2), Hydratation par apport d'eau - cl(3). 141

Sommaire

1	Introduction générale	4
1.1	Contexte	4
1.2	Contributions	9
2	État de l’art	13
2.1	Introduction	13
2.2	Le vieillissement et la fragilité	14
2.3	Les objets connectés pour la surveillance des personnes âgées	18
2.4	Réseaux de neurones profonds pour les technologie eHealthcare	35
2.5	Conclusion	41
3	De la taxonomie à la mise en place du dispositif d’acquisition	44
3.1	Introduction	44
3.2	Définition des situations à risques	45
3.3	Choix des mesures	54
3.4	Technologies d’acquisition et de transmission pour la solution connectée	59
3.5	Solution retenue	64
3.6	Synchronisations des données	80
3.7	Conclusion	89
4	Analyses de situations à risques sur des données Lifelog	91
4.1	Introduction	91
4.2	Les données Lifelog	91
4.3	Présentation du dataset NTCIR 13	93
4.4	Mise en correspondance des concepts avec la taxonomie de risques et réannotation des données NTCIR 13	96
4.5	Reconnaissance des risques sémantiques sur les données du corpus NTCIR 13	98
4.6	Conclusion	108
5	Dataset Birds	109
5.1	Introduction	109
5.2	Dataset BIRDS	109
5.3	Annotation des données	115
5.4	Pré-traitement des données	117
5.5	Conclusion	123

6	Analyse des données BIRDS	125
6.1	Introduction	125
6.2	Méthodes de traitement des données	127
6.3	Les réseaux de neurones auto-encodeurs avec attention	132
6.4	Expérimentations et résultats	136
6.5	Conclusion	143
7	Conclusion générale et perspectives	146
	Annexe A	151
	Bibliographie	153
	Liste des acronymes	173
	Table des matières	176

Remerciements

Ce manuscrit n'échappe pas à une règle, celle des remerciements. Un exercice important, car il résume assez bien 3 ans de travail, de vie et de rencontres.

Mes premiers remerciements s'adressent à mes rapporteurs, Monique Thonnat et Mihai Mitrea pour avoir accepté de rapporter cette thèse. Je remercie également mes examinateurs et le président du jury pour leurs disponibilités ainsi que d'avoir accepté de faire partie de mon jury de thèse.

Je remercie naturellement mes directrices de thèse Jenny Benois-Pineau et Hélène Amieva pour leurs encouragements, leur encadrement, leur soutien durant ces trois années et surtout de m'avoir permis de travailler dans un environnement pluridisciplinaires fort enrichissant. Je remercie également l'entreprise InflexSys d'avoir permis que cette thèse se réalise. Cette thèse m'a permis de travailler avec des chercheuses et chercheurs venant d'horizons très différents notamment dans le cadre du projet UBX/Waterloo. Merci à Marion Pech, sans qui je n'aurais pas pu faire cette immersion dans le monde de la neuropsychologie. Merci à Laura Middleton, Sam et Max qui ont enrichi notre taxonomie et qui ont apporté une coloration différente à notre projet.

J'aimerais également remercier mes collègues et amis de l'équipe AIV. Ils ont été présents pour moi au LaBRI et parfois en dehors. Ils ont su m'apporter leur aide, leurs conseils ainsi que leur soutien au quotidien lorsque cela était nécessaire : Abraham, Pierre-Etienne, Attila, Karim, Rupayan et Arthur. Merci également à Akka Zemmari et Pascal Desbarats pour leurs conseils et leur expertise sur les sujets de Deep Learning et traitement de données. Je n'oublie pas non plus tous les stagiaires qui sont passés dans notre bureau Rock, Milto, Azif, Meghna et bien d'autres. Vous avez su nous apporter votre aide et vos bonnes énergies.

Dans une tonalité un peu différente, mais non moins importante, je tiens à remercier tous mes collègues et surtout amis du LaBRI et de l'Afodib qui ont rendu mon quotidien sûrement plus plaisant. Parmi eux, je citerai Simon, Rohan, Alex, Christelle, les deux Remi, Paul, Lamine, Luis, Julien, Jason, Mathias, Abraham et les nombreux autres que j'oublie. Vous m'avez fait découvrir vos univers respectifs et pour cela, je vous en suis reconnaissante. Un merci particulier aux personnes qui ont pris de leur temps pour m'aider à l'annotation des datasets.

J'aimerais remercier les dirigeants et tous mes collègues d'InflexSys, pour leur bienveillance et leur bonne humeur. Un grand merci à Benjamin, Julien, Daly, Ye-

len, Nicolas et Antoine pour leur soutien technique et sans qui l'application BIRDS n'aurait pas vu le jour. Un merci pour Christine qui a admirablement pallié à mes talents de couturière. Je ne peux pas non plus oublier de remercier Olivier Champarnaud, qui a toujours accédé à mes demandes de relectures malgré son planning très chargé. Merci également à tous mes collègues pour ces parties de Baby et de pauses café Mélanie, Natasha, Sebastien, Olivier, Catherine, etc. Merci à Coco, Sylvain, Mathieu et Julien pour m'avoir fait découvrir l'univers Magic.

Un grand merci à l'administration du LaBRI et de l'école doctorale qui m'ont offert un lieu de travail de qualité pour mes recherches et qui ont toujours su répondre à mes questions et m'accompagner dans mes multiples démarches administratives.

Je remercie également ma cohorte de relecteurs et relectrices : Olivier, Alex, Didi et Nico, vos retours et vos remarques constructives m'ont été tutélaires.

Pour finir au mieux cet exercice, je me dois d'avoir quelques mots à ma famille en particulier ma petite soeur, mon père. Je vous remercie de votre soutien, vos encouragements et surtout votre affection inébranlable.

Merci enfin à Nicolas, mon fiancé, sans qui la vie serait bien morne, merci de me supporter et de me soutenir au quotidien. Tu me pousses à être la meilleure version de moi même et pour cela je t'en saurais pour toujours gré.

Chapitre 1

Introduction générale

La population mondiale continue de vieillir de manière significative. En France, une étude statistique réalisée par l’Institut National de la Statistique et des Études Économiques (INSEE) montre que près d’une personne sur trois aura plus de 60 ans en 2050, contre 21.5 % en 2007 [Blanpain et Chardon \(2010b\)](#). En raison de ce changement démographique, de nouveaux modèles de vieillissement où la technologie trouve un rôle sont en train d’émerger.

Des services sont nés à partir de ces modèles permettant la surveillance et prévention des événements indésirables tels que la chute, les situations de stress et les oublis majeurs.

Pour ce faire, les acteurs des nouvelles technologies mettent au point des dispositifs visant à améliorer la qualité de vie des personnes âgées. De nombreux projets sont alors menés dans le monde [Rajagopalan et al. \(2017\)](#), [Hernández-Peñaloza et al. \(2018\)](#), etc, avec des architectures comprenant des objets connectés intégrant différents types de capteurs (physiologiques, contextuels, ambiants, etc) et différentes techniques de détection à base d’objets connectés.

La plupart de ces algorithmes de détection nécessitent de grandes quantités de données qui peuvent être difficiles à obtenir, en particulier des données décrivant des événements anormaux qui sont, par définition, des événements rares.

Le point crucial pour tous ces modèles est le manque de données représentant les nombreuses situations à risques avec plusieurs profils de personnes et ce, dans des situations écologiques à leur domicile.

1.1 Contexte

Cette thèse de doctorat s’inscrit dans le cadre de la détection d’événements à risques pour des systèmes de suivi de personnes fragiles. Ce travail a été réalisé dans le cadre du dispositif CIFRE (Convention Individuelle de Formation par la Recherche) de l’ANRT (Agence Nationale de la Recherche et de la Technologie) et a été sponsorisé par l’entreprise Inflexsys, qui est une entreprise innovante française qui accompagne le Laboratoire Bordelais de Recherche en Informatique (LaBRI) ainsi

Bordeaux Population Health (BPH) dans la réalisation du projet notamment sur les aspects conception et développement du dispositif de captation des données. Plus tard, nous a rejoint l'équipe de Kinesiologie de l'université de Waterloo (Canada). Ils nous ont aidé à étendre la taxonomie des risques et à pouvoir mettre en place une expérience miroir. Dans ce contexte, l'objectif principal de cette thèse est d'étudier et de proposer des solutions de détection d'événements à risques pour le suivi de personnes âgées fragiles à partir de signaux multidimensionnels issus de différents types de capteurs portés sur la personne.

Avec le vieillissement, apparaissent des fragilités et des pathologies chroniques pouvant entraîner des déclin fonctionnels physiques et cognitifs favorisant l'entrée en dépendance et responsables d'une réduction de l'autonomie. Assister des personnes fragiles par la prévention des situations à risques devient alors un vrai défi et représente une réelle opportunité pour améliorer leur vie au quotidien.

En complément de ce vieillissement, une augmentation des besoins de soins de longue durée a été enregistrée. Selon l'Organisation Mondiale de la Santé (OMS), il s'agit là d'une des conséquences directes du nombre croissant de personnes âgées souffrant de maladies chroniques [Mousavi et Saba \(2010\)](#).

De ce fait, les structures de santé existantes sont confrontées à de multiples défis. Dès lors, l'utilisation des nouvelles technologies de la santé "smartHealth" s'impose. Le concept de "smartHealth" repose sur l'accès à de meilleurs outils de diagnostic ce qui encourage le changement du soin curatif au soin préventif.

De plus, l'utilisation des nouvelles technologies dans la vie de tous les jours, en particulier les objets connectés est aujourd'hui admise et même encouragée par notre société et notamment les pouvoirs publics.

1.1.1 La fragilité : mal des sociétés évoluées

Parmi les préoccupations soulevées par l'allongement de l'espérance de vie, la question de la qualité de vie accompagnant cet allongement de l'espérance de vie se pose, notamment eu égard à la prévalence des maladies dégénératives chroniques et à l'utilisation accrue, avec l'avancée dans l'âge, de services de santé et de médicaments. Une catégorie médicale assez récente vient enrichir le répertoire des « qualités » des processus de vieillissement : la « fragilité ». Le terme fragilité est utilisé dans plusieurs disciplines et avec diverses acceptions. En gériatrie, il fait référence à la diminution des réserves physiologiques et la réduction des capacités fonctionnelles chez les personnes vieillissantes. Le syndrome de fragilité résulte d'une réduction multisystémique des aptitudes physiologiques limitant les capacités d'adaptation à un stress physiologique. C'est un état qui peut évoluer vers une rupture d'état d'équilibre et entraîner des complications. Selon la définition de [Fried et al. \(2001\)](#), 5 critères peuvent être utilisés pour détecter les sujets dits "fragiles": un sentiment

d'épuisement, un ralentissement de la vitesse de marche, une faiblesse musculaire, une perte de poids involontaire (10 livres au cours de la dernière année) et une faible activité physique. Ce sont des résultats composites de plusieurs systèmes organiques. Un sujet âgé est considéré comme fragile s'il présente au moins trois de ces cinq critères [Fried et al. \(2001\)](#). La fragilité chez le sujet de plus de 65 ans est corrélée au risque de déclin fonctionnel, d'institutionnalisation, et de décès. Parmi d'autres facteurs, la fragilité contribue à augmenter le risque de chute. La fragilité augmentant avec l'âge, les personnes âgées deviennent alors particulièrement sujettes aux chutes. Bien que peu dangereuse pour une jeune personne en bonne santé, une chute peut en revanche entraîner de sérieuses blessures à une personne âgée, et peut parfois même être fatale.

Les répercussions de la chute chez les personnes âgées sont multiples [Araújo Neto et al. \(2017\)](#). La chute peut avoir un impact négatif sur la mobilité, la plupart étant des fractures, le plus souvent du poignet et de la hanche, ou un traumatisme crânien, ce dernier pouvant entraîner des lésions cérébrales. Outre la douleur et l'inconfort, les blessures physiques réduisent la mobilité, parfois pendant longtemps ou même de façon irréversible, ce qui entraîne une plus grande dépendance. Par ailleurs, une chute, même sans blessure physique grave, peut entraîner l'anxiété, la dépression et la peur de tomber à nouveau, augmentant ainsi le risque de récurrence, symptômes communément connus sous le terme "syndrome post-chute". Cette peur entraîne alors une réduction des interactions et des activités sociales, ce qui augmente considérablement le niveau de fragilité, et donc les chances de tomber à nouveau [Araújo Neto et al. \(2017\)](#).

1.1.2 Solutions proposées pour les populations fragiles

Depuis de nombreuses années, nombreuses sont les études consacrées aux systèmes de détection de situations à risques et notamment la chute dans la littérature [Friedrich et al. \(2021\)](#), [Vizitiu et al. \(2021\)](#), [Cohen et al. \(2021\)](#). Un système de surveillance peut être défini comme un dispositif d'assistance dont l'objectif principal est de donner un aperçu de l'activité d'une personne à des fins de santé, soit sous forme de statistiques générales telles que des indicateurs d'activité, soit sous forme d'alertes en cas de situations dangereuses. Les systèmes de détection destinés aux personnes âgées peuvent servir deux objectifs principaux, à savoir estimer le niveau de fragilité et alerter lorsque la personne est en danger, notamment en cas de chute. La détection précoce de la fragilité est considérée comme étant d'un intérêt majeur par la communauté médicale à des fins de prévention, et il a été démontré que les stratégies thérapeutiques ciblées réduisent considérablement les conséquences de la fragilité [Lucine, Meek](#). Il a aussi été démontré que la peur de tomber est susceptible d'être diminuée par la conscience d'être surveillé par un détecteur de chute robuste. Pour ces raisons, la surveillance de l'activité physique des personnes âgées dans les maisons de retraite est essentielle, car cela permet de concentrer les ressources sur les personnes particulièrement vulnérables à la fragilité et aux chutes.

Les innovations en matière des soins de santé, de capteurs et d'analyse de données par l'intelligence artificielle offrent des possibilités d'améliorer les soins de santé personnalisés et de réduire les coûts souvent onéreux des soins médicaux.

La réduction des situations à risques à l'aide de la surveillance diagnostique, de capteurs non invasifs, de données sonores et visuelles et de fusion de ces données offre aux individus fragiles des informations précises en matière de prévention.

Les objets connectés sont depuis plusieurs années appliqués au diagnostic de santé. Avec la miniaturisation des capteurs, la taille des objets connectés est de plus en plus petite avec des coûts de plus en plus bas. Les données mesurées par les différents capteurs intelligents et les analyses de ces solutions "smartHealth" personnalisées facilitent la détection précoce et la compréhension des risques associés à la santé des personnes fragiles et à leur bien-être en général.

Les flux de données collectés sont traités par des systèmes à base d'intelligence artificielle pour fournir des solutions de soins personnalisées et intelligentes complétant ainsi la technologie des objets connectés [Knickerbocker et al. \(2018\)](#).

De nombreuses applications de santé basées sur l'Internet Of Things (IoT) sont développées pour surveiller la santé des patients à distance. Des plateformes de surveillance de la santé sont développées sur la base d'architectures IoT pour surveiller les maladies chroniques. Dans la littérature, certaines études se concentrent sur l'intégration de l'IoT dans les solutions de santé. Ces systèmes de soins de santé basés sur l'IoT réduisent les coûts de traitement des personnes âgées en développant des plateformes et des applications pour améliorer les services de santé. La fréquence cardiaque et la température corporelle sont des caractéristiques essentielles du corps humain pour déterminer l'état de santé. Les données de détection sont utilisées pour classer les conditions du patient comme normales et anormales. Des études supposent que d'ici 2025, il y aura environ 50 à 100 milliards d'appareils connectés à Internet. Les développeurs IoT utilisent largement Raspberry Pi et Arduino Uno pour déployer des applications IoT au stade de développement. Dans la littérature, de nombreuses études ont tenté d'intégrer l'IoT dans les systèmes de surveillance des soins de santé pour les patients âgés en raison du nombre croissant de la population âgée. Ci-dessous sont citées quelques solutions existantes sur le marché :

- La détection de chute à l'aide de capteurs contextuels [Assistance](#); [Vigeo](#) ;
- La détection précoce des maladies chroniques et la possibilité de retarder ou de prévenir leur progression [Lucine](#) ;
- La compréhension du comportement individuel, des traitements médicamenteux et de l'efficacité des traitements sur les activités de la vie quotidienne [Meek](#) ;
- Des soins personnalisés basés sur l'Acide DésoxyriboNucléique (ADN) du sujet, les diagnostics médicaux et les tendances de ses soins de santé par rapport à ses besoins.

L'utilisation de systèmes de surveillance entraîne le traitement d'une grande quantité de signaux temporels. De manière générale, nous constatons ces dernières années une prolifération de systèmes basés sur des capteurs à des fins médicales, ce

qui amène la problématique de savoir comment extraire le meilleur de ces données. Qu'ils soient unis ou multivariés, ces signaux sont généralement temporels. Il peut s'agir par exemple d'électrocardiogrammes, d'électroencéphalogrammes, d'accélération des parties du corps, d'enregistrements vidéo, de sons, etc.

1.1.3 Traitement des données issus des systèmes eHealthcare et la part donnée au Deep Learning

Selon la tâche à accomplir, le traitement des signaux temporels peut répondre à plusieurs enjeux. Tout d'abord, les signaux peuvent être très redondants ou de grande dimension. Par exemple, la détection d'événements à partir d'une vidéo peut être facile car les images contiennent beaucoup d'informations, mais se révèle être coûteuse en termes de calcul et exige donc certaines étapes de pré-traitement permettant de se ramener à des signaux utilisables pour les modèles de détection. Deuxièmement, les signaux peuvent être bien adaptés à la tâche (par exemple, électroencéphalogramme pour la surveillance de l'anesthésie, accélération du corps pour l'analyse de la marche) ou beaucoup plus difficiles à interpréter. C'est le cas par exemple de l'utilisation des radars pour la surveillance des personnes, ou du son pour la détection des chutes. En conséquence, alors que certains modèles reposent sur de simples heuristiques pour accomplir leurs tâches prédictives, exécuter les mêmes tâches avec des modèles similaires et des signaux issus de capteurs différents peut échouer.

Aujourd'hui, nous assistons à une expansion de l'intérêt porté au eHealthcare et à une plus grande diversité des algorithmes utilisés. Un autre aspect important de notre cadre de travail est l'application de modèles étudiés en environnement contrôlé avec les applications sur le terrain. Cela peut s'exprimer de différentes manières. Premièrement, de nombreuses applications exigent de travailler en temps réel ou au moins en temps court. C'est évidemment le cas des systèmes de détection des chutes, mais aussi de toute tâche de surveillance ayant une valeur essentielle dans la détection précoce des événements, par exemple lors du contrôle de l'anesthésie pendant une intervention chirurgicale, ou qui nécessite une réponse rapide pour une utilisation pratique, ce qui est le cas de méthodes de diagnostic médical. Cependant, ces systèmes souffrent de l'explosion des données. Pour surmonter ces problèmes, il est nécessaire d'intégrer directement les microcontrôleurs à très faible consommation, les technologies de capteurs avancées et les algorithmes analytiques. La complexité des opérations typiques appliquées à l'apprentissage automatique et au traitement du signal, par exemple les inversions et décompositions de matrices, est estimée à $O(n^3)$. Cette complexité implique que s'il y a 100 échantillons, une opération algébrique simple nécessite environ 1 million d'instructions de boucle interne.

Deuxièmement, la plupart des méthodes traditionnelles d'apprentissage machine tiennent généralement compte d'hypothèses commodes, la première étant que l'entraînement ait accès à un ensemble de données étiquetées suffisamment important, ce qui, dans les applications réelles, est déjà discutable. Lorsque des systèmes innovants sont conçus, la collecte de données est la plupart du temps coûteuse, que

ce soit en temps ou en argent. Les données peuvent être fastidieuses à étiqueter pour plusieurs raisons. La difficulté d'étiquetage peut être due à la complexité des signaux à interpréter, certains requérant un regard expert (par exemple les données médicales), ou à la grande variété de signaux qui nécessitent d'être étiquetés afin d'englober la plus large diversité possible de signaux d'entrée (par exemple les signaux d'activité de la vie quotidienne pour la détection de chute). Cette limitation peut conduire à une petite quantité de données étiquetées, ou à une distribution déséquilibrée des classes, c'est-à-dire lorsqu'une classe est sous-représentée par rapport aux autres. Par conséquent, la rareté des données et le déséquilibre des classes peuvent devenir un problème majeur lorsque nous souhaitons mettre en place un modèle d'apprentissage fiable.

Enfin, pour pallier les problèmes de pénurie de données ou de coûts liés à la maîtrise d'un environnement, les données peuvent être collectées dans un environnement contrôlé ou même synthétisées, ce qui induit alors une dissemblance avec les données dites opérationnelles, à savoir les données que nous visons à traiter avec le modèle final. Ce phénomène pouvant se produire dans un très large éventail de sujets, le problème général du traitement des différences entre l'environnement de création des modèles ou des données et la tâche finale souhaitée fait l'objet d'un intérêt croissant, et relève d'un domaine de l'apprentissage machine appelé apprentissage par transfert. Dans notre cadre, nous rencontrons une situation de transfert typique avec deux ensembles de données similaires, l'un étant construit dans un environnement contrôlé et l'autre étant collecté sur le terrain.

1.2 Contributions

Dans ce manuscrit nous présentons les contributions suivantes de notre travail de recherche :

Chapitre 2 : État de l'art La première contribution de cette thèse consiste en un aperçu des systèmes de surveillance eHealthcare existants conçus pour la détection de situations à risques. Cet examen est mené tout au long de la chaîne de traitement, depuis la génération du signal jusqu'à la décision finale. Cette revue bibliographique permet d'abord de situer le système BIRDS parmi cette vaste gamme de systèmes et, à l'aide d'un ensemble de critères établis pour l'utilisation pratique de tels systèmes, elle met en évidence les avantages des systèmes portés. Une partie de cette étude est consacrée à la classification des données multimodales, soulignant l'utilisation plus fréquente de méthodes de classification supervisée avancées.

Chapitre 3 : Les situations à risques et le choix de mesures Le chapitre explore les situations à risques que rencontrent les personnes fragiles au quotidien. Dans cette étude pluridisciplinaire, les situations à risques ont été définies par les neuropsychologues travaillant avec des populations âgées soumises à des fragilités et

des kinésiologues, nous avons participé à la définition de ces situations sur le plan de faisabilité. Ainsi un choix approprié a été porté sur des mesures adaptées à notre taxonomie. Une sélection de dispositifs adaptés a été réalisée en tenant compte de l'étude de l'état de l'art sur les dispositifs IoT. Nous finissons par exposer notre taxonomie avec une description des différentes situations à risques et nous parlons du choix de mesures permettant la détection de celles-ci.

Chapitre 4 : Analyse de situations à risques sur des données Lifelog Ce chapitre explore des méthodes bien connues d'apprentissage profond appliquées sur un dataset Lifelog présentant des données multimodales. Ce chapitre peut être considéré comme une preuve de concept de notre projet et de notre taxonomie. Les modèles sont entraînés sur des sous-ensembles de données du dataset Lifelog. Ce dataset présentant des données hétérogènes collectées par une personne saine et décrivant la vie d'une personne active, nous avons pu tirer des situations qui peuvent s'apparenter à des situations à risques du quotidien des personnes fragiles. Après une description brève du dataset, nous présentons la réannotation qui a été réalisée en mettant en place un concept matching avec nos situations à risques. Ainsi nous avons pu les réannoter en tenant compte de notre taxonomie. Sur les données visuelles, nous avons testé des réseaux de neurones convolutionnels bien connus (Alexnet et ResNet). Une procédure d'augmentation des données est également proposée et améliore considérablement la précision des modèles. Sur les données physiologiques, nous avons testé le réseau de neurones récurrent bien connu le Long Short-Term Memory (LSTM) (Long Short Term Memory). Dans la mesure où les résultats obtenus témoignent des problèmes de classes très déséquilibrées et d'une fréquence d'enregistrement trop basse dû à un contexte d'enregistrement différent du notre problématique, nous avons arrêté les expérimentations sur les données physiologiques.

Chapitre 5 : Conception et mise en place du dispositif BIRDS Après la définition de la taxonomie et un choix de mesures adaptés à celle-ci, nous nous sommes penchés sur la conception du dispositif intégrant des systèmes multi-capteurs. Dans ce genre de projet, la phase de conception du dispositif est importante car c'est l'étape qui tient compte des différents paramètres liés à l'acceptabilité et à la facilité d'utilisation du dispositif par des personnes âgées souvent néophytes à ce genre de technologies. Pour répondre à cette problématique des enquêtes réalisées par nos partenaires neuropsychologues sur l'ergonomie et la facilité d'utilisation, nous ont permis d'améliorer le dispositif et l'application de collecte tout au long du projet. Ainsi, plusieurs solutions ont été testées avant d'aboutir à une solution finale. Nous évoquerons aussi le choix des dispositifs IoT. Le développement de l'application de collecte de données réalisé avec l'entreprise InfleXsys a également constitué une étape importante de la mise en place du dispositif BIRDS. Enfin, nous parlerons de la synchronisation des données des différents dispositifs de la solution BIRDS à travers un algorithme corrigeant en temps réel le déphasage entre les différents dispositifs de captation de données.

Chapitre 6 : Le dataset BIRDS Dans ce chapitre, nous décrivons le protocole d'enregistrement des données BIRDS en mettant l'accent les scénarios d'enregistrement pour chaque situation à risques. Ensuite, nous décrivons l'interface d'annotation des données BIRDS, ainsi que les évolutions qui lui ont été apportées au fil de l'évolution de la taxonomie. Nous finissons ce chapitre par la description des prétraitements réalisés sur les données.

Nous évoquerons le problème des données manquantes et de leur estimation par la moyenne et nous décrirons aussi le rééquilibrage des classes réalisé afin de limiter la classe *Sans risques* par rapport aux classes décrivant les situations à risques enregistrées.

Chapitre 7 : Deep Neural Network (DNN) pour la reconnaissance des situations à risques Dans ce chapitre, nous présenterons la détection de situations à risques sur nos données multimodales utilisant des réseaux de neurones profonds. Dans un premier temps nous décrirons les différentes architectures utilisées. Ensuite, nous évoquerons les expérimentations et l'optimisation de nos réseaux. Enfin, nous présenterons les résultats obtenus sur nos données BIRDS et nous les comparerons à celles de la littérature en se basant sur un dataset disponible en libre accès.

Enfin, le chapitre 8 conclut ce travail de recherche et propose quelques perspectives.

Chapitre 2

État de l'art

2.1 Introduction

Le vieillissement de la population mondiale est un des faits marquant du XXI^{ème} siècle. Ceci nous pousse à étudier et à comprendre le phénomène pour assurer un vieillissement dans les meilleures dispositions possibles. Pour cela, différentes technologies d'accompagnement permettant un vieillissement en bonne santé ont été étudiées.

Les approches visant à la détection automatique des situations à risques chez les personnes âgées remontent à la fin du siècle dernier, en commençant par les systèmes de détection de chutes [Williams et al. \(1998\)](#). Malgré la multiplication de systèmes de détection de situations à risque, apportant des solutions avec des architectures présentant différents capteurs, aucun système n'a pu s'imposer comme architecture de référence jusque-là.

Des revues de littérature présentent des méthodes de détection de situations à risques différentes [Cohen et al. \(2021\)](#); [Rucco et al. \(2018\)](#); [Bell et al. \(2020\)](#); [Singh et al. \(2020\)](#), certaines étant spécifiques aux capteurs ou d'autres plus axées sur les systèmes d'analyse de données ou encore sur les situations à risque. Dans ces revues de littérature, on parle des différents systèmes de détection en les regroupant sous trois catégories : les dispositifs portables, les capteurs ambiants et les systèmes basés sur des caméras portées.

Dans ce chapitre, nous parlerons d'abord du vieillissement et nous définirons le concept de fragilité. Dans un second temps nous présenterons les systèmes IoT en mettant l'accent sur les dispositifs IoT destinés aux personnes âgées. Ensuite nous ferons état des méthodes de détection de situations à risques destinées aux personnes âgées en général. Enfin, nous présenterons les principaux concepts pour comprendre les réseaux de neurones artificiels et les réseaux de neurones profonds sur des tâches de classification des signaux et des images pour la détection des situations à risques sur des données eHealthcare.

2.2 Le vieillissement et la fragilité

Dans cette section, nous aborderons le phénomène du vieillissement, de ces répercussions économiques et sociales mais aussi le concept de fragilité.

2.2.1 Le vieillissement

Ces dernières années, nous avons assisté à une amélioration substantielle des conditions sociales et sanitaires, avec une augmentation significative de l'espérance de vie. De même, une augmentation significative du nombre de "personnes âgées" a été constatée. Les personnes âgées de plus de 65 ans sont au nombre de 702,9 millions en 2019 dans le monde. Ce nombre devrait atteindre 1548,9 millions en 2050, soit une augmentation de 120 %. Pour les personnes âgées de plus de 80 ans, elles sont au nombre de 53,9 millions en 2019 et devraient atteindre 109,1 millions en 2050, soit une augmentation de 102,6 %. En France, une étude statistique réalisée par INSEE montre que près d'une personne sur trois aura plus de 60 ans en 2050, contre 21.5 % en 2007 [Blanpain et Chardon \(2010b\)](#) (figure 2.1). Le vieillissement de la population est incontestablement un succès démographique, porté par de nombreux changements dans la société. Mais si l'espérance de vie continue d'augmenter, la durée de vie en bonne santé, définie comme le nombre d'années de vie vécues en bonne santé, quant à elle, stagne depuis des années, [Pérès et al. \(2013\)](#) suggérant une compression de la morbidité et un recul de l'âge d'entrée en dépendance et dans la vieillesse. Cependant, de récents travaux aux États-Unis [Seeman et al. \(2010\)](#) mais également en France [Cambois et al. \(2013\)](#) suggèrent que cette tendance favorable ne se confirmerait pas pour les générations à venir et pour la première fois est envisagée la possibilité que la génération actuelle pourrait vivre moins bien que la précédente. En effet, si d'une part, la population a gagné en longévité, d'autre part, le vieillissement a favorisé une forte prévalence progressive des pathologies chroniques liées à l'âge et donc des années vécues en incapacité [McPhail \(2016\)](#). Des maladies telles que le diabète de type 2, l'insuffisance cardiaque chronique, les maladies cardiovasculaires et la démence sont fréquentes chez les personnes âgées. Ces maladies chroniques nécessitant un traitement et un suivi régulier, représentent un poids important pour le système de santé. Cela est particulièrement évident dans la prise en charge des personnes atteintes de démence, une maladie qui est devenue une charge croissante pour la société, la famille et les sujets âgés eux-mêmes. Le modèle de l'individu malade qui caractérise le plus la société moderne n'est pas tant l'individu atteint d'une maladie spécifique, aiguë et curable à court ou moyen terme, mais plutôt un patient chronique et complexe, souffrant le plus souvent de différentes maladies en même temps. La présence simultanée de plusieurs maladies est ce qui caractérise réellement la gériatrie aujourd'hui. Cela conduit également à la polypharmacie et à un risque majeur chez les sujets âgés de subir une réaction indésirable aux médicaments. Cette condition peut être associée à la fois au risque d'interventions fragmentées, davantage axées sur le traitement d'une seule maladie.

2. État de l'art

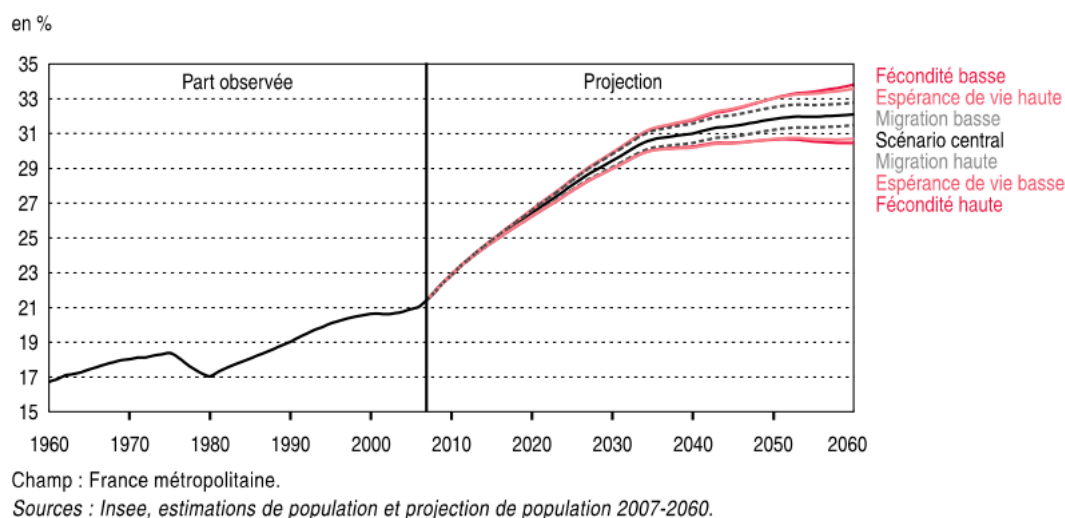


FIGURE 2.1 – Évolution des personnes âgées de 60 ans et plus en France de 1960 à 2060. Blanpain et Chardon (2010a).

Pour relever le défi du vieillissement de la population et répondre aux besoins des personnes âgées, il faut comprendre au mieux le vieillissement, la fragilité, le handicap, ainsi que des interventions préventives appropriées. Il s'agit notamment de la gestion des patients souffrant de fragilité, en leur assurant des dispositifs d'accompagnement dans leur vie quotidienne. Il faut donc mettre l'accent sur la prévention des complications liées à la fragilité et sur la promotion d'un vieillissement sain et actif afin de retarder l'apparition du handicap et de la dépendance.

2.2.2 Du vieillissement à la fragilité

Le vieillissement, qu'il soit "normal" ou "pathologique", est un processus qui modifie l'état de santé d'un individu. Il aboutit à l'échelle d'une population à la formation d'un groupe hétérogène incluant trois catégories spécifiques, à savoir, les personnes âgées robustes, les personnes âgées malades et/ou dépendantes, et une catégorie "intermédiaire", les personnes âgées fragiles. Dans la prise en charge des patients âgés, la fragilité est apparue comme l'un des syndromes cliniques les plus pertinents, qui va augmenter de façon exponentielle avec le vieillissement de la population.

Bien qu'il n'existe pas de définition stricte, il est clair que la fragilité décrit un état de santé particulier lié au vieillissement, dans lequel plusieurs systèmes de l'organisme perdent progressivement leurs réserves physiologiques intégrées. La définition sur laquelle s'accorde la communauté scientifique est que la fragilité est un syndrome gériatrique qui se caractérise par une diminution de la force, une faible activité physique et une perte de poids associée à des complications de santé. La fragilité se manifeste par une vulnérabilité biologique (liée à l'âge) aux facteurs de

stress et une diminution des réserves physiologiques. Les cinq critères de fragilité reconnus et couramment utilisés pour l'identification [Fedarko \(2011\)](#) sont les suivants : sentiment d'épuisement, ralentissement de la vitesse de marche, faiblesse musculaire, perte de poids involontaire (10 livres au cours de la dernière année) et une faible activité physique. Ces signes seraient la résultante de l'affaiblissement de plusieurs systèmes organiques.

La fragilité suscite un intérêt croissant en raison de capacité à prédire des événements de santé négatifs, tels que la dépendance, l'institutionnalisation, une diminution de la qualité de vie et la mortalité [Buckinx et al. \(2016\)](#). Pour définir le vieillissement en bonne santé, l'OMS a introduit la notion de "capacité intrinsèque" pour faire référence aux besoins en soins intégrés des personnes âgées. Cette notion inclut les manifestations cliniques communes du déclin des capacités physiques et mentales, qui sont de puissants prédicteurs de la mortalité et de la dépendance à l'égard des soins à un âge avancé [Beard et al. \(2016\)](#). La stratégie générale consiste à promouvoir ces capacités et à prévenir la fragilité au cours du vieillissement. Pour les personnes fragiles, il est impératif d'identifier les moyens les plus efficaces de restaurer les fonctions. Il a été démontré que l'inactivité, en particulier, contribue de manière importante à l'une des principales composantes de la fragilité, à savoir la perte de masse musculaire (appelée sarcopénie) et le déclin fonctionnel [Lazarus et al. \(2018\)](#). La figure 2.2 illustre trois sous-populations de personnes âgées : celles dont les capacités sont relativement élevées et stables, celles dont les capacités diminuent et celles qui subissent des pertes substantielles de capacités. Si l'on répond aux besoins de ces sous-groupes de population, la capacité fonctionnelle de la plupart des personnes âgées pourraient se voir améliorée, c'est ce qui constitue le défi de nos sociétés.

Le large éventail de situations à risques rencontrées par les personnes fragiles a été identifié dans [Avila-Funes et al. \(2008\)](#), [Vemulapalli et al. \(2017\)](#). D'après ces études, les accidents les plus fréquents chez les personnes âgées sont les chutes. Il s'agit de la troisième source de mortalité la plus fréquente après la crise cardiaque et le cancer chez les personnes âgées. Néanmoins, les risques encourus par les personnes âgées diffèrent en fonction de leurs antécédents médicaux. En dehors des cas de chutes des personnes âgées, il existe d'autres situations plus spécifiques comme dans le cas des personnes souffrant de maladie de Parkinson dus au gel de la marche qui précipite la chute, les risques spécifiques des personnes atteintes de la maladie d'Alzheimer en plus des diabétiques, les risques pour les personnes souffrant d'insuffisance cardiaque ou respiratoire et d'autres maladies chroniques. Par exemple, pour les personnes atteintes de la maladie de Parkinson, les risques les plus urgents sont les chutes liées aux troubles de la marche, qui sont très fréquentes, pour les personnes atteintes de la maladie d'Alzheimer, il y a les situations de stress, la perte d'orientation et la perte de direction. Pour les personnes diabétiques, les hyperglycémies et les hypoglycémies sont dangereuses. Pour les personnes âgées sans maladie particulière, les situations à risques sont souvent liées au contexte et à leurs conditions physiologiques, par exemple une personne en état hypotonique peut trébucher dans la salle de bain ou

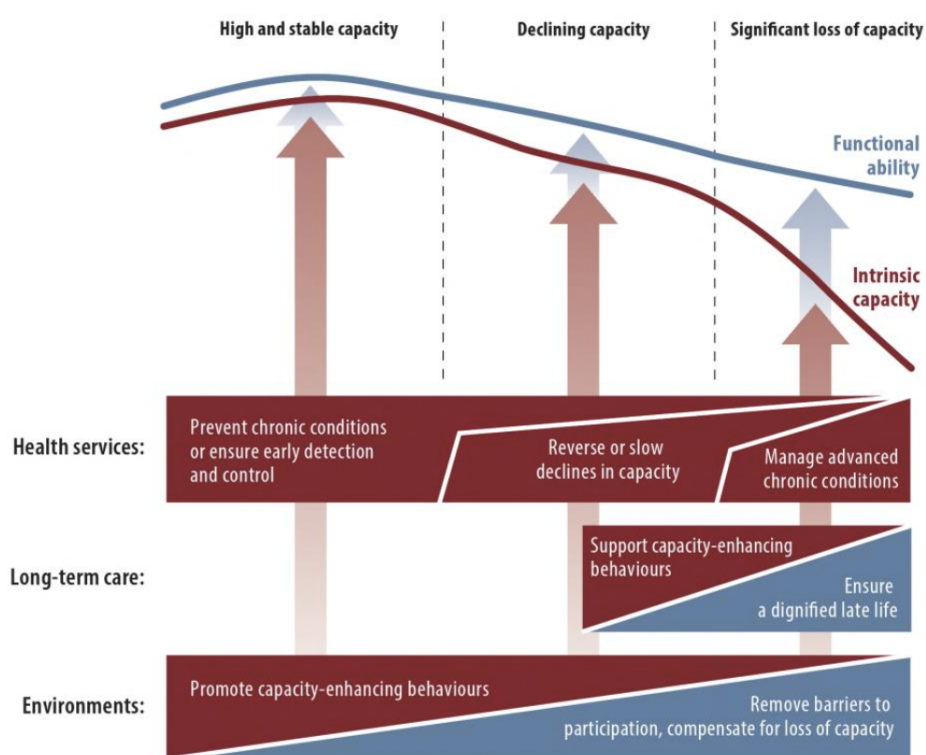


FIGURE 2.2 – Un cadre de santé publique pour un vieillissement sain [Organization \(2015\)](#).

dans la cuisine et donc avoir un risque de chute, ou étant dans un état de stress et cuisinant elle peut oublier le feu sur la cuisinière, c'est-à-dire avoir un risque d'accident domestique.

2.3 Les objets connectés pour la surveillance des personnes âgées

Les personnes âgées de 65 ans et plus sont victimes chaque année en France de 550 000 Accident de vie courante (AcVC) avec recours aux urgences. Elles contribuent pour plus des trois quarts aux 20 000 décès annuels [Ricard et Thélot \(2008\)](#). Le nombre d'accidents croit chez les personnes fragiles.

Différents risques touchant la population fragile ont été identifiés dans la littérature. Nous pouvons distinguer les chutes, les risque de chute dus à l'état fragile qui sont très fréquentes chez la population fragile et donc bien étudiées. Les autres risques peuvent être : les accidents de la route, la fatigue, la déshydratation, les facteurs environnementaux (par exemple, la température ambiante, la surexposition au soleil), les mauvais traitements, les abus, les crimes et le suicide et les risques d'accidents domestiques. Ces risques (et d'autres) sont associés à [Yared et al. \(2015\)](#) : (1) un coût financier lourd pour les collectivités ainsi que pour l'individu (2) un risque d'institutionnalisation (3) impliquent des blessures physiques et psychologiques aboutissant à un état de dépendance.

Ces risques et leurs conséquences concernent largement les personnes âgées fragiles. C'est pourquoi la recherche et l'industrie proposent diverses solutions pratiques pour détecter, prévenir et aider les situations à risque, et pour en atténuer les conséquences.

Dans cette section, nous évoquerons les différentes situations à risques étudiées dans la littérature, nous parlerons aussi des études réalisées pour chacune de ces situations à risques. Nous présenterons les différentes architectures de système d'aide aux personnes âgées, aussi que les méthodes de classifications proposées.

2.3.1 Les systèmes de détection de chute

La chute est définie généralement comme l'action de tomber, c'est-à-dire de perdre son équilibre et d'être attiré vers le sol par son poids sous l'effet de la force gravitationnelle. La définition la plus fréquemment citée dans la littérature est celle de l'OMS qui définit la chute comme l'action de tomber ou de glisser au sol indépendamment de sa volonté [OMS](#). A cause de la dangerosité, mais aussi du coup important que peut constituer la chute. Un nombre important de recherches sur les systèmes d'aide aux personnes âgées se concentre sur la détection de la chute. Les approches visant à la détection automatique des situations de chutes chez les personnes âgées remontent à la fin du siècle dernier [Williams et al. \(1998\)](#). Malgré les progrès de l'apprentissage automatique au cours de ces dernières années, aucun

algorithme automatique de détection de chute n'a pu s'imposer comme un standard de l'industrie, ni dans la recherche. Cela s'explique notamment par le nombre élevé des systèmes avec de fausses alarmes (positives ou/et négatives). Jusqu'à présent, de nombreuses méthodes ont été proposées pour automatiser le processus de détection de chute. Celles-ci peuvent être divisées en deux catégories :

- Détection de chute par capteurs portés ;
- Détection de chute capteurs ambiants.

Détection chute par capteurs portés

Comme de nombreuses chutes se caractérisent par une accélération du mouvement du corps humain sur l'axe vertical, l'approche la plus simple consiste à fixer un seuil préalablement déterminé au-delà duquel toute valeur supérieure est considérée comme un moment de chute. La définition d'un seuil de mouvement entre la chute et la non-chute associée à un capteur d'accélération porté est l'un des premiers dispositifs de détection de chute mis en place. Le seuil est fixé par exemple à $1.7g$ ($g = ma$) pour l'accélération, les valeurs au-dessus représentent une chute et à $0g$ la personne est allongée [Cao et al. \(2016\)](#).

Même si la plupart des chutes se produisent dans le plan "frontal" (vers l'avant ou vers l'arrière), la direction de la trajectoire de chute est évidemment variable d'une chute à une autre.

Pendant une chute, il y a une période temporaire de "chute libre", pendant laquelle la vitesse verticale diminue puis augmente linéairement avec le temps en raison de l'accélération gravitationnelle. En mesurant la vitesse verticale des mouvements contrôlés de la personne (se lever, se pencher, s'asseoir), nous pouvons distinguer ces vitesses de celles qui se produisent lors d'une chute, ce qui dépasserait un seuil fixe approprié.

La difficulté de l'approche par seuillage réside dans le choix du seuil, si celui-ci est trop bas, l'appareil détectera des événements de non-chute ("faux positifs"), mais si le seuil est trop élevé, il ne détectera pas les événements de chute ("faux négatifs"). Ce seuil dépend également de l'état de santé et de l'âge du sujet ainsi que de la variabilité de la chute (se heurter à un objet, perte de conscience, tomber sur un lit, faux pas, etc).

Plusieurs méthodes pour la définition d'un seuil fixe pour la détection de chutes ont été proposées. La plupart d'entre elles utilisent des données de capteurs de mouvement. Ci-après, nous présenterons deux méthodes fondatrices de la détection de chutes par seuillage ainsi que leurs limites.

Les travaux de [Bourke et Lyons \(2008a\)](#) ont été parmi les premiers traitant de la détection de chutes à l'aide d'un seuil fixé. Les données expérimentales ont examiné des exemples de divers types de chutes et de non chutes. D'abord par des simulations sur 10 jeunes personnes à l'aide de deux accéléromètres tri-axiaux placés sur leur poitrine et leur hanche. Ensuite dans des conditions écologiques sur des personnes âgées de 77.4 ± 4.3 ans, ils ont enregistré des situations de vie courante (s'asseoir,

s'allonger sur un lit, monter et descendre les escaliers) avec les deux accéléromètres cités précédemment.

Le traitement des données a consisté en un calcul d'un seuil minimal Lower Peak Value (LPV) (Lower Peak Value) et un seuil maximal Upper Peak Value (UPV) (Upper Peak Value) pour la chute. À partir de là, ils ont comparé chaque valeur du signal mesuré avec la LPV puis à la UPV. La LPV détecte la pré-chute et UPV confirme la chute.

Les résultats montrent plusieurs fausses détections (faux positif et faux négatifs), des faiblesses sur des données écologiques et le seuil fixé ne correspond pas à un seuil de détection de chute pour les personnes âgées. Néanmoins la méthode présente un coût calculatoire très faible ce qui est un réel avantage pour des dispositifs à faible capacité de traitement.

[Bourke et Lyons \(2008b\)](#) se sont intéressés à la détection de chute à l'aide de mesures de gyroscopes. Ils ont appliqué leur algorithme de détection sur des données d'une cohorte composée de dix jeunes individus sains qui ont porté un gyroscope au niveau du plexus solaire, ils ont simulé des chutes ainsi que des situations de vie courantes pouvant s'apparenter à des chutes. Le traitement des données pour la détection de chutes consistait en une comparaison des mesures du gyroscope (vitesse angulaire), de sa dérivée (accélération angulaire) ainsi que de son intégrale (position angulaire) avec des seuils fixés FT1, FT2 et FT3. L'hypothèse conduite pour établir les valeurs des seuils FT1, FT2 et FT3 est que les valeurs minimales des pics représentant des chutes correspondent aux seuils. La vitesse angulaire du sujet est calculée par la racine carrée de la somme des carrés des deux vitesses. L'article montre que la méthode obtient de bons résultats sur la détection de chute, nous pouvons également observer que son coût calculatoire reste relativement faible. Toutefois, les valeurs des seuils fixés en ne tenant compte que des chutes de sujets jeunes dans des conditions restreintes, sont peu représentatives des situations de chutes des personnes âgées.

Les méthodes de détection et de prédiction de chutes par apprentissage automatique sont très nombreuses. Nous remarquons toutefois que l'accent est mis sur l'apprentissage supervisé dans la majorité des cas.

Il est possible de retrouver dans la détection et la prédiction de chutes par capteurs portés (capteurs de mouvement et/ou capteurs physiologiques) :

- Des approches statistiques telles que la classification naïve bayésienne [Howcroft et al. \(2017\)](#), les Hidden Markov Model (HMM) (Modèle de Markov caché) [Tong et al. \(2013\)](#) ;
- Des approches par machine learning incluant la classification par k-Nearest Neighbor (kNN) [Rosero-Montalvo et al. \(2017\)](#), la classification par Support Vector Machine (SVM) [Wang et al. \(2015\)](#) et les réseaux de neurones profonds tels que les CNN [Fakhrulddin et al. \(2017\)](#) et Convolutional Long Short-Term Memory (ConvLSTM) [Nait Aicha et al. \(2018a\)](#).

[Howcroft et al. \(2017\)](#) ont présenté une analyse complète de la prédiction du risque de chute en utilisant deux types de capteurs portables (accéléromètres, se-

nelles intérieures à détection de pression), à quatre emplacements différents pour d'accéléromètre (tête, bassin, jambe gauche et droite). Les données obtenues ont été testées sur trois approches machine learning (réseau neuronal, SVM, classification bayésienne naïve). Leurs travaux ont conclu que la fusion de données fournissait les meilleures données d'entrée pour la prédiction du risque de chute. Ils ont obtenu les meilleurs résultats de leur expérience en utilisant une combinaison d'accéléromètres sur le bassin, la tête et la jambe gauche et un réseau neuronal.

[Tong et al. \(2013\)](#), ont proposé une méthode basée sur un modèle de Markov caché HMM utilisant un accéléromètre triaxial pour prédire les chutes. L'algorithme arrivait à prédire les chutes accidentelles 400 ms avant l'impact. Cependant, l'algorithme n'a pas été testé sur des situations écologiques.

[Nait Aicha et al. \(2018a\)](#), proposent l'utilisation des informations biomécaniques et de la fusion de données pour accroître l'efficacité des méthodes de machine learning à apprentissage profond dans la détection de risques de chutes. Ils ont utilisé des données existantes de 296 personnes âgées pour comparer la performance de trois architectures de modèles d'apprentissage profond CNN, LSTM et une combinaison de ces deux approches ConvLSTM. Les résultats montrent que les modèles ConvLSTM évaluent efficacement le risque de chute sur la base de données de capteurs portables et d'informations biomécaniques.

[Ozcan et al. \(2013a\)](#) présentent un système de détection des chutes et de classification des activités à l'aide de caméras portables. La caméra est portée sur le sujet, ce qui améliore la mobilité du dispositif contrairement à d'autres systèmes de vision qui opèrent dans des zones confinées. De plus, comme les images capturées ne concernent pas le sujet, les préoccupations relatives à la protection de la vie privée sont atténuées. L'algorithme de détection de chute proposé utilise les histogrammes des orientations et des contours ainsi que le calcul de l'optical flow pour la classification des activités. Les tests ont été effectués sur des données pré-enregistrées de simulation de chutes. Les résultats présentés montrent le succès de la méthode. Toutefois, le champ de test est limité à des données simulées.

Détection de chute par capteurs ambiants

Les capteurs ambiants ont été introduits et sont vite devenus les supports de détection de chute les plus utilisés dans les espaces communs tels que les hôpitaux ou les maisons de repos [Rougier et al. \(2011\)](#); [Ozcan et al. \(2013b\)](#). Les capteurs ambiants utilisés pour la détection de chutes sont très nombreux. Nous pouvons citer la caméra-vidéo, RGB-Depth caméra, le radar, etc.

Bien que les systèmes à caméra vidéo fournissent des informations détaillées sur le mouvement et la position du sujet et donc le risque de chute, ils souffrent de plusieurs inconvénients tels que l'intrusivité du capteur, les coûts souvent plus élevés que pour les autres capteurs et l'acceptation par les utilisateurs. Un autre inconvénient majeur de ces systèmes est leur incapacité à suivre les personnes en dehors de la portée de visibilité de la caméra.

Dans les systèmes basés sur la vision, les caméras 2D sont les capteurs les plus utilisés dans la détection de chutes. Les approches basées sur la vision se concentrent sur l'exécution en temps réel de l'algorithme en utilisant des plates-formes informatiques standard et des caméras à bas prix. Plusieurs méthodes sont utilisées pour la classification des risques de chutes mais beaucoup de travaux de recherches s'appuient ces dernières années sur des méthodes Deep learning. Ceci est dû à l'efficacité et la précision de ces méthodes pour l'analyse des données vidéos.

Les travaux de recherche réalisés par [Li et al. \(2017\)](#) décrivent l'utilisation des réseaux neuronaux convolutifs pour la détection des chutes dans un environnement de vidéosurveillance. Le CNN est directement appliqué à chaque image de la vidéo pour apprendre les caractéristiques de déformation de la forme humaine qui décrivent différentes postures de l'humain et déterminent si une chute se produit. Les résultats expérimentaux montrent que notre approche proposée fonctionne en temps réel et atteint une précision moyenne de 99,98 % pour une validation croisée de 10 fois pour la détection des chutes. Ces résultats sont prometteurs même s'ils ne sont pas testés sur des cas de chutes réelles, la méthode est intéressante.

La détection de chutes par mesures radar a été introduite en 2011 par [Liu et al. \(2011\)](#), et leur utilisation n'a pas cessé de croître dans les travaux de recherche. La détection des risques de chutes se fait par le calcul de l'accélération des données du radar et des algorithmes de classification pour éviter les fausses alarmes. Les travaux de [Liu et al. \(2011\)](#) ont été l'un des premiers exemples publiés sur l'utilisation d'un système radar commercial à impulsions Doppler à 5,8 GHz pour la détection des chutes. Les données expérimentales ont traité des exemples de divers types de chutes accidentelles et de mouvements de vie courante qui ont été effectués par huit sujets. Le traitement des données a consisté en un calcul des spectrogrammes, suivi d'une étape de présélection pour identifier l'événement de chute en examinant les courbes d'accélération. Un classificateur kNN simple avec 3 voisins et un classificateur SVM ont été utilisés. Les résultats obtenus avec les deux classificateurs sont très prometteurs pour la méthode. Toutefois, les données expérimentales utilisées ne sont pas assez représentatives des situations de chutes des personnes âgées qui sont la cible directe de ce genre de dispositif.

[Jokanovic et al. \(2016a\)](#) et [Jokanovic et al. \(2016b\)](#) ont introduit l'utilisation d'algorithmes d'apprentissage profond et de réseaux neuronaux convolutifs (CNN) pour la détection des chutes sur des données radar. L'intérêt de cette approche est que l'étape d'extraction des caractéristiques est contournée et que les spectrogrammes et les étiquettes de classe associées sont directement fournis en entrée au CNN, après une étape de pré-traitement pour réduire le bruit et appliquer une échelle de gris aux images. Les CNN impliquent des coûts de calcul plus élevés pour être formés, mais la tendance technologique d'augmenter la puissance de calcul à moindre coût est susceptible de rendre cette approche de plus en plus courante. Le document a fourni une validation préliminaire de l'idée, montrant de bons résultats de classification pour quatre mouvements et des résultats améliorés par rapport à la méthode

conventionnelle d'utilisation des caractéristiques dans le classificateur [Rivera et al. \(2014\)](#) et le classificateur SVM.

La disponibilité sur le marché de capteurs caméra RGB-Depth peu coûteux a encouragé à les appliquer à la détection de chutes. L'information sur la profondeur, a apporté de nombreux avantages dans le développement de solutions basées sur la vision. Les données de profondeur fournissent des informations 3D qui peuvent apporter plus d'information sur la position du sujet.

[Bian et al. \(2015\)](#) ont proposé une méthode pour extraire le mouvement des articulations utiles à partir de cartes de profondeur. Plus précisément, un algorithme d'arbre de décision aléatoire peut extraire les articulations de la tête et de la hanche. La méthode de classification considérée pour la détection des événements de chutes/non chutes est SVM. La performance a été évaluée par des tests de classification sur 4 catégories de chutes et un ensemble de mesures liées aux situations de vie courantes sont évaluées. Les tests sont faits sur un nombre total de 380 échantillons classés avec une précision de 97,6 %. Ces résultats sont prometteurs pour la RGB-D caméra et le faible coût calculatoire de la méthode est prometteur pour son application dans l'industrie. Cependant, on peut reprocher à la méthode l'utilisation de données de personnes jeunes et en bonne santé pour l'apprentissage ainsi que l'immobilité du capteur.

[Gasparrini et al. \(2014\)](#) ont proposé un système où le capteur caméra RGB-D placé au plafond de la pièce du lieu de vie, pour fournir une vue d'ensemble de la scène. La phase de pré-traitement et de segmentation est exploitée pour extraire la carte de profondeur du premier plan à partir d'un cadre de fond statique et pour segmenter les objets sur la scène. Certaines caractéristiques extraites des formes d'objets permettent de détecter la présence d'une personne, et il est possible de suivre les mouvements et de détecter une chute en tenant compte de la distance par rapport au sol. L'algorithme a été évalué sur un ensemble de données enregistrées auprès de 4 acteurs effectuant 20 tests, dont 10 impliquent la présence de plusieurs sujets dans la zone. La même configuration de capteurs a été utilisée par [Liciotti et al. \(2015\)](#). Un algorithme de soustraction d'arrière-plan basé sur un mélange de gaussiennes renvoie une image d'avant-plan qui contient les personnes et les objets en mouvement. L'algorithme de segmentation exploite un ensemble de paramètres pour distinguer les personnes des objets, à savoir la hauteur de chaque personne, la taille de chaque tête et la distance tête-épaules. Les chutes sont détectées lorsque la valeur de profondeur de la tête est proche du niveau du sol. D'autres méthodes basées sur les caméras RGB-Depth, exploitant une combinaison de données de chutes réelles et simulées ont été proposées [Liciotti et al. \(2015\)](#).

Détection de chute par fusion de données

La fusion de capteurs implique l'utilisation de signaux multiples et différents provenant de divers appareils de mesures incluant des capteurs visuels, inertiels ou bien physiologiques [2.3](#). La combinaison des mesures issues de différents capteurs

aide au développement d'algorithmes plus robustes et plus fiables grâce à la complémentarité et/ou à la redondance de l'information notamment pour les technologies DL qui peuvent être utilisés pour surveiller la santé des sujets et détecter les chutes [Gutiérrez et al. \(2021\)](#).

[Li et al. \(2018\)](#) ont combiné les données d'accéléromètre des smartphones et les données de caméra de profondeur de Kinect, ainsi que les signaux de la caméra du smartphone.

Des recherches se basent aussi sur la fusion partielle, bien que plusieurs capteurs soient déployés, un seul capteur est utilisé pour prendre la décision finale, comme les travaux de [Ma et al. \(2019\)](#). Ils ont utilisé une caméra RGB et une caméra thermique pour mener leurs expériences, la caméra thermique étant utilisée uniquement pour la localisation des visages. Les chutes ont finalement été détectées uniquement sur la base des données collectées par les caméras RGB ordinaires.

La deuxième approche de la fusion de capteurs est connue sous le nom de fusion de caractéristiques. Dans cette approche, l'extraction de caractéristiques se fait sur des signaux provenant de différents capteurs. Toutes les caractéristiques sont ensuite fusionnées en longs vecteurs de caractéristiques et utilisées pour entraîner des modèles de classification. Dans de nombreux produits commerciaux de dispositifs portables, des capteurs tels que des accéléromètres, des gyroscopes et des magnétomètres sont intégrés dans un seul dispositif. Les données de ces capteurs sont synchronisées de manière homogène avec la même fréquence et transmises avec des modules sans fil intégrés [Ren et Peng \(2019\)](#).

La fusion de décision est la troisième approche, où une chaîne de classificateurs est utilisée pour arriver à une décision. Un arrangement typique est d'avoir un modèle de classification qui prend l'entrée d'un type de capteur, un autre modèle qui prend l'entrée d'un autre capteur, et à son tour les sorties de ces deux modèles sont utilisées comme entrée à un troisième modèle de classification qui prend la décision finale.

La dernière approche est la fusion de données. Pour ce faire, il faut d'abord fusionner les données de différents capteurs, puis procéder à l'extraction de caractéristiques à partir des données fusionnées. Cette méthode s'oppose à la fusion de caractéristiques où les données de ces capteurs sont homogènes et synchrones avec la même fréquence. La fusion de données peut être appliquée à différents capteurs ayant une fréquence d'échantillonnage et des caractéristiques de données différentes. Les données provenant de divers capteurs peuvent être synchronisées et combinées directement pour certains capteurs de types différents.

[Queralta et al. \(2019\)](#) ont appliqué un réseau de neurones récurrent un réseau à mémoire à long court terme (LSTM), dans laquelle des capteurs portables comprenant un accéléromètre, un gyroscope et un magnétomètre ont été intégrés. Leur système a atteint un taux de rappel moyen de 95 % tout en fournissant une solution en temps réel de détection des chutes fonctionnant sur des plateformes en nuage.

[Delgado-Escano et al. \(2020\)](#) présente un modèle basé sur l'apprentissage profond indépendant des jeux de données qui, en employant une approche d'apprentissage multi-tâches, utilise les informations inertielles brutes en entrée pour résoudre si-

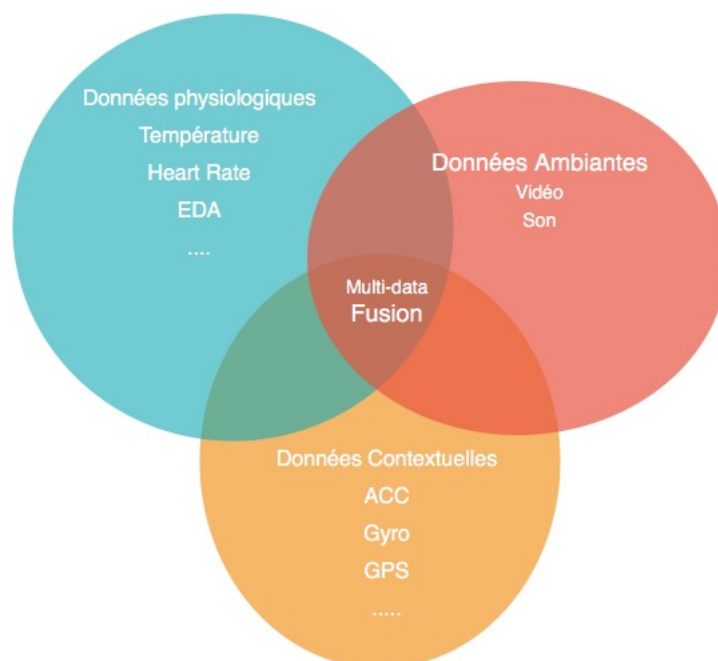


FIGURE 2.3 – Différents types de capteurs et le principe de fusion de capteurs.

multanément deux tâches : la détection de chute et l'identification du sujet. Les scores de détection de chute ont une précision de plus de 98 % et le modèle permet d'identifier correctement les personnes avec une précision moyenne de 79,6 %.

2.3.2 Les systèmes de détection de déshydratation

La déshydratation est un problème courant chez les personnes âgées entraînant de sérieuses complications sur le plan de la santé. La déshydratation peut causer un épisode confusionnel, et en cas extrême, des convulsions ou même la mort. Elle peut se présenter sous trois formes : hypertonique (perte d'eau, causée par un apport insuffisant, la transpiration/l'évaporation ou les vomissements), isotonique (perte proportionnelle de sel et d'eau, souvent causée par la diarrhée) et hypotonique (perte de sel, souvent causée par les diurétiques) [Bennett \(2000\)](#). Les personnes âgées présentent un risque plus élevé de déshydratation hypertonique pour une multitude de raisons entourant la réduction de l'apport en liquide et l'excès d'élimination. La sensation de soif diminue avec l'âge [Bennett \(2000\)](#). Les médicaments tels que les diurétiques, qui sont couramment prescrits en cas d'insuffisance cardiaque et de certains troubles rénaux, peuvent provoquer une perte de liquide et de sel [Rowat et al. \(2012\)](#). En outre, certains patients âgés atteints de démence oublient de boire. Les patients âgés peuvent initialement avoir besoin d'un apport supplémentaire en liquide pour éviter la déshydratation, mais les comorbidités et le déclin des fonctions organiques lié à l'âge peuvent accroître la vulnérabilité à la surhydratation notam-

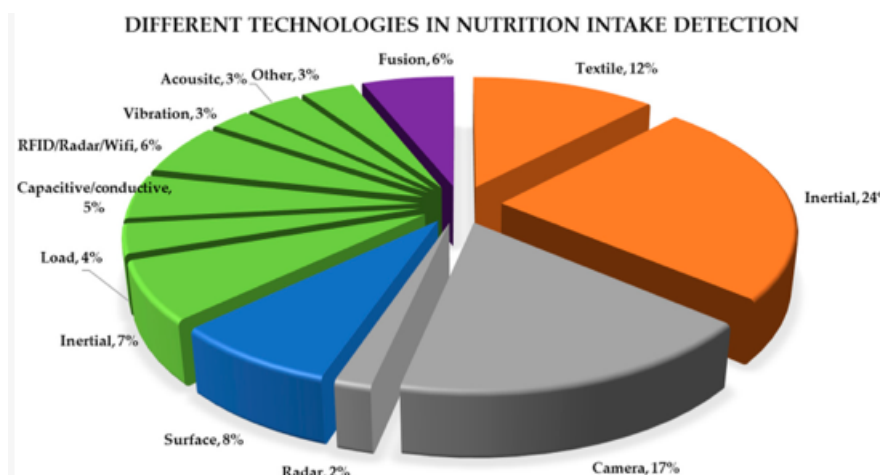


FIGURE 2.4 – Répartition des approches de détection de l’absorption de liquides et de nourriture en fonction de la technologie utilisée dans l’état de l’art. L’orange représente les dispositifs portables, le violet la fusion, le vert les conteneurs intelligents, le bleu les surfaces avec capteurs intégrés et le gris les approches basées sur la vision et l’environnement [Cohen et al. \(2021\)](#).

ment chez les personnes fragiles [Aubry et al. \(2018\)](#). On parle de surhydratation lorsque l’organisme a trop de liquide en cas de rétention d’eau par l’organisme et/ou de difficulté à éliminer l’excès d’eau. La méthode la plus utilisée dans le milieu médical pour surveiller le niveau d’hydratation est l’osmolarité du sérum/plasma sanguin. Cette méthode souvent considérée comme invasive malgré sa popularité est contestée aujourd’hui. Plusieurs systèmes de surveillance utilisant des technologies différentes permettent aujourd’hui de surveiller l’apport hydrique oral au quotidien. L’état de l’art sur les technologies de surveillance de déshydratation est assez large. Pour mieux synthétiser les différentes technologies, nous avons décidé de les regrouper en tenant compte de la nature de capteurs utilisés. Nous évoquerons les systèmes s’appuyant sur des technologies portables, des technologies environnementales et les systèmes fusionnant différents types de technologies. La figure 2.4 présente la répartition des approches de surveillance de l’absorption de liquides en fonction de la technologie utilisée.

Détection du risque de déshydratation avec les technologies portables

Des dispositifs portables sont proposées pour détecter l’ingestion de nourriture ou de liquide, notamment les mesures inertielles du poignet ou du corps ou les mesures basées sur le textile.

Plusieurs articles de recherche ont utilisé des unités de mesure inertielle (IMU) sur le poignet afin de mesurer l’ingestion de liquide [Cohen et al. \(2021\)](#). Les capteurs inertiels montés sur le poignet sont souvent utilisés en combinaison avec des algorithmes d’apprentissage machine (ML) pour détecter les moments de consommation.

Chun et al. (2019) propose une technique de segmentation adaptative sur un flux continu de données inertielles capturées avec standard monté sur le poignet pour détecter passivement les gestes de consommation de liquide. L'approche a été évaluée dans le cadre d'une étude avec 30 participants où 561 actions de consommation ont été enregistrées. En utilisant une méthode LOPO (leave-one-participant-out), ils ont pu détecter les épisodes de consommation d'alcool avec une précision de 90,3 %. Ordóñez et Roggen (2016) ont présenté des méthodes basées sur des réseaux d'apprentissage profonds appliqués sur deux ensembles de données existants où les sujets disposaient de 7 IMU et de 12 accéléromètres (placés sur les bras, le tronc, les jambes et les pieds) ou de 10 accéléromètres sur chaque bras pour classer les activités de vie quotidiennes, y compris la boisson. Ils ont obtenu un score F1 de 90 % en utilisant des couches convolutionnelles et LSTM à 8 couches pour classer "boire du café". Liaqat et al. (2020a), présentent un dispositif portable non invasif pour recueillir les données de conductance de la peau, les données collectées ont été traitées par différents algorithmes d'apprentissage automatique pour prédire le niveau d'hydratation du corps humain dans différentes postures corporelles. Les résultats expérimentaux comparatifs ont démontré que la forêt d'arbres décisionnels, avec une précision de 91,3 %, a obtenu de meilleures performances par rapport aux autres algorithmes d'apprentissage automatique pour prédire l'état d'hydratation du corps humain. Huang et al. (2020) ont testé plusieurs algorithmes d'apprentissage automatique. Un algorithme Adaboost binaire a été utilisé pour la reconnaissance des événements de consommation d'alcool avec une précision de 94,42 %, la RF a été utilisée pour le repérage des gestes (soulever, siroter, relâcher) avec une sensibilité de 90 % et une précision de 92 %, et enfin un algorithme SVM linéaire a été utilisé pour l'estimation de la consommation de liquide avec un Mean Absolute Percentage Error (MAPE) de 40,11 %.

Détection du risque de déshydratation avec les technologies environnementales

Les capteurs installés dans l'environnement sont une approche courante pour fournir aux systèmes basés sur l'environnement la capacité d'extraire les habitudes de consommation individuelles. Les approches basées sur la vision artificielle utilisent des caméras et des algorithmes de vision par ordinateur ou d'apprentissage profond pour détecter les activités de consommation d'alcool. De nombreuses études dans ce domaine ont utilisé Microsoft Kinect qui calcule la profondeur et prend des images RGB (rouge vert bleu). Chua et al. (2014) ont appliqué un algorithme de caractéristiques d'apprentissage profond sur des images pour détecter une main saisissant une tasse. Jayatilaka et Ranasinghe (2017) ont proposé une approche de surveillance de la consommation de liquides qui comprend un gobelet intelligent Radio frequency identification (RFID) capable de surveiller les épisodes de consommation d'alcool sur la base de modèles sous-jacents dans les variations de données RSSI et de phase. Les résultats prometteurs ont été obtenus pour la reconnaissance des épisodes de

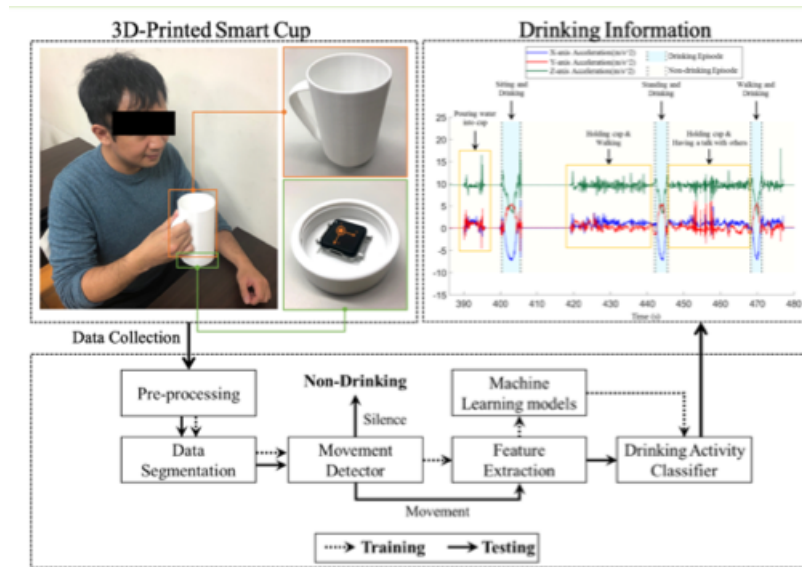


FIGURE 2.5 – Système de collecte et de traitement de données de détection d’une tasse intelligente, où sont annotés les épisodes de consommation (s’asseoir et boire, se tenir debout et boire et marcher et boire) et les épisodes de non-consommation (verser de l’eau dans la tasse, tenir la tasse et marcher et tenir la tasse et discuter avec d’autres personnes). Liu et al. (2020).

consommation d’alcool pour les données. Liu et al. (2020) présente un dispositif avec un accéléromètre intégré sur un gobelet intelligent imprimé en 3D (figure 2.5). Le système de surveillance de la consommation de liquides proposé, équipé d’un seul accéléromètre, reconnaît les événements et les épisodes de consommation d’alcool sur la base de modèles d’apprentissage automatique. Des expériences de validation de concept ont été réalisées sur onze jeunes participants en bonne santé afin de valider ce système dans un environnement de laboratoire. La meilleure performance du système utilisant le modèle kNN avec une taille de fenêtre de 4s atteint 89,92 % et 85,88 % de scores dans les approches d’évaluation définies par événement et par cadre. Les résultats montrent que le gobelet intelligent peut fournir un signal de détection fiable et faciliter l’identification des activités de consommation. Iosifidis et al. (2012) ont créé des volumes 3D à partir d’images de vidéos de consommation de nourriture et de boissons (figure 2.6), puis ont effectué une quantification vectorielle floue et une analyse Laser Doppler Anemometer (LDA). Dans les sous-espaces discriminants de faible dimension, les classes d’activité sont linéairement séparables, avec un taux de classification de 93,3 %.

Détection du risque de déshydratation avec la fusion de données issues de différentes technologies

Plusieurs chercheurs ont tenté de combiner les données de plusieurs sources de mesures afin d’obtenir des modèles de détection plus robustes et plus fiables. Plu-

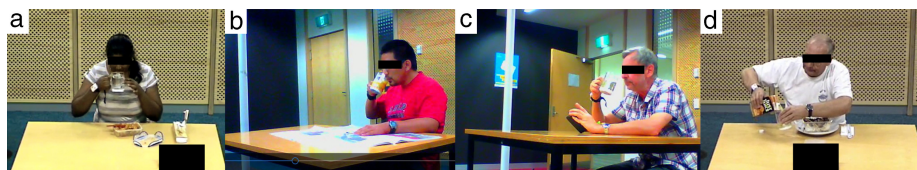


FIGURE 2.6 – (a) une jeune participante qui boit de l'eau tout en déjeunant ; (b) une jeune participante qui boit du jus d'orange tout en lisant le journal ; (c) un participant plus âgé qui boit du vin blanc froid tout en lisant un message au téléphone et en déjeunant ; et (d) un participant plus âgé qui se fait resservir du jus d'orange tout en dégustant son déjeuner Jayatilaka et Ranasinghe (2017).

sieurs études sur la détection du risque de déshydratation avec l'utilisation des données provenant de différents capteurs ont été menées. Cela concerne des données de capteurs portés ou environnementaux avec parfois la fusion de données de capteurs de ces deux catégories. Du et al. (2019) présente HydraDoctor qui est un système de surveillance de consommation de liquides en temps réel. Le dispositif comporte un accéléromètre de poignet commercial intégré dans une smartwatch pour détecter les événements de consommation d'alcool, puis déclenche des lunettes intelligentes (Google Glasses) pour enregistrer une vidéo. La vidéo est utilisée pour classer le type de récipient et de boisson à partir d'une base de données. Avec 11 sujets et 6 types de liquide, le système avait une précision de 85 % pour la détection des boissons et pouvait classer le type de liquide avec une précision de 84,3 % Du et al. (2019). Seiderer et al. (2017) ont combiné un pèse-personne, une smartwatch et un smartphone pour surveiller la consommation d'aliments et de boissons, la solution a été présentée sans rapporter de résultats d'évaluation sur la détection. Garcia-Constantino et al. (2020) ont fusionné les données d'un accéléromètre portable, de capteurs de contact placés dans la cuisine sur des ustensiles et les surfaces de contact. Cette étude a impliqué 30 participants qui ont tous effectué la même action : entrer dans la cuisine, préparer une boisson chaude, consommer la boisson et partir. Les actions de consommation ont été détectées avec une précision de 95 %.

Nombreuses sont les architectures proposées dans la littérature pour la détection de la prise de boissons, allant du type de capteurs au nombre de capteurs utilisés. Malgré l'inhomogénéité des systèmes de collecte, la plupart des méthodes semblent s'accorder sur l'utilisation des méthodes Machine Learning pour la détection de la prise de fluide. Quelques recherches s'intéressent à la détection en temps réel, les résultats obtenus sont prometteurs même s'il reste encore à améliorer la précision et à optimiser le système de collecte en termes énergétique. Des résultats sortants sur les différents systèmes de collecte, les technologies portables se basent souvent sur les technologies IoT, car alliant efficacité et praticité. Les capteurs d'intérêt dans ces technologies IoT sont les capteurs inertiels et plus spécifiquement les accéléromètres pour la mesure de l'accélération du bras. L'accélération du bras lors de la prise de boissons est assez différentiable des autres activités qu'on peut effectuer ce qui donne d'assez bons résultats de détection. D'autres mesures intéressent les scientifiques

pour la détection de la prise de boisson, nous pouvons citer l'activité électrodermale, la vidéo, etc. Souvent pour améliorer les résultats, la fusion de données est sollicitée. Une grande partie des techniques de fusion examinées fait appel à un accéléromètre de poignet et à une autre technologie. En combinant les données d'accélération avec d'autres mesures, les résultats sur les systèmes de détection semblent être plus prometteurs. Il s'agit d'un domaine qui peut être exploré davantage afin d'augmenter la précision des systèmes et de réduire les faux positifs.

2.3.3 Détection du risque de chute chez les personnes âgées

En plus de la détection de la chute, il est important de s'intéresser à la prévention de la chute. Ceci relève de la préservation de la santé physique et psychologique des personnes âgées. Chez les personnes âgées fragiles, cette prévention est réalisée par des exercices sportifs aidant au maintien de l'équilibre et du tonus musculaire. Les systèmes d'intervention de prévention des chutes se concentrent sur le suivi et le soutien des patients qui n'ont pas encore subi de chute, mais qui peuvent être considérés comme présentant un risque de chute. Les systèmes d'intervention de prévention des chutes visent à surmonter les facteurs intrinsèques de risque de chute tels que la vision, l'équilibre, la force musculaire et le déclin cognitif [Montero-Odasso et al. \(2012\)](#). D'autres facteurs peuvent provoquer la chute en plus des facteurs environnementaux, les facteurs physiologiques accroissent le risque de chutes notamment chez les sujets atteints de démence [Taylor et al. \(2014\)](#). La figure 2.8 illustre la relation entre la démarche et les chutes chez les personnes âgées atteintes de démence. La force, l'équilibre et l'entrée sensorielle dans les facteurs physiques et sensoriels, ainsi que l'attention, l'intégration sensorielle et la planification motrice dans les facteurs cognitifs, ont un impact direct sur la performance de la démarche et entraînent donc un risque accru de chutes dans cette population. D'autres facteurs physiques, sensoriels et cognitifs, tels que le contrôle inhibiteur et le jugement, peuvent également contribuer au risque de chute ; cependant, leur impact n'est pas médié par la démarche. Le gel de la marche est une perturbation grave de la marche, fréquente aux stades intermédiaire et avancé de la maladie de Parkinson, qui affecte la mobilité et augmente le risque de chute. Des capteurs portables ont été utilisés pour détecter et prédire le blocage de la marche, dans le but ultime de prévenir ce phénomène ou d'en réduire les effets à l'aide de dispositifs de surveillance et d'assistance à la marche. [Borzì et al. \(2021\)](#) ont utilisé des dispositifs portables intégrant des capteurs inertiels 2.7 et l'apprentissage automatique pour détecter et prédire l'apparition d'épisodes de gel de la marche dans la maladie de Parkinson. Ils ont combiné deux modèles, un arbre de décision pour la sélection des caractéristiques et SVM pour la classification. Les résultats obtenus avec la combinaison des deux méthodes sont : une sensibilité de 84,0 %, une spécificité de 88,3 % et une précision de 87,4 %.

Plusieurs études ont été menées sur la détection de chute, mais aussi sur la détection de la pré-chute ou encore la prévention de chutes spécifiques. Toutefois, il est possible de constater que très peu d'études se sont intéressées à la détection du risque

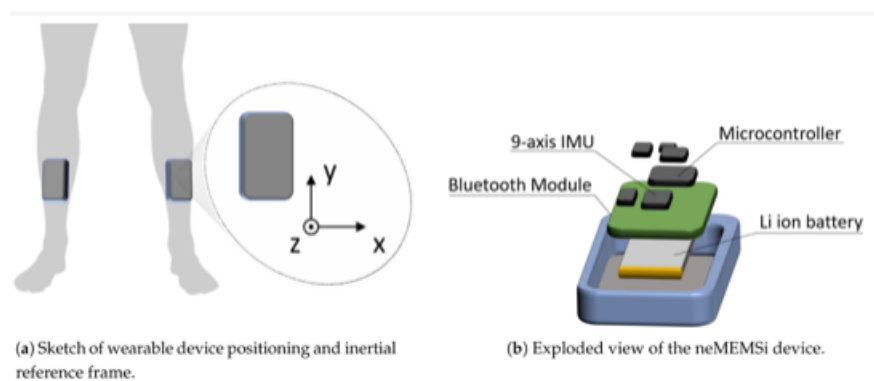


FIGURE 2.7 – Dispositif et schéma du positionnement du dispositif portable [Borzì et al. \(2021\)](#).

de chute pouvant être relié à des événements physiologiques ou environnementaux.

[Nouredanesh et al. \(2021\)](#) [Hilbe et al. \(2010\)](#) ont proposé une alarme "Bed-exit" pour prévenir les chutes du lit du patient. Leur système utilise des capteurs de pression intégrés aux barrières latérales du lit du patient pour suivre sa tentative de sortie du lit. Si la valeur du capteur de pression dépasse un certain seuil, une alarme est envoyée aux cliniciens afin d'empêcher une chute.

[Shany et al. \(2011\)](#) ont décrit l'utilisation de capteurs portables et de méthodologies pour l'évaluation du risque de chute dans des environnements supervisés et non supervisés. Cependant, ils n'ont pas abordé le test et la validation des méthodologies d'évaluation du risque de chute et les mises en œuvre dans le monde réel sur des personnes tombant fréquemment.

[Uddin \(2019\)](#), présente un système multimodal de prédiction de l'activité humaine. Le système inclut de multiples capteurs portables (Electrocardiogram (ECG), magnétomètre, accéléromètre et gyroscope) couplé à un réseau de neurones récurrent pour la classification des événements. Le système analyse douze activités humaines différentes à partir d'un ensemble de données publiques appelé MHEALTH. Enfin, le Recurrent Neural Networks (RNN) entraîné a été utilisé pour prédire l'activité sous-jacente sur des données de capteurs inconnues. Les performances de prédiction obtenues atteignent les 99,99 %.

2.3.4 Risque d'oubli de prise de médicaments

Plus de 80 % de la population âgée souffre d'au moins une maladie chronique et 50 % en a au moins deux [Goulding et al. \(2003\)](#). L'OMS a estimé que, dans les pays développés, environ 50 % des patients atteints de maladies chroniques ne prennent pas leurs médicaments tels qu'ils devraient l'être [Brown et Bussell \(2011\)](#).

Pourtant, les médicaments sont souvent utilisés comme principale approche pour prévenir et gérer efficacement les maladies chroniques. Cependant, les taux estimés d'adhésion à un régime de médication à long terme sont d'environ 50 %, et il n'y a

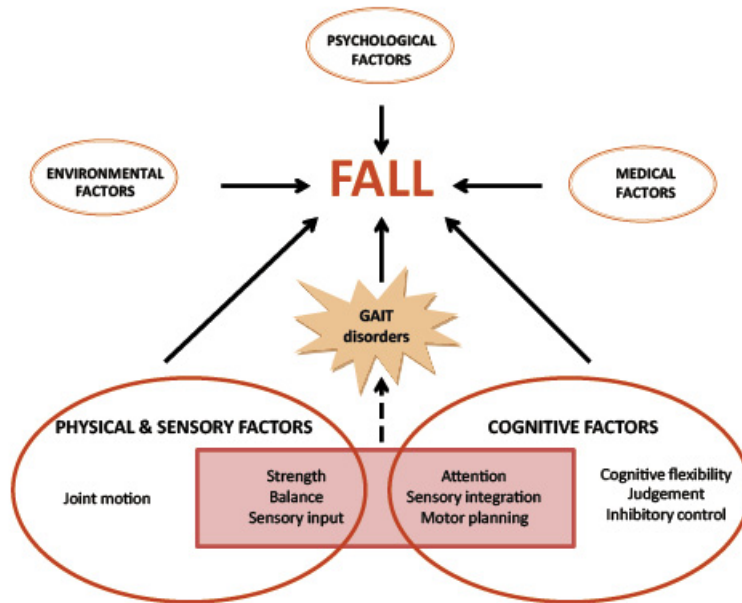


FIGURE 2.8 – La relation entre la démarche et les chutes chez les personnes âgées atteintes de démence. Taylor et al. (2014).

aucune preuve de changements significatifs au cours des 50 dernières années Organization et al. (2003). La non-adhésion du patient au traitement a été définie comme la prise de moins de 80 % des doses prescrites ou le dépassement des doses prescrites Organization et al. (2003).

Le développement d'un système non intrusif de surveillance de la prise de médicaments en temps réel peut simplifier le processus de prise de médicaments en détectant les activités de prise de médicaments et en les suivant Chang et al. (2019). De nombreuses études suggèrent que la facilité d'utilisation, l'utilité et les avantages perçus sont étroitement liés à l'acceptation, à la satisfaction et à l'intention de l'utilisateur d'utiliser un système de surveillance de la santé mobile, ce qui peut avoir une incidence directe sur l'observance thérapeutique Wu et al. (2007).

La surveillance de la prise de médicaments peut être classée en deux grandes catégories : directe et indirecte. Dans le premier cas, il s'agit d'observer une personne qui prend des médicaments ou de procéder à des tests de médicaments en laboratoire. La seconde implique l'auto-déclaration, le comptage des pilules, le suivi des renouvellements de médicaments et le suivi électronique à l'aide de vêtements intelligents, de caméras ou de piluliers équipés de systèmes de surveillance des événements médicamenteux Ma et al. (2018). Les méthodes directes sont les plus précises, mais peu discrètes, plus longues et plus coûteuses. Les méthodes indirectes sont des outils relativement peu coûteux, efficaces et moins gênants pour surveiller et signaler la prise de médicaments. Les dispositifs intelligents à porter sur soi ont été utilisés dans l'observation indirecte de l'adhésion aux médicaments de nombreuses façons.

De nombreux systèmes de détection de prise de médicaments ont été développés incluant différents types de capteurs avec des algorithmes de détection donnant des résultats très encourageants.

[Boonnuddar et Wuttidittachotti \(2017\)](#) présente une application développée pour smartphone Android qui demande à ces utilisateurs de confirmer leur adhésion à la prise de leur traitement. Elle utilise également une carte Arduino et un capteur à cellule de charge pour suivre la prise des médicaments du patient. Le système peut suivre l'état de la prise de médicaments en utilisant le capteur de poids des médicaments, puis alerter le patient à l'avance afin que les soignants reçoivent les bonnes informations et puissent aider le patient à prendre ses médicaments au bon moment. Cette recherche propose donc un nouveau prototype pour un centre de soins pour personnes âgées ou une famille dont les membres ont besoin de contrôler et de suivre les médicaments du patient. L'efficacité du système de notification est de 96,88 %.

[Moshnyaga et al. \(2016\)](#) développe un système intelligent basé sur Kinect pour le contrôle discret de l'adhésion aux médicaments des personnes souffrant de troubles de la mémoire. C'est un système destiné aux patients souffrant de démence. Contrairement aux piluliers, le système ne se contente pas de rappeler au patient l'heure de la prise de médicaments, mais il guide également vocalement le patient à travers les étapes de la prise de médicaments, en alertant le soignant en cas de problème.

[Li et al. \(2014\)](#) ont conçu un système de surveillance thérapeutique composé d'un pilulier cylindrique à 7 compartiments, d'un bracelet et d'un ordinateur qui communiquent tous entre eux sans fil. Le pilulier est composé d'un microcontrôleur Arduino, d'un moteur, d'un émetteur-récepteur ZigBee et d'un lecteur RFID. En outre, chaque compartiment est équipé d'une diode et d'une photodiode pour détecter le retrait de la pilule. La fonction de l'unité centrale est de commander le moteur de manière à ce qu'il fasse tourner le compartiment vers l'utilisateur lorsqu'il est temps de prendre un médicament et lorsque le bracelet RFID est détecté à proximité du pilulier. Le bracelet intègre un Inertial Measure Units (IMU), une étiquette RFID et une LED. Il est donc utilisé pour collecter les données de mouvement associées au prélèvement et à la prise de pilules. En outre, l'étiquette RFID et la LED sont utilisées pour la détection de la proximité RFID dans le pilulier et pour le rappel, respectivement. Une fois que le bracelet est détecté à proximité du pilulier, l'IMU est activé pour échantillonner les données de mouvement afin de détecter d'éventuels gestes de la main liés à la prise de pilules. Enfin, afin de vérifier le retrait de la pilule, la photodiode émet une tension à mesurer par le microcontrôleur, à un niveau donné en fonction de la lumière reçue de la LED et du blocage qui peut ou non résulter de la présence de la pilule dans ce compartiment.

[Cheon et al. \(2020\)](#), présente une approche utilisant les données de capteurs enregistrées par une Apple Watch pour détecter la prise d'un traitement en pilules dans des flacons d'ordonnance standard. Le modèle final a obtenu une précision moyenne validée par recoupement et des scores F1 de 80,27 % et 80,22 %, respectivement. Deux cas solutions sont présentées, lesquelles sont des dispositifs portables tels que l'Apple Watch, contribuant aux efforts visant à améliorer l'adhésion des patients

aux médicaments.

2.3.5 Analyses

De nombreuses études s'intéressent à la détection de situations à risques pour les personnes âgées touchant tous les segments de cette population avec de nombreuses situations abordées. A travers ces différentes études, nous pouvons constater que les dispositifs développés restent spécifiques à une situation à risques. Ce type de dispositif ne répond pas aux attentes des personnes fragiles qui ont besoin d'un système facile à utiliser dans leur quotidien. Les dispositifs mis en place rassemblent plusieurs capteurs avec des potentiels de reconnaissance élevés, des performances des algorithmes et des algorithmes de reconnaissance de plus en plus robustes.

2.4 Réseaux de neurones profonds pour les technologies eHealthcare

2.4.1 Réseaux de neurones profonds

Ces dernières années, les réseaux de neurones artificiels ont dominé et dépassé les méthodes classiques de traitement de données dans différents domaines (séries temporelles, traitement d'images, vidéos etc). Ces avancées ont sans aucun doute apporté de grands avantages à différents secteurs de la société, allant de la recherche à l'industrie. Les réseaux de neurones artificiels Artificial Neural Network (ANN) sont des systèmes informatiques inspirés des réseaux de neurones biologiques qui constituent le cerveau des animaux. Ils ont été décrits pour la première fois en 1943 dans [McCulloch et Pitts \(1943\)](#), en présentant un modèle simplifié des neurones biologiques et en établissant un lien direct entre les relais téléphoniques et les connexions synaptiques entre les neurones biologiques. Après la théorie de l'assemblage cellulaire de [Hebb \(1949\)](#), celui qu'on considère aujourd'hui comme le premier modèle formel d'un réseau de neurones artificiel le perceptron multicouches Multi Layer Perceptron (MLP) a été mis au point par [Rosenblatt \(1958\)](#). Malgré un fort engouement autour des réseaux de neurones artificiels dans les années 1960, les architectures encore rudimentaires et les machines de calcul peu performantes n'ont pas permis d'aller au-delà de l'exécution de simples opérations de calcul. Après une période avec peu de travaux sur les réseaux de neurones, à la fin des années 1980, un regain d'intérêt a été observé avec l'arrivée de nouvelles architectures et de nouvelles techniques d'apprentissage comme la rétropropagation qui a été testée pour la première fois en 1986 [Rumelhart et al. \(1986\)](#). Actuellement, nous assistons à une explosion de l'utilisation des réseaux de neurones et ce, grâce à la combinaison de quatre éléments : i) la capacité de calcul, ii) les ensembles de données à grande échelle, iii) les réseaux neuronaux profonds et iiiii) des algorithmes de calcul bien plus performants. Dans les CNN, la principale opération d'extraction de caractéristiques est la convolution (d'où leur nom). Le premier CNN, le Neocognitron, a été proposé par [Fukushima et Miyake \(1982\)](#). Cependant, dans la littérature, la première application des réseaux convolutifs est attribuée à [LeCun et al. \(1998\)](#) pour la reconnaissance de chiffres manuscrits. La capacité de calcul limitée à l'époque où ces approches ont été publiées a détourné l'attention d'une grande partie de la communauté des chercheurs, mais pas de tous. En 2012, la nouvelle ère de l'apprentissage profond a commencé avec AlexNet [Krizhevsky et al. \(2012a\)](#), un CNN utilisé pour battre l'état de l'art dans le défi ImageNet en réduisant considérablement l'erreur de classification dans un ensemble de données à grande échelle avec plus d'un million d'images et un millier de catégories. Contrairement aux MLP, un CNN peut exploiter la structure des données. L'intuition sous-jacente est que les données d'entrée - telles que les images, les clips vidéo ou sonores - contiennent une structure topologique locale qui ne dépend pas de l'emplacement spécifique donné par un système de référence, c'est-à-dire

les coordonnées de l'image. Cela signifie que les objets peuvent apparaître à n'importe quelle position dans l'image. Ensuite, un CNN utilise cette compréhension pour appliquer le même détecteur à n'importe quel endroit de l'image. Le détecteur de motifs est dans ce cas appliqué par des couches de convolution. En outre, les CNN comportent d'autres opérations que la convolution, comme les couches de mise en commun, les couches d'activation et les couches entièrement connectées. La conjonction de toutes ces opérations dans une chaîne peut générer des caractéristiques utiles qui permettent l'identification de modèles. Cependant, malgré toutes les couches nécessaires à la construction d'un CNN, les couches de convolution et de mise en commun sont considérées comme les plus importantes pour extraire des caractéristiques multiples et réduire la complexité de calcul, respectivement.

2.4.2 L'apprentissage automatique au service des personnes âgées fragiles

L'objectif ultime des systèmes IoT pour le suivi des sujets fragiles consiste en une prise de décision en temps réel, pour la détection et la prévention des risques. La tendance générale pour une telle détection consiste à développer des approches d'apprentissage automatique supervisé [Boll et al. \(2019\)](#). La complexité des opérations typiques appliquées à l'apprentissage automatique et au traitement du signal, telles que l'inversion et la décomposition de matrices, est estimée à $O(n^3)$. Cette complexité implique que s'il y a 100 échantillons, une opération algébrique simple nécessite environ 1 million d'instructions de boucle interne. Il convient de noter que la prise en compte d'instructions d'opérateurs supplémentaires, tels que les produits en virgule flottante, n'est pas incluse. Cependant, avec la popularisation des dispositifs portables, des algorithmes légers ont été mis en œuvre pour les composants de traitement de base des techniques en ligne, tels que les filtres de bruit, l'extraction de caractéristiques et les détecteurs de pics [Andreu-Perez et al. \(2015\)](#). Aujourd'hui, les classificateurs les plus efficaces sont les réseaux neuronaux profonds. Ces classificateurs sont lourds à former, mais sont assez légers dans la prise de décision car ils nécessitent des opérations matricielles qui sont exécutées en parallèle et peuvent être mises en œuvre sur des dispositifs mobiles [Merenda et al. \(2020\)](#). Néanmoins, avant qu'une architecture conçue puisse être allégée et transférée à un dispositif portable, par exemple par quantification du réseau [Hubara et al. \(2017\)](#), il est nécessaire de concevoir une solution efficace de réseaux neuronaux profonds offrant une grande précision dans la classification des signaux multimodaux provenant des dispositifs portables. Les progrès simultanés des processeurs graphiques Graphics processing unit (GPU) ont permis de traiter plus efficacement de gros volumes de données dans de grands référentiels. En outre, cette technologie permet également d'effectuer des transformations et des opérations parallèles intensives, ce qui profite à de nombreux algorithmes. Support Vector Machines - SVM [Corinna et Vapnik \(1995\)](#), ont proposé pour la première fois en 1995, un algorithme d'apprentissage automatique supervisé appliqué aux problèmes de classification et de régression. À son niveau fondamental,

une machine à vecteurs de support est une méthode de classification ou de prédiction et est formée par une combinaison de techniques statistiques et d'apprentissage automatique. Ce puissant classificateur discriminant classe formellement les cas en identifiant et en optimisant une frontière de séparation appelée hyperplan. L'approche SVM présente de nombreux avantages. Ce classificateur peut surmonter le problème de la haute dimensionnalité, un des problèmes typiques de l'apprentissage automatique lorsqu'il y a un grand nombre de variables d'entrée correspondant au nombre d'observations disponibles. Un autre avantage du SVM est que moins de données d'entraînement doivent être saisies par rapport aux autres algorithmes, tandis que les erreurs et la complexité peuvent être minimisées.

L'enregistrement des données pendant la surveillance de sujets fragiles peut être effectué à l'aide de capteurs et de caméras vidéo portables, comme cela a été proposé dans [Meditskos et al. \(2018\)](#) pour les patients atteints de maladie d'Alzheimer. Les données des capteurs sont des séries temporelles qui peuvent être traitées efficacement par des RNN [Pandey et Janghel \(2019\)](#). Il s'agit de familles de réseaux neuronaux qui contiennent des couches cachées pour le traitement séquentiel des données. Les réseaux neuronaux récurrents disposent d'une "mémoire" qui saisit les informations sur ce qui a été calculé jusque-là et exécutent la même tâche pour chaque élément d'une séquence, la sortie dépendant des calculs précédents. Cette "mémoire" est utile dans de nombreuses applications comportant des données "temporelles", telles que l'analyse de textes, de la parole ou de séquences d'ADN. Les RNN peuvent non seulement apprendre les dépendances temporelles locales et à long terme des données, mais aussi s'adapter à des séquences d'entrée de longueur variable. Le RNN est généralement alimenté par des ensembles d'apprentissage qui présentent de fortes interdépendances et une représentation significative pour conserver des informations sur ce qui a résulté de toutes les étapes temporelles précédentes. Le résultat obtenu par le réseau au temps $t-1$ affecte le choix au temps t . C'est la "mémoire" du RNN. Bien que le RNN soit un algorithme simple et puissant, il souffre du problème des gradients qui disparaissent et explosent. Pour combattre ces problèmes, une conception particulière des RNN a été proposée par [Hochreiter et Schmidhuber \(1997\)](#), à savoir les architectures de mémoire à long terme et à court terme (LSTM). Actuellement, les RNN tels que les LSTM et les réseaux GRU (Gated Recurrent Unit) ont attiré l'attention des chercheurs. Les GRU ont suscité l'intérêt des chercheurs dans le domaine de la santé. Dans l'article [Che et al. \(2018\)](#), une nouvelle méthode GRU-D, basée sur les réseaux neuronaux récurrents est utilisée pour la prédiction des conditions des patients avec différents ensembles de données de santé (l'architecture de cellule GRU-D est représentée dans la figure 2.9). [Friedrich et al. \(2021\)](#) présentent une recherche sur une architecture de réseau de neurones qui est une combinaison de deux réseaux de neurones, LSTM et de CNN (voir figure 2.10) entraîné sur des données mesures inertielles (IMU) pour prédire le score d'évaluation de la mobilité des personnes âgées. Ces résultats apportent des informations utiles aux médecins pour diagnostiquer les changements dans la mobilité et la performance physique des personnes âgées et pouvoir alerter sur l'état d'avancement de la fragilité.

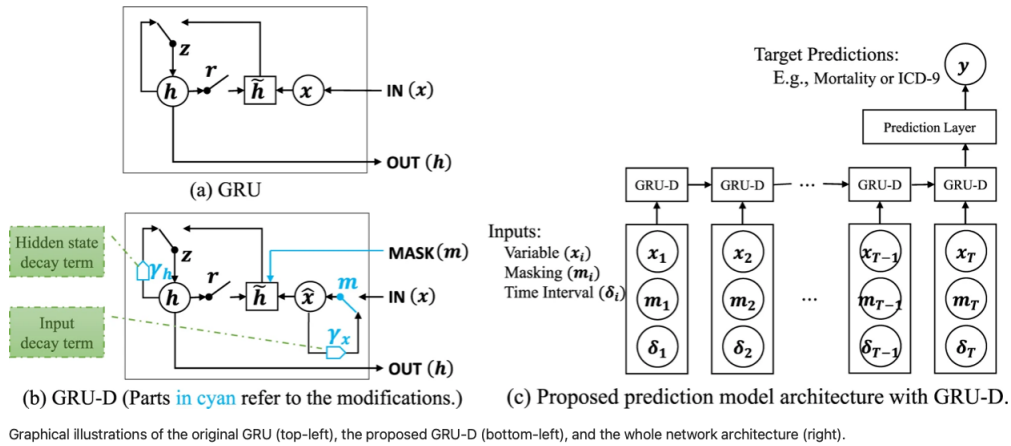


FIGURE 2.9 – Illustrations graphiques de la GRU originale (en haut à gauche), de la GRU-D proposée (en bas à gauche) et de l’architecture complète du réseau (à droite) [Che et al. \(2018\)](#).

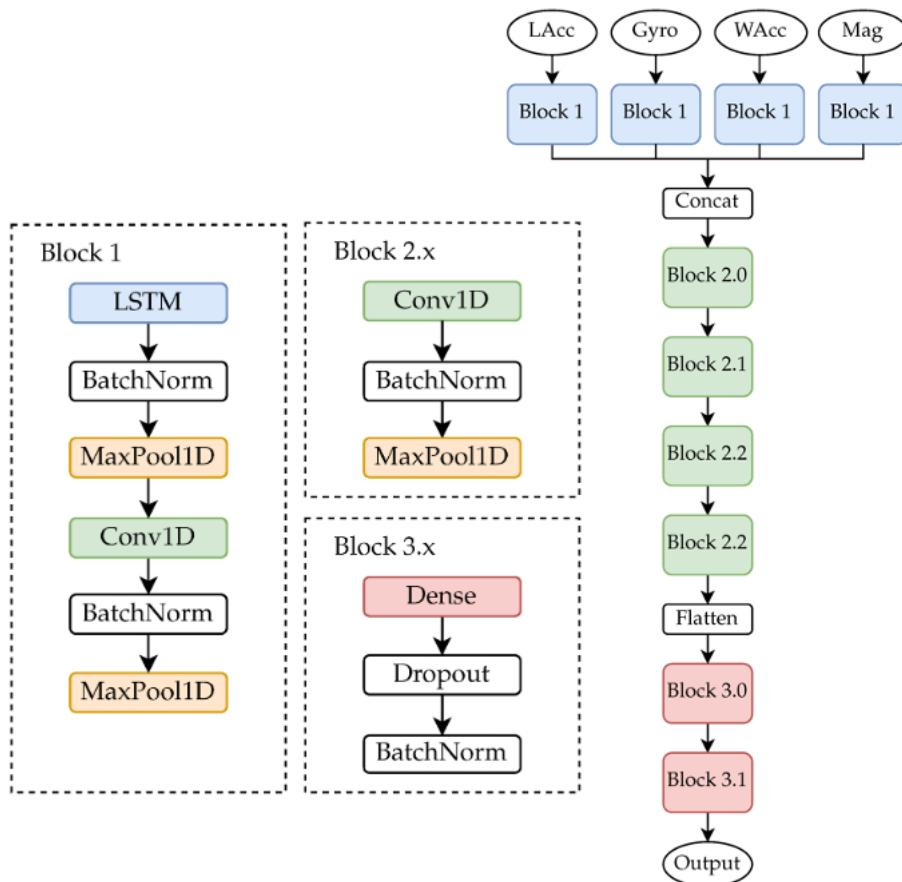


FIGURE 2.10 – (à droite) l’architecture du réseau présenté par [Friedrich et al. \(2021\)](#), (à gauche) Détails d’architecture des blocs constituant le réseau de neurone.

2. État de l'art

Dans le traitement des données visuelles, les réseaux neuronaux convolutionnels (CNN) ont gagné en popularité. Les CNN appartiennent à la classe des modèles de reconnaissance visuelle d'inspiration biologique. Le premier réseau de neurones convolutif (CNN) fut introduit par [LeCun et al. \(1990\)](#) en 1989. Il a été conçu pour la reconnaissance d'images et un des premiers tests effectués avec l'approche CNN a été réalisé sur la reconnaissance de lettres manuscrites [Lecun et al. \(1998\)](#).

L'idée a été de passer l'image dans une succession de filtres convolutifs apportant une description réduite et pertinente de l'image. Ces caractéristiques sont par la suite envoyées à un perceptron multicouches, composé de couches cachées et d'une couche de sortie complètement connectées permettant la classification du chiffre présent dans l'image (cf Figure 2.11).

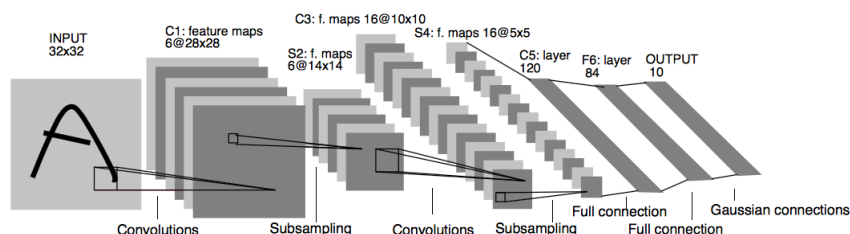


FIGURE 2.11 – Architecture d'un réseau CNN pour la reconnaissance de lettres manuscrites (Figure extraite du [LeCun et al. \(1998\)](#)).

Les CNN utilisent des filtres convolutifs pour combiner des fonctions de bas niveau afin de créer des classificateurs précis et efficaces [LeCun et al. \(2015\)](#). Les résultats obtenus par cette approche sont encourageants pour l'avenir et offrent de nouvelles perspectives pour la détection, par exemple de chute à partir de données visuelles [Li et al. \(2017\)](#). De nombreuses études ont appliqué les CNNs pour détecter les situations à risque, en particulier les chutes [Nait Aicha et al. \(2018b\)](#). Les CNNs représentent également un bon cadre pour la fusion de données [Martin et al. \(2020b\)](#). [Nait Aicha et al. \(2018b\)](#) propose l'utilisation d'informations biomécaniques et la fusion de données pour augmenter l'efficacité des méthodes d'apprentissage profond dans la détection des risques de chute. Ils ont utilisé des données existantes provenant de 296 personnes âgées pour comparer les performances de trois architectures de modèles d'apprentissage profond CNN, LSTM et une combinaison de ces deux approches ConvLSTM. Les résultats montrent que les modèles ConvLSTM évaluent efficacement le risque de chute à partir des données des capteurs portables et des informations biomécaniques. Il est donc intéressant de combiner ces types de réseaux dans le traitement de données multimodales complexes. Pour l'analyse des informations vidéo temporelles, les réseaux convolutionnels 3D conviennent car ils permettent de travailler avec des volumes spatio-temporels, c'est-à-dire des morceaux d'images vidéo successives, afin de prendre en compte la cohérence temporelle et les singularités pour la classification vidéo. [Ji et al. \(2012\)](#), est considéré comme un article fondateur pour les réseaux neuronaux CNN 3D appliqués à la reconnais-

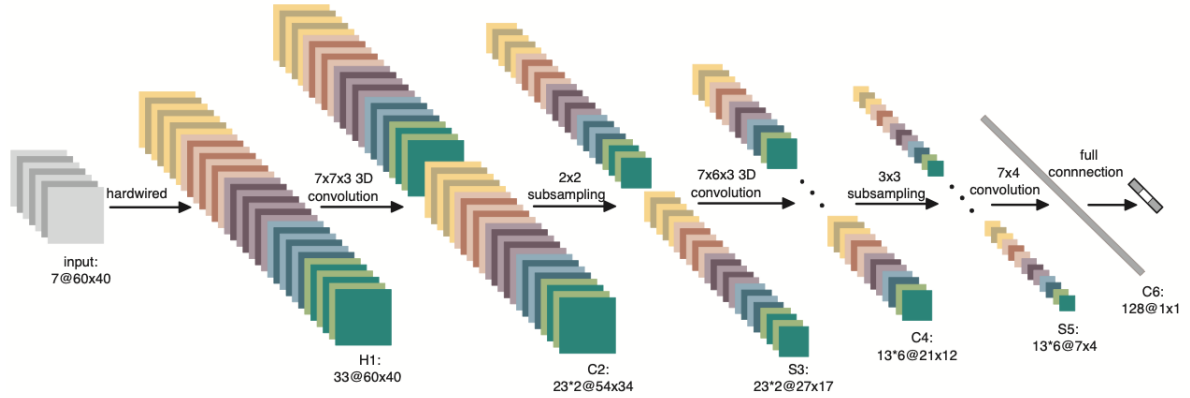


FIGURE 2.12 – L’architecture CNN 3D pour la reconnaissance d’actions humaines, présentée par [Ji et al. \(2012\)](#).

sance d’actions sur des données réelles. Ce modèle extrait les caractéristiques des dimensions spatiales et temporelles en effectuant des convolutions 3D, capturant ainsi l’information de mouvement encodée dans plusieurs images adjacentes. Cette architecture est décrite dans la figure 2.12, elle consiste en 1 couche câblée, 3 couches de convolution, 2 couches de sous-échantillonnage, et 1 couche de connexion complète. Une approche des CNN appelée Inflated 3D (I3D) pour Inflated 3D ConvNet qui est basée sur l’inflation de 2D ConvNet est introduite [Carreira et Zisserman \(2017\)](#). Elle utilise des filtres et des noyaux de mise en commun de ConvNets de classification d’images très profonds et les étend en 3D le long de l’axe temporel, ce qui permet d’apprendre des extracteurs de caractéristiques spatio-temporelles transparents à partir de vidéos. [Martin et al. \(2020b\)](#) utilise un réseau inspiré du réseau I3D pour reconnaître une action sportive dans le cadre d’une activité de tennis de table. Ainsi, dans une architecture générale à concevoir pour l’exploration de données multimodales, il est intéressant d’utiliser des CNN 3D pour la composante vidéo des données.

Les réseaux à deux flux ont été très populaires pour certaines tâches de classification vidéo telles que les tâches de reconnaissance d’action. [Simonyan et Zisserman \(2014\)](#) ont conçu un réseau à deux flux, l’un pour la composante spatiale de la vidéo tandis que l’autre flux est destiné à l’information de mouvement, c’est-à-dire que ce flux capture l’information temporelle et utilise finalement la fusion tardive pour le calcul des scores de classe. Il existe d’autres approches de fusion à deux flux pour la reconnaissance d’actions vidéo, telles que l’introduction de couches de fusion convolutives et de couches de fusion temporelle [Feichtenhofer et al. \(2016\)](#). Un ConvNet de fusion spatio-temporelle applique des ConvNets à deux flux, qui capturent des informations à court terme à une échelle temporelle fine ($t \mp L2$), à des entrées temporellement adjacentes à une échelle temporelle grossière ($t + T\tau$). Les deux flux sont fusionnés par un filtre 3D capable d’apprendre des correspondances

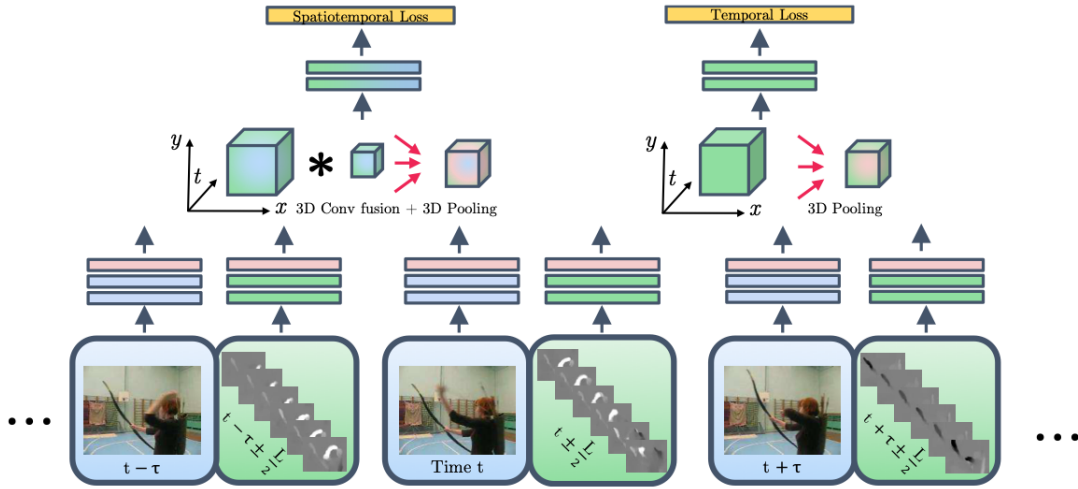


FIGURE 2.13 – Architecture présentée par Feichtenhofer et al. (2016) pour la fusion des informations spatio-temporelles pour la reconnaissance d'action sur les données vidéo.

entre des caractéristiques hautement abstraites du flux spatial et du flux temporel, ainsi que des combinaisons locales pondérées en x, y, t (cf figure 2.13). Les caractéristiques résultantes du flux de fusion et du flux temporel sont mises en commun en 3D dans l'espace et le temps pour apprendre des caractéristiques spatio-temporelles et purement temporelles pour reconnaître la vidéo d'entrée. Das et al. (2020) étudie l'aspect spatio-temporel de la reconnaissance des activités de la vie quotidienne. En considérant deux propriétés décrivant les activités de la vie quotidienne à savoir les motifs spatio-temporels subtils et les motifs visuels similaires variant dans le temps. Ils proposent un réseau de neurones appelé Video Pose Network (VPN) (Video Pose Network) avec deux composants clés que sont un encastrement spatial et un réseau d'attention. L'intégration spatiale projette les poses 3D et les indices RVB dans un espace sémantique commun. Cela permet au cadre de reconnaissance d'action d'apprendre de meilleures caractéristiques spatio-temporelles. Les résultats obtenus avec le VPN encouragent dans l'idée d'inclure l'attention pour la reconnaissance des actions de la vie quotidienne.

2.5 Conclusion

Dans ce chapitre, nous avons évoqué les systèmes IoT dans la surveillance des personnes âgées et/ou en situation de fragilité. Nous avons fait l'état de l'art concernant les dispositifs de détection de situations à risques pour les personnes âgées. Nous avons parlé des différentes situations à risques abordées par les recherches les plus récentes en précisant des choix des capteurs utilisés. Nous avons aussi abordé les différentes méthodes de détection utilisées en mettant l'accent sur les méthodes usant des réseaux de neurones profonds pour la classification d'événements suivant

des données multicateurs avec des séries temporelles.

Dans le chapitre suivant, nous définirons les différentes situations à risques étudiées et la mise en place de notre dispositif d'acquisition des données avec le choix des capteurs adéquats et des technologies adoptées.

Chapitre 3

De la taxonomie à la mise en place du dispositif d'acquisition

3.1 Introduction

Nombreuses sont les technologies aujourd'hui visant à soutenir la vie indépendante des personnes. Le spectre de surveillance est très grand allant du plus commun telle que la détection de chute au plus pointu en prenant l'exemple de la détection des crises épileptiques grâce à un bracelet de collecte de signaux multimodaux [Regalia et al. \(2019\)](#). Par conséquent, les capteurs portables (Wearable Sensors (WS)) suscitent un grand intérêt et leur développement est poussé pour en faire des outils efficaces pour la prévention, la détection précoce des chutes et le suivi des activités de la vie quotidienne. Un progrès considérable a été rendu possible grâce au développement de méthodes et de technologies de collecte de données à distance pour un envoi de données plus facile et moins coûteux d'un point de vue énergétique [Sabesan et Sankar \(2015\)](#). En ce qui concerne les applications de surveillance des activités de la vie quotidienne et de la vie autonome chez les personnes âgées, plusieurs études ont montré l'efficacité des systèmes WS et IoT d'un point de vue technologique [Hegde et al. \(2017\)](#)-[Hegde et Sazonov \(2014\)](#)-[Bellagente et al. \(2018\)](#). Pour la surveillance à distance, les applications basées sur les smartphones et les smartwatches sont fréquemment utilisées pour les personnes âgées [Bellagente et al. \(2018\)](#)-[Kheirkhahan et al. \(2019\)](#)-[Khojasteh et al. \(2018\)](#). Le principal avantage de ces dispositifs est qu'ils sont non invasifs et confortables pour les utilisateurs.

Au vu des enjeux pour la société, la recherche dans ce domaine est très active et ses implémentations industrielles telles que les montres connectées sont populaires.

Diverses solutions technologiques se révèlent être des outils potentiellement intéressants pour favoriser l'autonomie à domicile, et les résultats récents concernant la prévention des chutes sont prometteurs. Néanmoins, des défis et des obstacles à l'adoption plus large des applications WS et IoT existent toujours [Baig et al. \(2019\)](#).

Si la détection des chutes est assez bien étudiée, la détection de risques plus sémantiques reste un défi. Dans l'article [Yebda et al. \(2020\)](#), nous avons introduit les "risques sémantiques" qui sont aujourd'hui au centre de l'intérêt des psychologues et des kinésioles. Leur détection nécessite la collecte de données visuelles ainsi que des mesures par capteurs. Dans cette mesure, l'enregistrement des données s'ap-

parente à un "journal de vie". La méthodologie consiste aujourd'hui à reconnaître différentes situations pouvant présenter des risques pour les personnes dites fragiles et partir de là, regrouper les capteurs nécessaires à la reconnaissance de celles-ci. Pour finir, nous présenterons la solution retenue pour la détection de ces situations à risques.

3.2 Définition des situations à risques

Plusieurs études ont été menées pour définir les situations à risques les plus courantes rencontrées par les personnes fragiles [Vemulapalli et al. \(2017\)](#). D'après ces études, les accidents les plus fréquents chez les personnes âgées sont les chutes. Il s'agit du troisième événement de santé le plus fréquent après les infarctus et les cancers [Rubenstein \(2006\)](#). Néanmoins, les risques encourus par les personnes fragiles diffèrent en fonction de leur propre historique médicale. En dehors des cas de chutes, il existe d'autres situations plus spécifiques comme les chutes "Parkinsoniennes", les risques spécifiques des personnes atteintes de la maladie d'Alzheimer en plus des diabétiques, les risques des personnes souffrant d'insuffisance cardiaque, rénale ou respiratoire et d'autres maladies chroniques. Les situations de ces patients diffèrent en fonction de la maladie. Par exemple, pour les malades de Parkinson, les risques les plus urgents sont les chutes parkinsoniennes, qui sont très fréquentes, pour les malades d'Alzheimer, il y a les situations de stress, la perte d'orientation et la perte du chemin. Pour les personnes diabétiques, les hyperglycémies et les hypoglycémies sont dangereuses. Compte tenu des nombreuses études sur les situations à risques pour les personnes fragiles, nous avons choisi de concevoir un système hétérogène pouvant être utilisé dans de nombreuses situations de fragilité. Ainsi, si un dispositif générique doit être conçu avec une couverture de cohortes hétérogènes de sujets fragiles, il doit détecter le plus grand nombre possible de situations à risque.

Sans faire référence à une pathologie spécifique, les situations à risques telles que définies par les études psychologiques [Avila-Funes et al. \(2008\)](#) peuvent être décomposées en situations à risques immédiates et situations à risques survenant à long terme.

3.2.1 Les situations à risques immédiates

sont toutes les situations à risques pouvant survenir dans l'immédiat et ne nécessitant pas d'études longitudinales pour leur détection.

Risque de chute

Peuvent être considérés comme tous les événements précipitant l'avènement de la chute. Reprenant les chutes accidentelles, elles sont largement étudiées, et nous pouvons aujourd'hui déterminer avec une précision allant de 150 à 600 ms le début de la chute, qui est par définition l'élément qui précède la chute. Pour le risque

de chute, le sujet est moins étudié car il regroupe plusieurs facteurs devant être étudiés en même temps. Dans nos études avec l'équipe de kinésiologie de l'université de Waterloo (Canada) et de l'équipe de neuropsychologues de l'unité de l'Inserm BPH de Bordeaux nous avons dénombré trois types de risques de chutes : *le risque de chute physiologique, le risque de chute due à une action et le risque de chute environnementale*. La figure 3.1 Cuevas-Trisan (2017) résume les éléments de mesure de risque de chute chez les personnes âgées.

Le risque de chute physiologique : lié aux antécédents du sujet, il s'agit des facteurs intrinsèques à son état pathologique. Sont considérés comme tels tous les agents qui déclenchent la chute. Ces facteurs sont souvent multiples chez les sujets fragiles avec une association entre eux. Les causes cardiaques dominent, notamment les troubles du rythme supra-ventriculaires, les troubles de conduction, l'infarctus du myocarde, l'embolie pulmonaire ou la sténose aortique serrée. Parmi les causes vasculaires, l'hypotension orthostatique, impliquée dans 10 à 15 % des chutes, est de loin la plus fréquente. Les causes neurologiques sont dominées par les accidents vasculaires cérébraux. Les épisodes confusionnels exposent à un risque de chute par l'altération de la vigilance. Enfin, les causes métaboliques sont représentées par les hypo et hyperglycémies surtout iatrogènes peuvent aussi s'observer lors de cachexie. Les sujets fragiles sont très exposés au risque de chute physiologique contribuant à une forme de stress et d'hypervigilance de la part du sujet lui-même ou de ses proches. Ce risque est accru notamment en situation d'accroupissement et en position assis debout (orthostatisme).

Le risque de chute environnementale : mis en cause, le port de chaussures inadaptées, trop lâches, maintenant mal le pied, à semelle usée ou glissante. Les chaises bancales, les tapis mal fixés, les objets mal rangés ou traînant au sol, les animaux domestiques, le sol mouillé ou glissant ou même monter les escaliers qui ne sont que des exemples parmi les plus fréquents des facteurs environnementaux précipitant la chute de la personne âgée fragile.

Le risque de chute due à une action : la marche est un mouvement constant qui peut être décrit comme une suite de rupture d'équilibre. C'est une activité motrice automatique alternant chute et parachute. Les modifications apportées par le vieillissement n'ont pas été étudiées de façon systématique. Plusieurs arguments tendent à montrer que les réactions posturales se maintiennent efficacement au cours du vieillissement physiologique mais la vitesse des réactions et les capacités d'adaptation aux situations extrêmes sont moindres. Les situations de risques de chute qui sont dues à des action et que nous avons intégrées dans notre étude sont le fait de trébucher, soulever (par exemple, une boîte), atteindre le plafond, se précipiter et balayer le sol. N'ayant presque aucun impact sur les personnes en bonne santé, elle peuvent représenter un vrai danger pour les personnes fragiles.

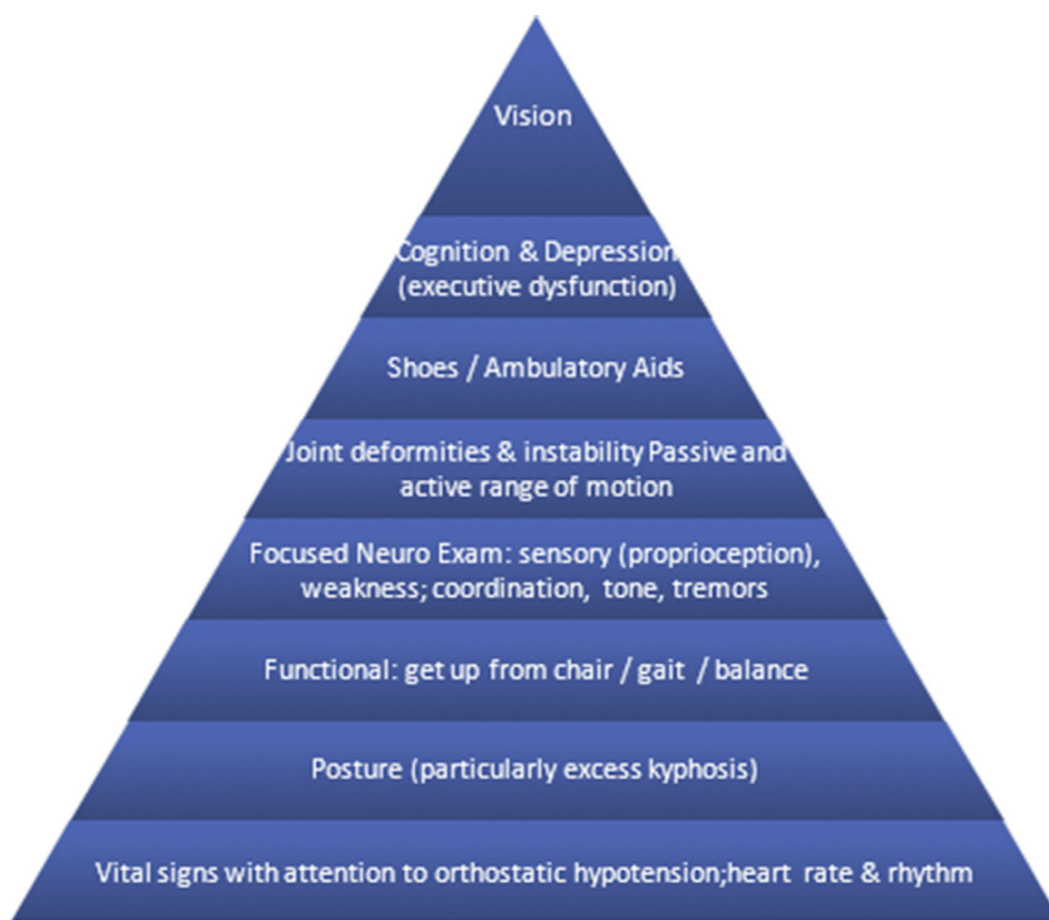


FIGURE 3.1 – Résumé des éléments de l'examen physique lors de l'évaluation du risque de chute chez une personne âgée Cuevas-Trisan (2017).

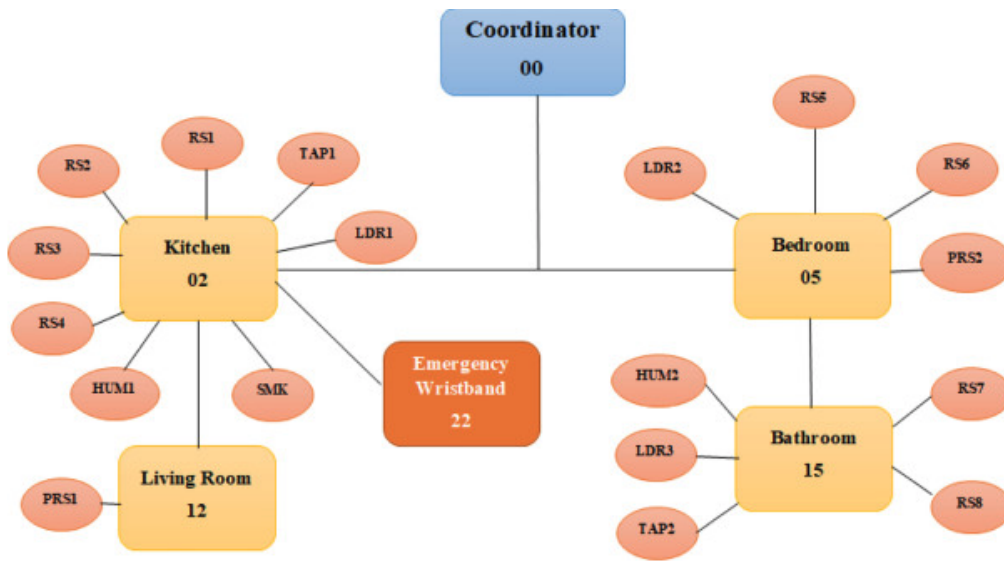


FIGURE 3.2 – Architecture du système et topologie du réseau présentée Sokullu et al. (2020).

Risque d'accidents domestiques

Situations très fréquentes chez les patients atteints de démence [MONIZ-COOK et al. \(1998\)](#), mais cela concerne également les personnes fragiles. On parle souvent de la perte des repères dans un lieu connu. Cette situation concerne aussi les oublis majeurs comme par exemple l'oubli d'un feu allumé, d'accidents pouvant survenir lors d'activités ménagères tel que le repassage ou le fait de se couper en cuisinant. [Sokullu et al. \(2020\)](#) présente le prototype d'un système AAL (Ambient Assisted Living Systems) basé sur l'IoT à faible coût qui peut reconnaître un ensemble d'activités domestiques dans le but de rassurer le patient en le suivant dans sa routine domestique sans intrusion. Ce système est basé sur des capteurs ambiants placés autour de la maison, ils collectent des mesures et les transmettent au coordinateur, où les informations sont évaluées, interprétées et transférées vers le cloud. L'architecture mise en place par [Sokullu et al. \(2020\)](#) est présentée dans la figure 3.2. Dans notre système, nous avons cherché à mettre au point une architecture que le sujet peut porter et ne nécessitant pas d'installation particulière donc avec des capteurs pouvant être portés, et nous informant essentiellement sur l'état physique et physiologique du sujet.

Risque d'intrusion d'une tierce personne

Tentative de fraude ou tentative d'intrusion par un tiers. La technologie de l'internet des objets (IoT) est utilisée pour renforcer la sécurité des personnes âgées vivant dans des environnements domestiques intelligents et vivant en autonomie. Le



FIGURE 3.3 – Illustration du risque d'intrusion sur le dataset NTCIR [Yebda et al. \(2020\)](#).

comportement quotidien des personnes âgées est collecté à l'aide de capteurs, puis évalué pour détecter tout comportement anormal. Les personnes âgées sont hélas souvent victimes d'escroquerie ou d'intimidation. Pour cette raison, l'intrusion d'une personne mal intentionnée dans le domicile de la personne âgée, est une situation à risques particulièrement redoutée par les personnes âgées ainsi que leurs familles. à risque notamment fragiles. On estime qu'environ 4 personnes âgées sur 1000 sont victimes de crimes violents non mortels [Reisig et Holtfreter \(2013\)](#). L'impact de la victimisation est accru en raison des déclin cognitif et fonctionnel associés au vieillissement [Beard et al. \(2009\)](#). Dans le cas d'intrusion d'une tierce personne, des données visuelles et sonores sont le plus souvent sollicitées. Néanmoins, l'état de stress de la personne pouvant être remonté par le signal Galvanic Skin Response (GSR) et le volume de pression sanguine sont aussi des mesures pouvant aussi nous informer sur le cas de stress de la personne. Dans notre étude sur le dataset NTCIR 13, nous nous sommes basés sur les données visuelles, nous avons notamment associé la présence d'une personne inconnue du sujet fragile à côté d'une porte à cette situation à risques. 3.3. Il est à noter que pour des questions de réglementation sur les données, nous avons décidé de ne pas prendre en compte, cette situation à risques dans notre corpus de données BIRDS.

3.2.2 Situations à risques survenant dans le temps

Risque de déshydratation

La déshydratation est fréquente chez les personnes âgées et peut nécessiter une hospitalisation. Ceci est en partie dû au fait que cette population perd peu à peu la sensation de soif et, par voie de conséquence, le réflexe de boire. En effet, il a été montré que même une déshydratation légère peut entraîner une déficience de la mémoire visuelle et de la mémoire de travail [Liaqat et al. \(2020b\)](#). Les professionnels de santé peuvent diagnostiquer une déshydratation par perte d'eau en prélevant un échantillon de sang et en mesurant l'osmolarité sérique, mais un test moins invasif serait utile. Plusieurs solutions sont proposées ces dernières années pour prévenir la déshydratation soit par la mesure du taux d'eau dans le corps, soit par la mesure de la quantité d'eau consommée. Parmi ces méthodes non invasives, nous pouvons citer l'analyse de l'impédance bioélectrique qui consiste à faire passer un courant lumineux dans le corps et à mesurer la résistance à ce courant. Une autre étude se basant sur l'utilisation de capteur Electrodermal Activity (EDA) pour la collecte de signaux GSR a été proposée, La méthode fonctionne sur le même principe qui consiste à envoyer un courant lumineux dans la peau [Liaqat et al. \(2020b\)](#). Cette méthode propose une solution simple et portable et se basant sur les réseaux de neurones afin d'analyser le signal GSR. D'autres études se basant sur l'analyse du signal d'accéléromètre porté au poignet montrent aussi de bons résultats. [Chun et al. \(2019\)](#) explique qu'une caractéristique essentielle de l'ingestion de liquide est l'ensemble des mouvements effectués par la main, le poignet et le bras pendant l'acte de boire. Dans le cas général où un récipient de liquide (par exemple, une tasse ou un verre d'eau) est posé sur une table, ces mouvements comprennent (1) la saisie du récipient, (2) son élévation vers la bouche, (3) l'inclinaison du récipient de manière à ce que le liquide s'écoule dans la bouche, (4) l'abaissement du récipient jusqu'à sa position initiale sur la table, (5) et enfin son relâchement. Lorsqu'ils sont capturés par des capteurs inertiels placés sur le poignet, ces mouvements constituent un marqueur comportemental stable de l'ingestion de liquide, la figure 3.4 montre cette segmentation du signal selon les cinq phases expliquées. Dans notre étude, nous avons convenu à mesurer uniquement la consommation d'eau et de la détection de cette consommation par l'utilisation des mesures multicapteurs associées à ce risque.

Oubli de prise de médicament

Une grande partie de la population dite fragile est soumise à un traitement médicamenteux. Une mauvaise prise de médicaments a un impact direct sur la santé du sujet qui peut se traduire par des admissions urgentes à l'hôpital et même parfois par des conséquences létales. Les méthodes de surveillance de la prise de médicaments se divisent en deux grandes catégories : les méthodes directes et les méthodes indirectes. Les méthodes directes comprennent l'observation directe du sujet lors de sa prise de médicaments, la détection en laboratoire du médicament dans les fluides

3. De la taxonomie à la mise en place du dispositif d'acquisition

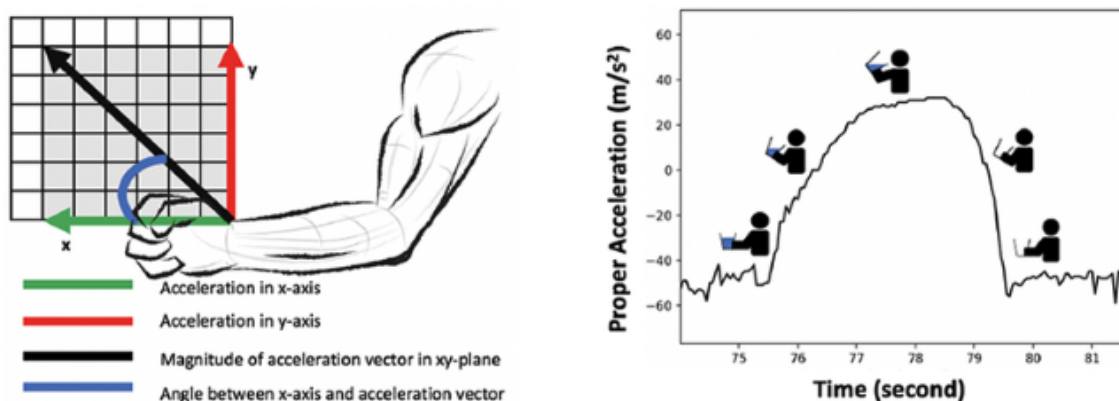


FIGURE 3.4 – À droite : Signaux bruts, mesurés par un accéléromètre placé sur le poignet. À gauche : le schéma d'un geste de prise de liquide du début à la fin. [Chun et al. \(2019\)](#).

biologiques du patient ou dans les biomarqueurs. Les méthodes indirectes quant à elles sont : la déclaration d'un patient, le comptage des pilules, le suivi de l'historique des renouvellements de médicaments et les systèmes de suivi automatique utilisant des caméras ou des capteurs sans fils sous forme de technologies IoT. L'utilisation de ces technologies connaît un grand essor ces dernières années et par conséquent de nombreuses études sur le sujet ont été réalisées proposant des capteurs à faible coût, qui peuvent enregistrer une série d'activités pendant la prise de médicaments et fournir un retour d'information en analysant la sortie des capteurs. En reconnaissant automatiquement les prises de médicaments parmi d'autres activités à l'aide des données collectées par les capteurs, ces systèmes pourraient détecter si un patient a pris ses médicaments dans une fenêtre temporelle souhaitable et pourraient donc être utilisés pour fournir des rappels en cas d'absence de prise de médicaments. Compte tenu de la facilité d'utilisation, il est préférable d'utiliser des capteurs intégrés dans un dispositif facile et léger à porter. [Fozoonmayeh et al. \(2020\)](#), propose une solution avec des capteurs intégrée dans une smartwatch pour la détection de prise de médicament sur des intervalles de temps bien définis. Des signaux d'accéléromètre et de gyroscope mesurés sur le poignet sur sujet 3.5 servent à reconnaître l'action de prise de médicament. Le tableau 3.1 récapitule les différentes situations à risques figurants dans la taxonomies citées ci-dessus. En analysant la taxonomie proposée, il est clair que les différentes situations à risques ne sont pas équivalentes en termes de charge sémantique. Ainsi, les chutes et les déclenchements de chutes peuvent être détectés à partir d'une simple analyse des signaux d'accéléromètres via un seuillage de la magnitude ou des méthodes d'apprentissage automatique [Cao et al. \(2016\)](#). Il s'agit de situations de risques "populaires" bien étudiées dans l'état de l'art. La deuxième catégorie de risques tels que les risques d'oubli majeur, d'accident domestique, d'intrusion d'une tierce personne nécessitent des données beaucoup plus riches, notamment des informations visuelles à enregistrer au domicile du sujet pour les détecter. De plus, ces situations sémantiques constituent un véritable défi : sera-t-il possible de reconnaître ces risques en utilisant uniquement des dispositifs

portables ?

Situations à risques	Scénario
Chute	tomber pour des causes accidentelles
Risque de chute	Physiologique : Accroupi, Assis debout
	Environnemental : Marcher sur un sol glissant, monter des escaliers, trébucher sur des obstacles.
	Actions : Trébucher, soulever (par exemple une boîte) Atteindre le plafond, se précipiter, balayer le sol
Risque d'accident domestique	Brûlures, Utilisation de couteaux, Manipulation d'ustensiles "dangereux", Repassage
Risque d'oubli de prise de médicaments	Sauter ou oublier son traitement
Risque d'intrusion	Quelqu'un qui tient la poignée de la porte d'entrée pendant quelques minutes.
Risque de déshydratation	Manque d'hydratation au quotidien

Tableau 3.1 – Tableau résumant la taxonomie envisagée dans le cadre du projet BIRDS.

3.2.3 Modèle des risques

Comme nous l'avons vu dans les sections précédentes, les risques sémantiques sont partagés entre deux grandes catégories :

- Les risques immédiats ;
- Les risques survenant dans le temps ;

Pour la détection de ces risques, il s'agit de détecter les événements sur l'ensemble des mesures prises par des capteurs portables. Dans le premier cas, l'événement même correspond à la situation à risques. Dans le deuxième cas la non-détection d'un événement, comme par exemple, "prise de médicaments" sur la durée d'un enregistrement journalier représente le risque. Dans les deux cas, il s'agit d'événements qui s'expriment à la fois sur la composante visuelle des données et sur les composantes sensorielles. Ainsi nous considérons un modèle de risques à deux composantes et proposerons dans les chapitres suivants leur détection séparée sur des données visuelles et une détection sur celle des autres capteurs.

Plus formellement, soit $\Delta t = [t_i, t_{i+1}]$ - la longueur de la fenêtre temporelle qui comprend des mesures visuelles I_i et les mesures des capteurs multimodaux X_i . Un événement à risque doit être perceptible par l'ensemble des capteurs de la fenêtre temporelle $W_{\Delta t}$. Dans le présent travail de recherche, nous nous sommes attachés à détecter les événements à risque dans les deux flux - les mesures visuelles (images) et les mesures multimodales de capteurs 1D (signaux). Dans ce qui suit, nous explicitons le choix des mesures des signaux.

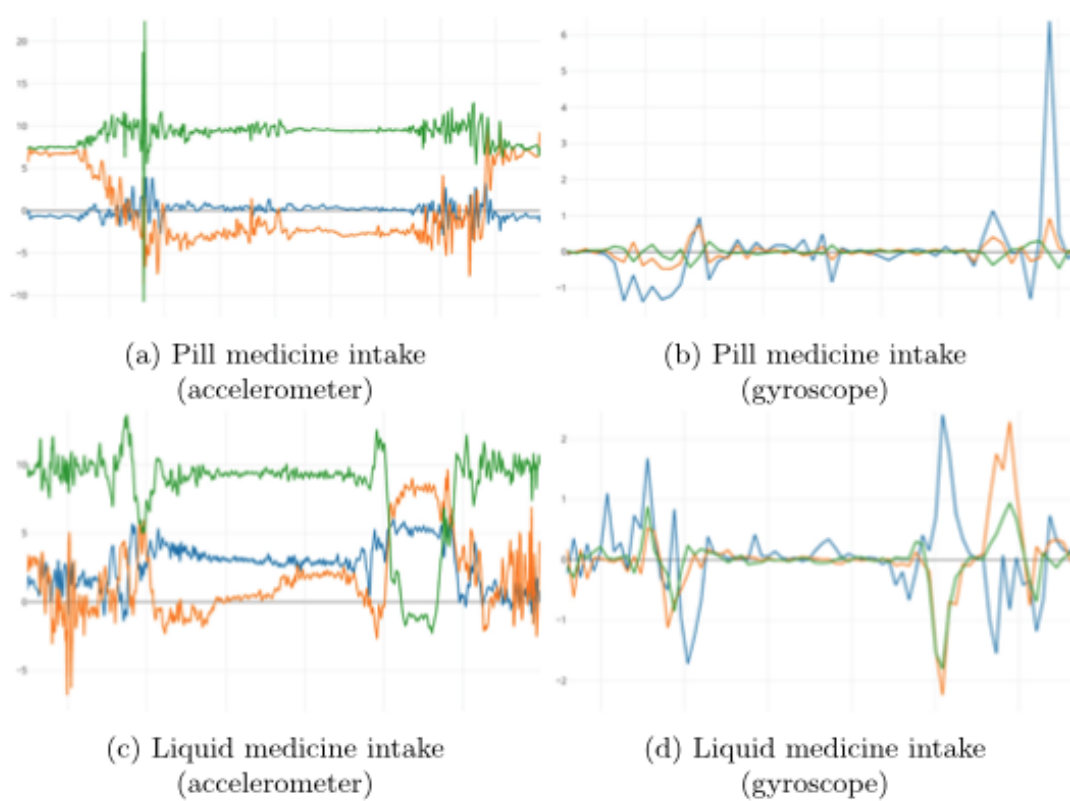


FIGURE 3.5 – Relevés par accéléromètre et gyroscope des activités de prise de médicaments Fozoonmayeh et al. (2020).

3.3 Choix des mesures

Afin de reconnaître les situations à risques mentionnées dans la section 3.2 afin d'assurer une bonne qualité de vie des personnes fragiles, il est utile d'avoir accès à leurs informations physiologiques, motrices et ambiantes en continu. Les informations physiologiques reflètent l'état de physiologique ainsi que l'état émotionnel de l'individu, tandis que les informations motrices aident à identifier sa position et avertir l'individu si l'état immédiat peut représenter un risque pour le sujet fragile.

Enfin, nous avons les informations ambiantes qui peuvent servir à alerter quant aux oublis majeurs et risques d'accidents domestiques [Rajagopalan et al. \(2017\)](#).

Dans ce qui suit, nous allons expliquer plus en détails nos choix de mesures en tenant compte des coûts matériels ainsi que de la taille du dispositif qui doit convenir aux situations écologiques dans lesquelles il va être porté.

3.3.1 Mesures physiologiques

Paramètres vitaux

Les paramètres vitaux tels la température, le pouls (fréquence cardiaque), la fréquence respiratoire et la pression artérielle, sont les signes physiologiques les plus importants qui indiquent l'état de santé d'un individu [Stevenson et al. \(2016\)](#).

Il s'agit souvent des premières informations prises en cas d'urgence et requises pour tout examen de santé. Les paramètres vitaux peuvent être utilisés pour mesurer les fonctions les plus élémentaires du corps pour identifier ou surveiller efficacement les problèmes de santé.

Il a été démontré que des changements rapides ou des différences graduelles dans les mesures vitales sont annonciateurs de situations d'urgence [Giacovelli et al. \(2009\)](#). Les plages normales pour les paramètres vitaux d'une personne varient en fonction de la démographie du sujet, de l'âge, du sexe, du poids, du stress et de l'heure de la journée. Le tableau 3.2 décrit les caractéristiques des différents signes vitaux :

Paramètres extra-vitaux

Plusieurs autres paramètres tels que l'activité musculaire, la glycémie et l'activité électrodermale, peuvent alerter des situations à risques auxquelles sont confrontées les personnes fragiles. Le tableau 3.3, présente les différents paramètres extra-vitaux et leurs caractéristiques [Luo et al. \(2017\)](#), [Soh et al. \(2015\)](#). Pour le choix de nos mesures physiologiques, nous nous sommes basés sur des mesures n'utilisant pas de capteurs intrusifs. Ceci est important pour que le sujet se sente à l'aise avec le dispositif. Nos choix se sont aussi appuyés sur le fait que nous utilisons un dispositif multi-capteurs, les capteurs physiologiques sont essentiellement là pour alerter des changements émotionnels soudains et compléter ses mesures avec les données vidéos.

3. De la taxonomie à la mise en place du dispositif d'acquisition

Signes vitaux	Description	valeurs normales
Température	Varie selon le sexe, l'activité physique, l'heure de la journée et chez les femmes, le stade du cycle menstruel.	Entre 36.5 °C et 37.3 °C. Avec une moyenne de 37 °C
Fréquence cardiaque	Mesure le nombre de battements du coeur par minute. Des mesures inférieures à 60 Beats per Minute (bpm) signifient des risques de bradycardie et des mesures supérieures à 100 bpm signifient des risques de tachycardie	Entre 60 bpm et 100 bpm.
Fréquence respiratoire	Nombre de respirations d'une personne par minute. Les mesures varient selon l'âge et l'état de santé de l'individu.	Entre 12 et 18 respirations par minute pour un adulte et entre 24 et 30 respirations pour un enfant.
Pression sanguine	Mesure la pression du sang dans les artères. Les variations sont souvent soumises à l'âge ainsi qu'à la santé et l'état de stress de l'individu.	120 mm/Hg pour la pression systolique et 80 mm/Hg pour la pression diastolique.

Tableau 3.2 – Les différentes caractéristiques des signes vitaux.

3.3. Choix des mesures

Signes extra-vitaux	Description	valeurs normales
Glycémie	Représente la quantité de glucose circulant dans le sang. Les valeurs normales fluctuent entre 70 mg/dl et 100 mg/dl	La valeur normale moyenne est de 1 g/dl chez l'adulte non-diabétique
Activité musculaire	Souvent utilisée dans le cas de maladies neuro-musculaires et neuro-dégénératives	Valeurs variables selon les situations de repos et de contraction musculaire
Niveau de CO_2	Informant sur la quantité de dioxyde de carbone dans le sang. Le déséquilibre entre l'oxygène et le dioxyde de carbone dans le sang humain est l'indication d'un trouble rénal, respiratoire ou métabolique.	La plage normale pour le CO_2 dans le sang humain est d'environ 23 à 29 mEq/L
Activité électrique cérébrale	Mesure l'activité neuronale. La modification anormale de l'activité neuronale peut entraîner des crises d'épilepsie. De plus, l'information est utilisée pour identifier les problèmes de santé mentale et les problèmes physiques tels que les problèmes au niveau du cerveau, de la moelle épinière ou du système nerveux des sujets	Diffère d'un individu à un autre.
Activité électrodermale	Mesure les changements du processus émotionnel non conscient d'un individu à travers la mesure des sécrétions de ses glandes sudoripares.	Diffère d'un individu à un autre.

Tableau 3.3 – Les différentes caractéristiques des signes vitaux.

Pour cela, les mesures physiologiques auxquelles nous nous sommes intéressées sont la fréquence cardiaque, l'activité électrodermale et la température corporelle. Les spécificités de ces mesures sont décrites dans le tableau 3.2 et 3.3. Il est aussi important de mentionner que le but de notre projet n'étant pas la conception de dispositif de détection de situations à risques mais leur détection, nous nous sommes contraints à utiliser des solutions préexistantes sur le marché.

3.3.2 Mesures motrices

La détection de chute est une des thématiques les plus abordées dans la littérature scientifique concernant les technologie d'aide au quotidien des personnes âgées. Les publications actuelles peuvent être classées en deux grandes catégories : (1) celles utilisant des technologies portable type IoT avec des mesures motrices (accélération et/ou vitesse angulaire, champs magnétique) et caméras portées et celles s'appuyant sur une technologie non portable notamment la vidéo (caméras fixes ou capteurs Kinect) ou des capteurs de pression au sol intelligents [Minvielle et Audiffren \(2019\)](#). Nous nous concentrons dans ce travail sur les technologies portables fournissant plus d'autonomie aux sujets fragiles et n'ayant pas besoin d'installation particulière. Les mesures motrices sont importantes dans le sens où elles nous fournissent des informations sur l'état physique du sujet. Nous pouvons par ces capteurs savoir si le sujet est en train de marcher, de monter des escaliers ou de s'accroupir. Cela nous indique si des situations à risques sont à envisager selon ses signaux.

Les mesures motrices sont essentielles pour la détection des chutes accidentelles. Celles utilisées le plus souvent dans le cas de la détection de chutes sont l'accélération [Bourke et Lyons \(2008a\)](#), la vitesse angulaire et le champs magnétique [Bourke et Lyons \(2008b\)](#). La combinaison de ces trois mesures est ce qu'on appelle l'unité de mesure inertielle.

3.3.3 La vidéo

Pour alerter quant aux oublis majeurs et risques d'intrusion, la prise en compte de la vidéo dans nos enregistrements pour compléter les données, nous a paru essentielle. Nous avons choisi d'utiliser une Camera portée donc produisant une vidéo égocentrique nous permettant de voir ce que le sujet voit. Cette approche est utilisée dans plusieurs études. Nous pouvons citer la collecte de données Lifelog [Bolaños et al. \(2017\)](#), la détection précoce de troubles dans le cadre de la maladie d'Alzheimer [Karaman et al. \(2014\)](#), [Chaabouni et al. \(2016\)](#).

Dans les vidéos enregistrées, des caméras vidéo égocentriques [Karaman et al. \(2014\)](#) sont utilisées dans plusieurs études, nous pouvons citer la collecte de données Lifelog [Bolaños et al. \(2017\)](#), la détection précoce d'une maladie d'Alzheimer [Karaman et al. \(2014\)](#), [Chaabouni et al. \(2016\)](#). Elle a au fur et à mesure remplacé l'identification des objets par des signaux RFID, les progrès réalisés dans la reconnaissance d'objets à partir de données visuelles ont permis aux chercheurs d'utiliser

des détecteurs d'objets visuels. L'utilisation de modèles de plus haut niveau permet la détection de scénarii complexes avec des vues sémantiques, telles que la présence d'objets liés à l'activité et la localisation de la personne dans certaines zones [Yebda et al. \(2020\)](#). En effet, les vidéos égocentriques donnent non seulement une information sur le mouvement quand la fréquence d'acquisition le permet mais elle fournit aussi des informations sur l'environnement pouvant aider à la reconnaissance de risques considérés comme risques sémantiques.

3.3.4 Choix de l'emplacement des capteurs

Emplacement des capteurs pour l'acquisition des données motrices Comme les informations motrices varient selon les différentes parties du corps humain pendant les mouvements, l'emplacement de l'appareil de mesure est très important, il reflète les caractéristiques clés du processus de chute. En effet, comme indiqué dans [Tong et al. \(2013\)](#), [Purwar et al. \(2007\)](#), [Bourke et Lyons \(2008c\)](#), [Tong et al. \(2009\)](#), le bras, le poignet, la hanche et la jambe ne sont pas les positions appropriées pour récupérer les informations motrices, en raison de leur fréquence de mouvement élevée et de leur complexité, bien que souvent plus confortable. La région la plus appropriée pour distinguer les chutes des autres mouvements avec les informations motrices est située au niveau du plexus solaire. Ce dernier subit les mouvements globaux du sujet, il n'est pas perturbé par les mouvements du reste du corps.

Emplacement des capteurs pour l'acquisition des données physiologiques Les données physiologiques nécessaires à la détection des situations à risques fixées dans les objectifs du projet sont essentiellement la température corporelle, la réponse électrodermale (EDA) et la fréquence cardiaque.

D'après de nombreuses études [Taylor et Machado-Moreira \(2013\)](#), l'endroit optimal pour la prise de mesures de l'activité électrodermale est la paume ou les doigts d'une main pour la forte concentration en glandes eccrines responsables des sécrétions en sueur du corps humain. Toutefois, des mesures comparatives de l'EDA prises au poignet et sur les doigts au même moment [Poh et al. \(2010\)](#), prouvent que l'erreur entre les deux résultats est tolérable. La mise en place des électrodes sur les doigts ou les paumes nécessite des connexions externes ce qui rendrait le dispositif encombrant et très sensible aux artefacts de mouvement et/ou de pression.

Pour des conditions de non-intrusivité, de compatibilité aux situations écologiques dans lesquelles le dispositif va être porté et en tenant compte des publications scientifiques sur la précision des mesures physiologiques [Mukhopadhyay \(2015\)](#), [Patel et al. \(2012\)](#), nous avons opté pour une solution multi-capteurs connectée au poignet.

3.4 Technologies d'acquisition et de transmission pour la solution connectée

Les progrès récents dans les domaines de la technologie des capteurs, de la micro-électronique et des télécommunications ont considérablement aidé au développement de solutions connectées. Les informations physiologiques, motrices, ambiantes ou de localisation, peuvent être facilement collectées.

Les données collectées peuvent être analysées et transmises aux contrôleurs à travers les différentes technologies de communication sans fil existantes dans les technologies smartHealth (Near Field Communication (NFC), WiFi, Bluetooth, ZigBee...)

3.4.1 Technologies capteurs

Les capteurs sont les éléments de base des systèmes d'acquisition de données. Ce sont des dispositifs électroniques transformant l'état d'une grandeur physique observée en une grandeur utilisable. Ils jouent un rôle essentiel dans le domaine des objets connectés. Les signaux analogiques correspondant aux informations physiologiques du corps, aux activités humaines et aux informations ambiantes sont collectés puis convertis en valeurs numériques pour les transmettre finalement aux émetteurs radio. Comme ces capteurs communiquent directement avec l'homme ou sont même implantés, la taille des capteurs et de l'électronique frontale est devenue une des principales préoccupations des spécialistes du domaine.

La miniaturisation des circuits électroniques basée sur l'utilisation de la microélectronique a joué un rôle essentiel dans le progrès de la technologie des capteurs. En outre, le développement du traitement du signal et surtout des systèmes micro-électromécaniques (Micro Electro Mechanical Systems (MEMS)) permet de surveiller les données sensorielles de manière non invasive. Les MEMS sont une technologie avancée pour la conception de capteurs basés sur des composants mécaniques et électromécaniques miniaturisés (c'est-à-dire des dispositifs et des structures) qui sont fabriqués en utilisant les techniques de microfabrication. En conséquence, les capteurs deviennent de plus en plus petits, de l'ordre de 1 à 100 micromètres. En effet, une nouvelle possibilité d'applications de soins de santé omniprésentes a été créée, avec des coûts médicaux réduits, une indépendance, des services de soins de santé plus confortables et de haute qualité. La technologie MEMS a été appliquée à la conception de différents types de capteurs tels que l'accéléromètre, la glycémie, la pression sanguine, le capteur de gaz carbonique (CO₂), l'ECG, l'EEG, l'EMG, le gyroscope, l'oxymétrie de pouls, ainsi que certains capteurs généralement utilisés dans les Wireless Sensor Networks (WSNs) récemment. Nous allons maintenant présenter quelques capteurs couramment utilisés dans les technologies eHealth et répandues dans les IoT et correspondant aux choix de mesures expliqués dans la section précédente. Il est possible de distinguer trois classes de capteurs : les capteurs physiologiques, les capteurs inertiels et les capteurs contextuels. Les capteurs

3.4. Technologies d'acquisition et de transmission pour la solution connectée

Capteur	Mesure	Champs d'intérêt
Capteurs physiologiques		
Activité électrodermale (EDA)	Mesure la capacité ou la résistivité du sujet	Mesure l'état de stress du sujet
Électrocardiographe (ECG) Photoplethysmogramme Photoplethysmogramme (PPG)	Mesure la fréquence cardiaque du sujet	Mesure les risques de tachycardie ou de bradycardie du sujet
Électroencéphalogramme (Électroencéphalogramme (EEG))	Mesure l'activité électrique du cerveau	Détection des risques épileptiques du sujet
Sphygmomanomètre	Mesure la pression sanguine	Permet la détection des situations d'hypertension ou d'hypotension artérielle
Electromyographe (Électromyographie (EMG))	Mesure l'activité musculaire du sujet	Permet la détection de chutes ou d'accidents neuro-musculaires.
Thermomètre	Mesure la température corporelle du sujet	permet de détecter des augmentations anormales de la température corporelle
Glucomètre	Mesure la quantité de glucose dans le sang du sujet	Détection d'hypoglycémie ou d'hyperglycémie
Capteurs inertiels et contextuels		
Accéléromètre triaxial	Mesure l'accélération du sujet sur les 3 axes (x,y,z)	Dans le même esprit que le gyroscope, ils servent à la détection de chute ou de mouvement du corps
Gyroscope triaxial	Mesure le moment angulaire du sujet sur les 3 axes (x,y,z)	/
Magnétomètre	Mesure le champs magnétique du sujet	Détection de l'orientation
Microphone	Mesure le signal acoustique ambiant	Détection des son de détresse
Caméra	Mesure les rayonnements électromagnétiques	Tentative d'intrusion, détection de risque de chute (environnemental et dû à une action) etc.

Tableau 3.4 – Les capteurs contextuels, inertiels et physiologiques les plus couramment utilisés dans les objets connectés.

3. De la taxonomie à la mise en place du dispositif d'acquisition

inertiels sont constitués de capteurs de mouvements (Accéléromètre Gyroscope et magnétomètre), capteurs de position Global Positioning System (GPS) et capteurs de mesures ambiantes (capteurs audio et video). Des capteurs appartenant à ces catégories sont présentés dans le tableau 3.4.

Les capteurs physiologiques capturent les signes vitaux ou extra-vitaux du sujet tels que la température, la fréquence cardiaque, la fréquence respiratoire et la pression artérielle ou l'électrocardiographie (ECG), le taux de glucose dans le sang, etc.

La Figure 3.7 représente l'accélération verticale gauche et verticale droite d'un corps humain. La Figure 3.6 illustre un signal représentant une réponse électrodermale sur trois plages temporelles (5 heures, 30 minutes, 4 minutes).

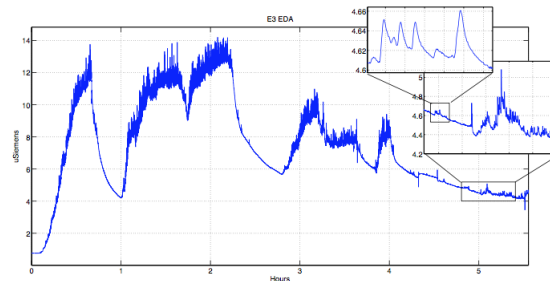


FIGURE 3.6 – Réponse électrodermale sur 3 plages temporelles (Figure extraite de Drover et al. (2017)).

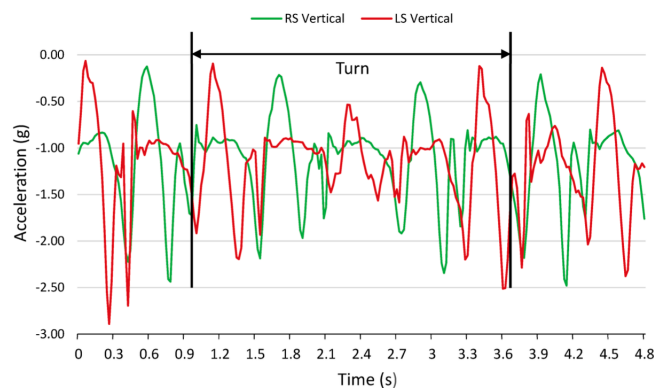


FIGURE 3.7 – Accélération verticale gauche et l'accélération verticale droite d'un corps humain en mouvement (Figure extraite de Garbarino et al. (2014)).

3.4.2 Les objets connectés et la connectivité

Les objets connectés sont en train de révolutionner notre vie quotidienne par la simplification d'accès à des informations que notre corps mesurait et interprétait sans pour autant avoir le recul nécessaire pour comprendre le chaînon qui va de la

collecte à l'analyse de ces mesures. Rendre ces mesures accessibles et analysables par l'homme et la machine, c'est pourvoir intervenir et aider au processus de décision. L'internet des objets (IoT) est appelé à devenir l'un des développements technologiques clés de notre époque, à condition que nous soyons en mesure d'exploiter pleinement son potentiel. Les domaines d'application de l'IoT sont multiples : maisons intelligentes, systèmes d'alerte, achats intelligents, gadgets intelligents, villes intelligentes, routes intelligentes, soins de santé, systèmes de détection d'incendie, systèmes d'identification des menaces, suivi et surveillance.

L'objectif principal de ces objets IoT, dont le nombre et les types ne cessent d'augmenter, est de produire des données utiles sur notre environnement afin de le rendre plus intelligent donc plus compréhensible par l'homme et la machine. Ceci est réalisé en fournissant aux environnements l'accès aux informations dont ils ont besoin par la collecte et l'analyse des données passées, présentes et futures. Ces données permettent de prendre des décisions optimales sur nous et nos environnements, éventuellement en temps réel. Les objets connectés produisent actuellement un volume important de données. Ces données sont produites par divers capteurs, elles sont ensuite transmises par des moyens de transmission sans fil sur des concentrateurs de données optimisant l'analyse de ces données par la fusion. Les objets connectés forment une famille extrêmement diversifiée, qui ne cesse de s'agrandir : domotique, maintenance prédictive, téléphonie, transfert et traitement de données, etc. Le choix de la technologie de télécommunication est fait en fonction d'un certain nombre de critères techniques tels que la portée, le débit et l'autonomie.

La multiplicité des réseaux de communication pose la question de leur compatibilité avec nos besoins. Dans ce qui suit, nous présenterons quelques solutions de communications sans fil pour les objets connectés à courte, moyenne et longue portée. Puis dans un deuxième temps, nous expliquerons notre choix pour le Bluetooth à faible consommation Bluetooth Low Energy (BLE).

Les technologies de courte portée

Bluetooth Le protocole Bluetooth est un standard de transfert de données sans fil [Hossain et Soh \(2007\)](#). Il utilise une bande de fréquence de 2.4 GHz, ce qui permet la transmission de petites quantités de données sur des courtes distances.

Zigbee Zigbee est un protocole de communication radio maillé, développé spécialement pour les applications domotiques. Cette technologie a pour but la communication de courtes distances avec une faible consommation en énergie à l'instar du protocole Bluetooth. Le principal intérêt du protocole Zigbee est sa très faible consommation en énergie, les nœuds sont conçus pour fonctionner plusieurs mois jusqu'à plusieurs années en autonomie complète grâce à une simple pile alcaline de 1,5 V [Gislason \(2008\)](#).

Le protocole NFC Le protocole NFC, ou protocole de communication en champs proche, est une technologie de communication sans fil à courte portée et à

haute fréquence. Il permet une communication directe entre deux appareils électroniques. Il est par exemple utilisé sur les supports d'abonnement de réseau de transport en commun ou par certaines entreprises pour les badges d'accès [Coskun et al. \(2013\)](#).

Les technologies à moyenne portée

Bluetooth Low Energy Le BLE est un protocole de réseau personnel sans fil à très basse consommation [Bronzi et al. \(2016\)](#). Comme la technologie Bluetooth originelle, le BLE ne permet de transférer qu'une quantité limitée de données à une distance moyenne de 60 mètres. La différence entre les dispositifs Bluetooth et BLE se situe au niveau de la consommation électrique nécessaire à la communication, qui est dix fois moindre pour BLE [Mackensen et al. \(2012\)](#).

Z-wave Le Z-Wave est un protocole de communication sans fil principalement dédié à la domotique [Gomez et Paradells \(2010a\)](#). Il permet une transmission de données avec des distances allant de 30m en intérieur à 100 m en plein air. Il fonctionne en réseau maillé, chaque appareil connecté pouvant relayer les informations émises par ses voisins, ce qui lui permet d'élargir sa portée. Tout comme le protocole Zigbee, l'utilisation de Z-Wave ne nécessite que très peu de puissance et les appareils peuvent donc communiquer pendant plusieurs années avec une simple pile.

Le protocole Wi-Fi Le Wi-Fi est un ensemble de protocoles de communication sans fil. Fonctionnant avec des ondes radio dans une bande de fréquence de 2,4 ou 5 GHz, il permet de relier des équipements informatiques et de téléphonie mobile dans un réseau sans fil haut débit [Gomez et Paradells \(2010b\)](#).

Les technologies à longue portée

Réseaux radio bas-débit Sigfox et LoRa sont deux protocoles de communication radio. Ils conviennent à des utilisations sur appareils à basse consommation, dotés ainsi d'une grande autonomie, qui transfèrent une faible quantité de données. Ils sont particulièrement conçus pour la transmission des données des objets connectés, sur des distances allant à 10 km en ville et 50 km en milieu rural [Nolan et al. \(2016\)](#).

Réseaux cellulaires mobiles Les réseaux téléphoniques sont des réseaux de communications sans fil, permettant la transmission de grandes quantités de données à très longues distances. Après la technologie Global System for Mobile Communications (GSM), on a vu lui succéder la 3G, la 4G puis la 5G.

Réseaux propriétaires Un protocole propriétaire est un protocole de communication dont les spécifications sont privées, ils sont notamment développés par des grands groupes.

La Figure René et Alexandre, donne un aperçu des portées et des débits possibles avec chaque technologie.

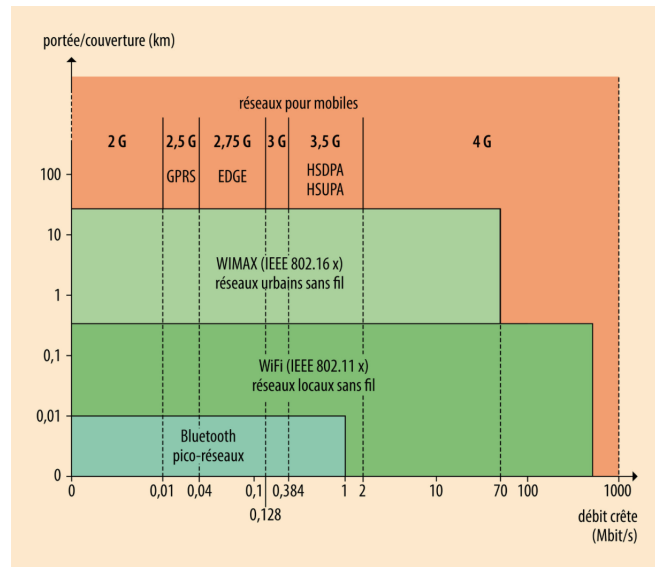


FIGURE 3.8 – Aperçu des portées et des débits possibles avec chaque technologie de transmission. (Figure extraite de René et Alexandre).

Le tableau 3.5 présente un comparatif des différentes technologies de transmission actuelles Majumder et al. (2017)

3.5 Solution retenue

Dans cette section, nous présentons un schéma de collecte de données répondant à notre taxonomie et aux technologies IoT existantes sur le marché. Le dispositif se veut facile à porter pour permettre le plus d'ergonomie possible nous permettant ainsi une meilleure adhésion des sujets de notre cohorte. Pour la mise en place de notre dispositif, nous sommes passés par deux solutions, la première ayant plusieurs limitations logicielles et matérielles, nous avons décidé de passer à une autre solution plus optimisée et avec des technologies plus récentes et plus adaptées à nos objectifs de recherche.

3.5.1 Architecture du dispositif

Pour notre projet, nous tirons parti des technologies existantes que nous complétons avec une architecture s'appuyant sur la fusion de données. Nombreux sont les systèmes où l'intelligence artificielle et les objets connectés sont utilisés ensemble pour offrir des solutions smartHealth. Ci-dessous sont citées quelques solutions existantes sur le marché :

3. De la taxonomie à la mise en place du dispositif d'acquisition

Wireless Tech.	Frequency	Range	Data Rate	Power (mW)	Maximum Nodes	Network Topologies	Security
RFID	13.56 MHz 860–960 MHz	0–3 m	640 kbps	200	1 at a time	peer-to-peer (P2P) passive	N/A
Bluetooth	2.4–2.5 GHz	1–100 m	1–3 Mbps	2.5–100	1 M + 7 S	P2P, star	56–128 bit key
BLE	2.4–2.5 GHz	1–100 m	1 Mbps	10	1 M + 7 S	P2P, star	128-bit AES
HomePlug GP	1.8–30 MHz	~100 m	4–10 Mbps	500	-	P2P, star, tree and mesh	128-bit AES
EnOcean	902, 928, 868 MHz	30–300 m	125 kbps	~0.05 with energy harvesting	-	P2P, star, tree and mesh	128-bit AES
ZigBee	2.4–2.5 GHz	10–100 m	250 kbps	50	65,533	P2P, star, tree and mesh	128-bit AES
WiFi	2.4–2.5 GHz	150–200 m	54 Mbps	1000	255	P2P, star	WEP,WPA, WPA2
DASH7	315–915 MHz	200 m–2 km	167 kbps	<1	-	P2P, star, tree and mesh	128-bit AES
Insteon	RF: 869.85, 915, 921 MHz powerline: 131.65 KHz	40–50 m	38 kbps (RF) 2–13 kbps (powerline)	-	64,000 nodes per network	P2P, star, tree and mesh	256-bit AES
Sigfox	868/902 MHz	10–50 km	10–1000 bps	0.01–100	-	P2P, star	No default encryption
NFC	13.56 MHz	5 cm	424 kbps	15	1 at a time	P2P	AES
Wireless HART™	2.4 GHz	50–100 m		10	-	P2P, star, tree and mesh	128-bit AES
6LoWPAN	2.4 GHz	25–50 m	250 kbps	2.23	-	P2P, star, tree and mesh	128-bit AES
ANT	2.4–2.5 GHz	30 m	20–60 kbps	0.01–1	65,533 in one channel	P2P, star, tree and mesh	64-bit key
Z-Wave	860–960 MHz	100 m	9.6–100 kbps	100	232	mesh	128-bit AES

AES: Advanced Encryption Standard.

Tableau 3.5 – Technologies de communication pour les maisons intelligentes (Tableau extrait de [Majumder et al. \(2017\)](#)).

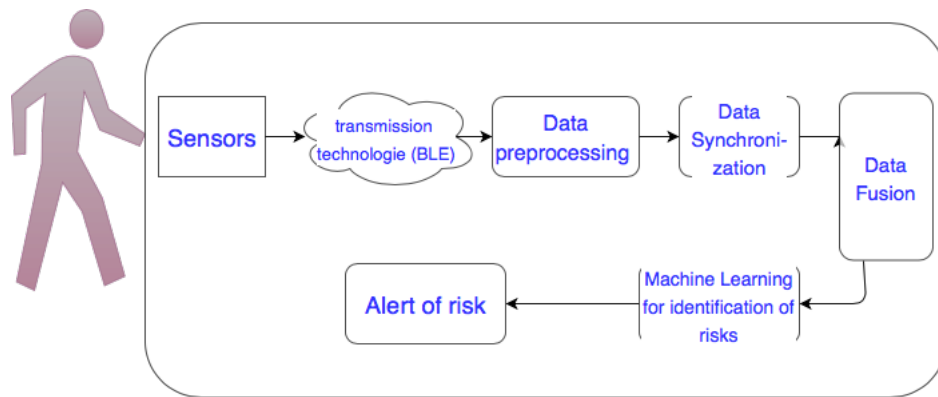


FIGURE 3.9 – Architecture du système pour le projet BIRDS.

- La détection de chute à l’aide de capteurs contextuels [Assistance](#); [Vigeo](#) ;
- La détection précoce des maladies chroniques et la possibilité de retarder ou de prévenir leur progression [Lucine](#) ;
- La compréhension du comportement individuel, des traitements médicamenteux et de l’efficacité des traitements sur les activités de la vie quotidienne [Meek](#) ;
- Des soins personnalisés basés sur l’ADN du sujet, les diagnostics médicaux et les tendances de ses soins de santé par rapport à ses besoins.

La Figure 3.9, représente l’architecture du système de détection des situations à risques envisagé pour notre projet.

3.5.2 Premier prototype

Dans cette sous-section, nous verrons les composants du premier prototype de la solution connectée proposée pour l’acquisition des données. Cette solution n’a finalement pas été retenue à cause des limitations techniques et logicielles qu’elle présentait.

Principe du système

La Figure 3.10 présente les capteurs composant le dispositif d’acquisition de surveillance de l’activité physiologique, motrice ainsi que du milieu ambiant du sujet. Le prototype développé est composé de plusieurs éléments :

- Camera GoPro Hero : permettant le support du système de capture vidéo ;
- Un bracelet connecté Microsoft Band 2 : permet la capture des données physiologiques et motrices du sujet ;
- Un MetaMotionC Pro : dispositif d’acquisition des mesures motrices du sujet ;
- Un smartphone : capture de son et de position GPS, il permet également l’enregistrement des mesures des différents capteurs. Il est considéré comme le contrôleur principal du dispositif.

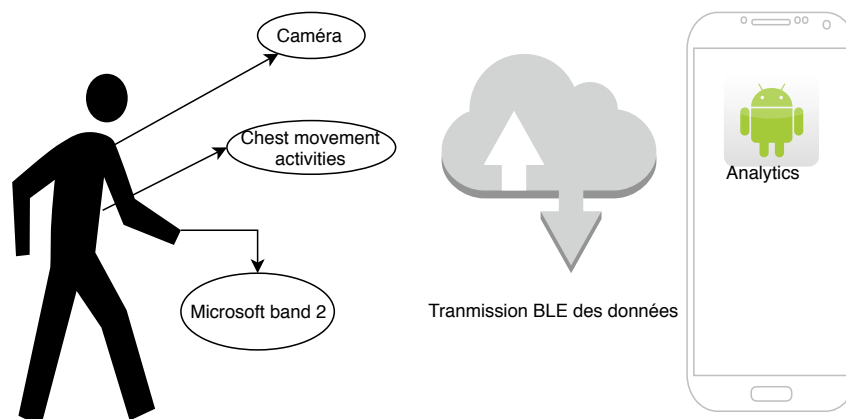


FIGURE 3.10 – Dispositif d'acquisition des données capteurs.

Les détails de chacun des composants de notre solution sont expliqués dans ce qui suit.

Microsoft Band 2

Pour la collecte des données physiologiques plusieurs solutions se sont offertes à nous, notamment le bracelet Empatica E3 [Garbarino et al. \(2014\)](#), Microsoft Band 2 [Vhaduri et Poellabauer \(2017\)](#), Apple Watch [Raja et al. \(2019\)](#), etc.

Pour des questions de coûts, de fonctionnalité et de son applicabilité aux projets scientifiques smarthhealth, nous avons porté notre choix sur le Microsoft Band 2.

Présenté comme une évolution du Microsoft Band, le microsoft Band 2 (cf Figure 3.11), est une montre connectée développée par Microsoft. C'est un dispositif multi-capteurs sans fil portable conçu pour une acquisition de données en temps réel et en continu. Le Microsoft Band 2 est composé de 12 capteurs :

Accélération : donne l'accélération X, Y et Z en unités g ($1\text{ g} = 9,81\text{ m/s}^2$) ;

Gyroscope : donne la vitesse angulaire en X, Y et Z en unités $^\circ/\text{s}$;

Thermopile : indique la température actuelle de la peau du porteur en degrés Celsius ;

Photoplethysmographie (PPG) : indique le nombre de battements par minute ;

Calorimètre : indique le nombre total de calories brûlées par l'utilisateur ;

Réponse électrodermale (EDA) : fournit la résistance cutanée de l'utilisateur en Ohms ;

Photo-détecteur : fournit l'intensité lumineuse ambiante (éclairage) en lux 2 Hz (Lumes par mètre carré) ;

Baromètre : fournit la pression atmosphérique brute ambiante en hPa (hectopascals) ;

Altimètre : récupère les données altimétriques courantes comme le gain total/perde totale ;

Capteur de contact : Fournit l'état actuel de la bande comme portée ou non portée.

Les caractéristiques techniques des capteurs Microsoft Band 2 sont indiquées dans le tableau 3.6.

Capteur	Unité	Fréquence	Composition
Accéléromètre	m/s^2	62/31/8 Hz	Accélération 3 axes x,y et z
Gyroscope	$^{\circ}/s$	62/31/8 Hz	vitesse angulaire sur 3 axes x,y et z
Thermopile	$^{\circ}C$	1 Hz	thermopile infrarouge
Photoplethysmographie	bpm (Beats per minute)	1 Hz	Photo-diodes
EDA	Ohm	0.2/5 Hz	Électrodes
Photodecteur	Lux (Lumes per meter)	2 Hz	/
Baromètre	hPa	1 Hz	/
Calorimètre	Kcals	inconstante	/

Tableau 3.6 – Cactéristiques techniques du Microsoft Band 2.

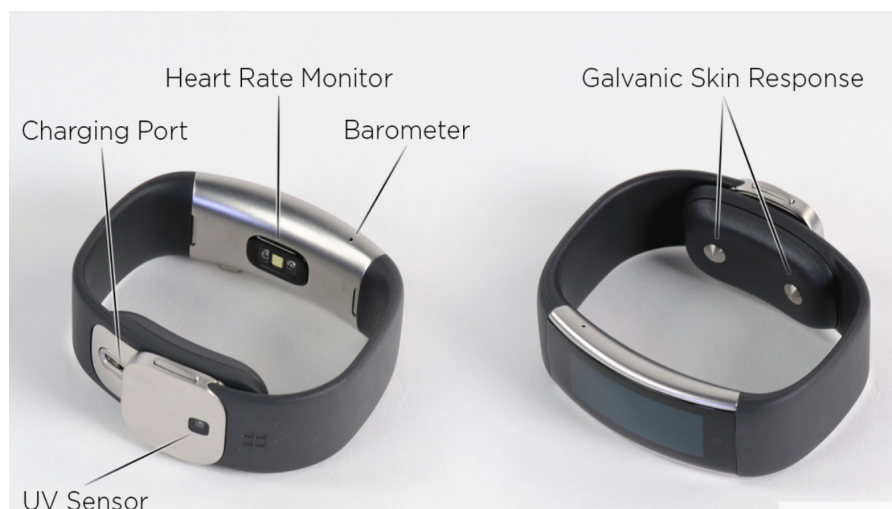


FIGURE 3.11 – Microsoft Band 2.

MetaMotionC Pro

MetaWearC Pro (cf Figure 3.12) est un dispositif multi-capteur connecté. Il est doté d'un BLE pour la transmission des données, ce qui permet une communication par smartphone et un traitement centralisé. MetaWearC Pro intègre avec son système de radio transmetteur des capteurs pour une architecture centralisée à pile. Tous les circuits ont été conçus en tenant compte de l'efficacité énergétique. Le MetaMotionC Pro comprend une multitude de capteurs accéléromètre, gyroscope, magnétomètre, capteur d'humidité, capteur de lumière, baromètre, thermomètre, etc.

Les besoins du projet requièrent l'acquisition des 3 mesures motrices au niveau de la poitrine (section 3.3). Le dispositif MetaMotionC Pro sera exclusivement utilisé dans le but d'acquérir ces mesures. Les mesures prises en compte sont l'accélération, la vitesse angulaire et le champs magnétique. Nous retrouvons dans le tableau 3.7, les spécifications techniques des capteurs impliqués dans ces prises de mesures.

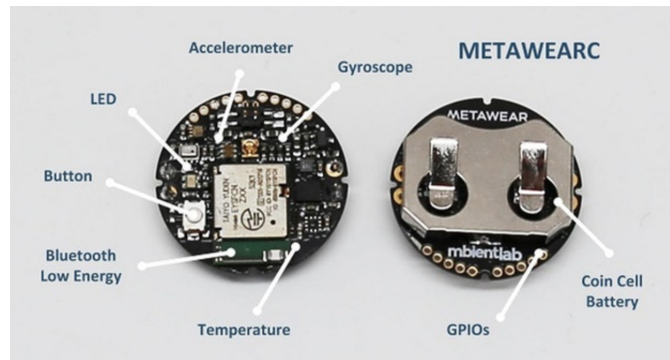


FIGURE 3.12 – MetaWearC Pro.

Caméra

Pour l'acquisition de la vidéo, plusieurs solutions connectées existent. Une de nos exigences sur la caméra est due à la nature particulière de notre dispositif qui intègre une caméra portée, cela entend la discrétion ainsi que sa commodité pour le quotidien des sujets. Avant de choisir un modèle de caméra composé uniquement d'une lentille et connecté par câble au téléphone, nous avons étudié d'autres solutions telles que : la ION Snapcam, la GoPro Hero(2018), la FrontRow FR et la YoCam. Pour des questions de coûts parfois élevés de certaines caméras, l'absence de SDK en libre accès pour d'autres, ou encore la faible qualité du streaming, notre choix s'est porté sur un modèle de camera constitué uniquement d'une lentille facilitant ainsi la portabilité du dispositif.

La quantité de données étant conséquente avec la vidéo, une lenteur dans l'envoi des données par le bluetooth a été constatée. Nous avons donc décidé d'utiliser une caméra avec une connexion filaire avec le smartphone. Nous nous sommes alors tournés vers un modèle de webcam classique incluant uniquement une lentille,

Capteur	Mesure	Unité	Fr d'acquisition min	Fr d'acquisition max	Sensibilité	Résolution
Accéléromètre	Accélération	m/s^2	12.5 Hz	800 Hz	$\pm 2, \pm 8, \pm 16$	2048 \sim 16384 counts/g
Gyroscopie	Vitesse laire	ang- /s	25 Hz	1600 Hz	$\pm 125, \pm 2000$	16 \sim 262 counts/°
Magnétomètre	Champs magné- tique	T (tesla)	10 Hz	300 Hz	$\pm 1300\mu T$ (x, y- axis), $\pm 2500\mu T$ (z axis)	$\pm 0.3\mu T$

Tableau 3.7 – Caractéristiques techniques du MetaMotionC Pro.

3. De la taxonomie à la mise en place du dispositif d'acquisition

figure 3.13. La vidéo sert à compléter les données physiologiques et les données de mouvement afin de mieux reconnaître les situations à risque. Pour cela la caméra n'enregistre que 3 secondes toutes les 10 secondes. La vidéo est encodée sur le smartphone à partir de l'application sous format MP4.



FIGURE 3.13 – Présentation de la caméra retenue pour la solution finale.

3.5.3 Limitations du premier prototype

Après plusieurs tests effectués sur le dispositif, nous avons relevé plusieurs limitations de notre première solution. Les données manquantes des capteurs constituent le principal problème des données des capteurs 1D.

Le flux de données est le flux continu de données mesurées depuis un capteur vers le réseau. Pour de nombreuses raisons, tels qu'une coupure de courant au niveau du nœud du capteur, un taux d'erreur binaire élevé des transmissions radio sans fil, ce flux de données du capteur peut être perdu ou corrompu. L'activité humaine sur les lieux de déploiement des capteurs peut être considérée comme des occurrences aléatoires d'interférences locales qui sont la raison de la perte de données des capteurs. Parallèlement, le taux d'échantillonnage du capteur est également compromis, ce qui entraîne le problème des données manquantes du capteur.

La synchronisation des données est une autre difficulté des données des capteurs. Sur la base de leur horloge interne, les capteurs transmettent les données aux contrôleurs. Mais, les horloges de ces capteurs ne sont souvent pas synchronisées entre elles ; c'est un autre défi dans la tâche d'analyse des données entre capteurs. De plus, en raison de la consommation d'énergie, le taux d'échantillonnage est différent pour chaque type de capteur.

Le dernier problème rencontré, a été la fin du support des applications et services du tableau de bord Microsoft Health, ce qui rendait notre dispositif presque obsolète. Cette situation est souvent rencontrée sur les technologies healthcare destinées au grand public. Ceci est dû au fait que ce sont des technologies qui n'ont pas bénéficié d'études préalables sur les interactions homme machine.

3.5.4 Implémentation de la solution finale

Dans un deuxième temps, sur la base de l'architecture générale du système telle que présentée dans la Figure 3.9, nous réalisons la mise en œuvre du système de détection de situations à risques pour le projet BIRDS. La figure 3.9 illustre la procédure de surveillance des situations à risques qui peut être résumée comme suit.

Le dispositif composé d'un ensemble d'objets connectés : (1) un dispositif connecté MetamotionRL qui mesure l'activité physique du sujet par des capteurs inertiels accéléromètre, gyroscope et magnétomètre (2) un bracelet connecté Empatica E4 est utilisé pour capturer les différents signaux physiologiques tels que Blood Volume Pressure (BVP), InterBeat Interval (IBI), Heart Rate (HR), température, l'activité électrodermale et les chiffres de l'accéléromètre du patient à partir du poignet (3) une caméra portée à l'épaule et branchée par une connexion filaire au smartphone pour un transfert optimal des données en temps réel.

Grâce à la technologie BLE, les données collectées sont transférées au smartphone Android qui est le contrôleur principal du dispositif. L'application BIRDS développée par notre équipe reçoit les données des différents dispositifs et les synchronise pour un meilleur traitement par la suite.

Ces données sont destinées dans un premiers temps à entraîner nos réseaux de neurones pour la détection des situations à risques. Ces données sont donc annotées suivant le journal de bord du sujet enregistrant ces données. S'en suit une étape de pré-traitement pour la normalisation et l'estimation des données manquantes. La dernière étape est le traitement de ces données pour une détection des situations à risques.

Les données pré-traitées ont été analysées dans la dernière étape de l'implémentation. Le résultat fourni par le système de traitement mis en œuvre va servir par la suite pour une implémentation du système sur l'application finale BIRDS.

Les détails de chacun des composants de notre processus de mise en œuvre sont expliqués dans ce qui suit.

Bracelet Empatica E4 Wearable

Présenté par Empatica, le bracelet Empatica E4 [Empatica \(2021\)](#) est un dispositif multi-capteurs sans fil portable conçu pour une rétroaction biologique informatisée et une acquisition de données continue et en temps réel. Le bracelet E4 présenté dans la figure 3.14 est équipé de quatre capteurs intégrés :

Un capteur de photopléthysmographie PPG pour lire le pouls du volume sanguin (BVP). La fréquence cardiaque, la variabilité de la fréquence cardiaque et d'autres caractéristiques cardiovasculaires peuvent être déterminées à partir de ce pouls du volume sanguin.

Capteur d'activité électrodermale (EDA) appliqué pour fournir l'excitation du système nerveux sympathique. Les caractéristiques liées au stress, à l'engagement et à l'excitation peuvent être déduites de cette EDA.

Un capteur accélérométrique à 3 axes est utilisé pour capturer l'activité basée sur le mouvement.

Un capteur thermopile infrarouge est utilisé pour mesurer la température de la peau.



FIGURE 3.14 – Présentation du bracelet E4 [Empatica](#) (2021).

L'envoi et le traitement de données en temps réel est possible grâce à une interface BLE. Dans un mode portable confortable et compact, le dispositif E4 permet de mener une étude de santé en dehors des conditions de laboratoire en acquérant des données en temps réel pour un enregistrement mobile. La qualité des données du bracelet E4 a été validé par cette étude [McCarthy et al. \(2016\)](#). Les détails techniques sont présentés dans le tableau 3.8.

Capteur	PPG	EDA	Température	Accéléromètre
fréquence d'échantillonnage	64 Hz	4 Hz	4 Hz	32 Hz
Unité de mesure	nanoWatt (nW)	microSiemens (μS)	Celsius	G-foces (g)
Portée	–	0,01 – 100 μS	–40 à 115 ° C	–2 à 2g
Résolution	0,9nW/Digit	1 Digit to 900 pico Siemens	0,02 ° C	8 bit de la plage sélectionnée
Description	Photodiodes	Électrodes	Thermopile infrarouge	Accéléromètre à 3 axes
Emplacement des capteurs	Poignet	Dos du poignet	Indifférent	Poignet

Tableau 3.8 – Caractéristiques techniques d'Empatica E4.

Les détails techniques des capteurs du bracelet E4 d'Empatica sont précisés dans le tableau. Le bracelet E4 est équipé d'un capteur PPG, qui éclaire la peau humaine et capte la lumière réfléchie, le signal en sortie est le signal BVP (Blood Volume Pulse). Le taux d'échantillonnage du capteur PPG est de 64 Hz. Dans son application, le signal PPG est l'entrée de l'algorithme qui mesure les battements de coeur et calcule ainsi les signaux IBI (Inter Beat Interval) qu'on peut obtenir en sortie.

Dans ce processus, au lieu de la combinaison traditionnelle rouge/IR, le bracelet E4 utilise le couplage des lumières verte et rouge pour augmenter l'efficacité de détection de l'onde de pouls. Le capteur est composé de 4 longueurs d'onde d'opération des LEDs (2 leds vertes, 2 leds rouges), et de 2 unités de photodiodes (14mm² de surface sensible). Les lumières créées par les LEDs vertes et rouges sont orientées vers la peau et absorbées par le sang de façon différente. Ensuite, le récepteur de lumière est utilisé pour mesurer la portion de la lumière réfléchie. L'intervalle de temps entre deux points diastoliques, qui sont des minima locaux du signal BVP, est appliqué pour calculer l'IBI.

MetaMotionLR

Le MetamotionRL est dispositif connecté portable et multi-capteurs. Il offre une surveillance continue avec des données de capteurs de mouvement et d'environnement en temps réel. Le dispositif MetamotionLR, présenté dans la figure 3.15 est équipé de 5 capteurs dont trois capteurs inertiels et deux capteurs environnementaux. Dans notre étude, nous allons nous intéresser uniquement aux capteurs inertiels qui sont :

Magnétomètre pour lire le champ magnétique, il nous indique le nord terrestre.

Il permet de donner des informations de position et d'orientation par rapport au champ magnétique terrestre sur les 3 axes X, Y, Z.

Accéléromètre utilisé pour capter le mouvement sur les 3 axes X, Y, Z. Il permet le calcul de de l'inclinaison d'un corps par rapport au sol et de mesurer le mouvement sur les 3 axes.

Gyroscope associé à l'accéléromètre, il nous indique l'orientation d'un corps. Il est complémentaire avec l'accéléromètre pour la mesure du mouvement.

Le Tableau 3.9 résume les spécifications techniques du dispositif MetamotionRL.

3. De la taxonomie à la mise en place du dispositif d'acquisition

Capteur	Accéléromètre	Magnétomètre	Gyroscope
Fréquence d'échantillonnage	0.001Hz – 100Hz	0.001Hz – 25Hz	0.001Hz – 100Hz
Unité de mesure	G-forces (g)	microTesla (μT)	Degré par seconde ($^{\circ}/s$)
Portée	$\pm 2, \pm 4, \pm 8, \pm 16$ g	$\pm 1300\mu T(x, y - axis), \pm 2500\mu T$ (z-axis)	$\pm 125, \pm 250, \pm 500,$ $\pm 1000, \pm 2000^{\circ}/s$
Résolution	16 bit	$0.3\mu T$	16 bit
Emplacement des capteurs	Plexus Solaire	Plexus Solaire	Plexus Solaire

Tableau 3.9 – Caractéristiques techniques du dispositif MetamotionR.

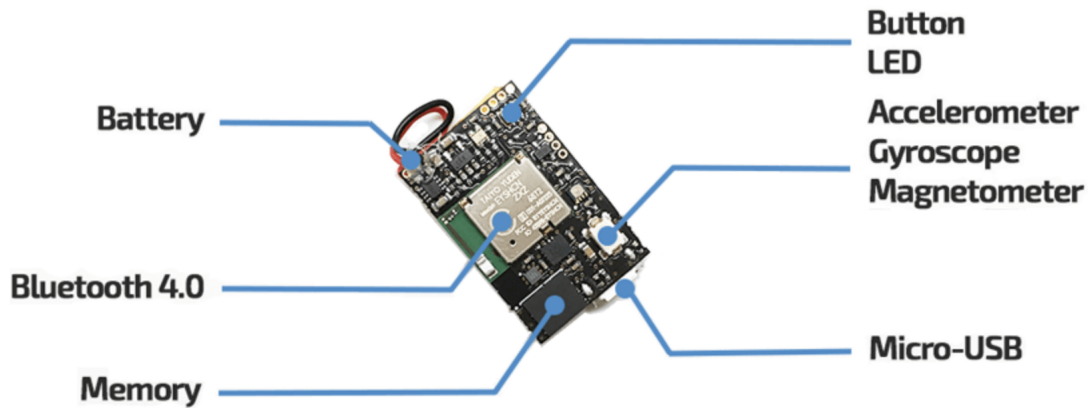


FIGURE 3.15 – Présentation du dispositif connecté MetamotionR.

Caméra USB

Pour la caméra, nous avons opté pour le même modèle que celui présenté pour le premier prototype. Ce choix a été motivé par le fait que les données récupérées sont directement envoyées sur le téléphone, et malgré le fil qui la connecte au téléphone, le dispositif reste facile à porter.

3.5.5 Acquisition des données

Les deux objets connectés utilisés (le MetamotionR et le EmpaticaE4) sont fournis avec des SDK pouvant être intégrés sur plusieurs interfaces. Pour la caméra, nous avons utilisé une solution logicielle permettant d'exploiter les données vidéo issues de caméra USB et de les encoder sous le format souhaité 3.16. Le tableau 3.10, présente succinctement tous les capteurs utilisés dans le dispositif BIRDS. Nous avons décidé de nous appuyer sur des smartphones avec des OS Android en raison de la



FIGURE 3.16 – Présentation de la solution BIRDS retenue.

diversité des équipements disponibles sur le marché, de la facilité de déploiement et de l'adoption généralisée du système. La fonction principale de la passerelle Android est de collecter les données en temps réel de les synchroniser et de les enregistrer avec une nomenclature définie pour fin de traitement.

Les données physiologiques et les données de mouvement qui sont collectées sont enregistrées dans des fichiers Comma-separated values (CSV) avec les timestamps synchronisés correspondant à chaque donnée. Le tableau 3.11 résume les dispositifs utilisés pour monter la solution BIRDS.

Dans ce qui suit nous allons présenter la synchronisation implémentée sur notre appareil de collecte de données.

Capteur	Mesure	Unité	Fr d'acquisition	Système d'acquisition	Emplacement	
Accéléromètre	Accélération	m/s^2	12.5 Hz	800 Hz	$\pm 2, \pm 8, \pm 16$	2048 \sim 16384 counts/g
Gyroscopie	Vitesse angulaire	$^\circ/s$	25 Hz	1600 Hz	$\pm 125, \pm 2000$	16 \sim 262 counts/ $^\circ$
Magnétomètre	Champs magnétique	T (tesla)	10 Hz	300 Hz	$\pm 1300\mu T$ (x, y-axis), $\pm 2500\mu T$ (z axis)	$\pm 0.3\mu T$

Tableau 3.10 – Tableau récapitulatif des capteurs utilisés par le dispositif BIRDS.

IoT	Capteurs	Fréquences
Empatica E4	Accéléromètre	32Hz
	EDA	4Hz
	PPG	64Hz
	Skin température	4Hz
MetamotionR	Accéléromètre	100Hz
	Magnétomètre	25Hz
	Gyroscope	100Hz
Caméra	Caméra vidéo	25Hz

Tableau 3.11 – Caractéristiques de la dispositif IoT MetaMotionR avec les capteurs retenus.

3.6 Synchronisations des données

L'étape de pré-traitement des données est une étape cruciale dans la détection des situations à risques, en particulier pour notre système pour lequel, la fiabilité des données d'entrée est extrêmement importante.

Cependant, la qualité des capteurs généralement utilisés pour les objets connectés reste faible et les moyens de transmission sans fil utilisés sont soumis au décalage d'horloge et aux interférences entre capteurs. De ce fait, les données peuvent contenir du bruit, des artefacts dûs aux mouvements du sujet, des valeurs manquantes et une désynchronisation [Shih et al. \(2015\)](#). Cela peut alors mener à de fausses détections et donc à de fausses alarmes. C'est pourquoi un pré-traitement des données brutes est nécessaire.

Le pré-traitement des données de notre système repose sur deux étapes :

- Synchronisation des données essentielles avant la fusion des données.
- Rééquilibrage des classes d'apprentissage pour la classification ;

Dans ce qui suit, nous expliquerons plus en détails les pré-traitements envisagés pour notre système, pour finir par conclure et exposer nos perspectives pour le travail qui reste à mener.

3.6.1 Choix de l'approche par fusion de données

La fusion de données multi-capteurs est un domaine de recherche bien établi, il existe une vaste littérature traitant du sujet [GUSTAVO HERNÁNDEZ-PEÑALOZA et ÁLVAREZ \(2017\)](#), [Hernández-Peñaloza et al. \(2018\)](#), [Luo et al. \(2017\)](#), [Banaee et al. \(2013\)](#)... La fusion de données sera essentiellement utilisée pour la détection de risques de chutes pour notre projet. Tout système de mesure basé sur des objets connectés s'appuyant sur un seul capteur, ou sur plusieurs capteurs considérés individuellement, souffre de plusieurs limitations, telles que :

- Défaillance d'un des capteurs qui peut entraîner une perte de la mesure ;
- Présence de bruit dû aux moyens de transmissions ;
- Non-déterminisme des sujets biologiques.

L'utilisation de données provenant de sources multiples assure une plus grande fiabilité du dispositif, une robustesse accrue face aux interférences environnementales et une amélioration de la précision des mesures.

3.6.2 Synchronisation des données

La fusion des mesures des différents capteurs est prévue pour l'analyse des situations à risques. Un système de fusion de données capteurs doit faire face à des fréquences de mesure de capteurs différentes et variables, à la latence de mesure des capteurs ainsi qu'à des temps de mesure asynchrones.

La synchronisation des données hybrides est une étape fondamentale avant toute fusion de données. Le défi est d'autant plus important pour nous car la fusion est prévue en temps réel. Ceci nécessite une synchronisation d'horloge évolutive puisqu'elle doit fournir une référence temporelle commune pour toutes les données reçues des capteurs physiologiques et des capteurs inertiels.

Les capteurs étant prévus pour un fonctionnement à longues périodes, pour conserver l'énergie limitée des batteries, la majorité des objets connectés sont programmés pour se mettre en veille en cas d'inactivité prolongée. Un maintien d'une référence temporelle commune est important pour le moment de remise en marche des capteurs.

Les deux exigences les plus importantes (et conflictuelles) pour un tel protocole de synchronisation sont une grande précision de synchronisation et une faible consommation d'énergie.

Pour décrire le problème de synchronisation, nous allons passer par les points suivants :

Délai de transmission aléatoire : D'après [PalChaudhuri et al. \(2004\)](#), celui-ci est la source d'erreur la plus importante à considérer dans les systèmes de synchronisation temporelle. Les types d'erreurs peuvent alors être considérés comme suit :

- Le délai nécessaire pour construire le message et le transférer à la couche réseau pour la transmission ;
- Temps d'accès [Maggs et al. \(2012\)](#), c'est le temps passé dans la couche Media Access Control (MAC) en attente de l'accès au canal de transmission. Le temps d'accès dépend essentiellement du schéma MAC utilisé ;
- Délai de propagation [Maggs et al. \(2012\)](#), c'est le délai entre l'émission et la réception du message contenant la mesure du capteur. Ce délai inclut le temps de retransmission dû à des erreurs ;
- Délai de traitement à la réception : Le temps passé à l'interface du récepteur pour traiter le message entrant ;
- La qualité de l'horloge : Les horloges utilisées sont des horloges quartz qui manquent de précision et souvent elles sont source de décalages dans le temps aberrants.

Le retard d'accès est un retard non-déterministe affectant la synchronisation globale du système [Maróti et al. \(2004\)](#). Lorsque le timestamp disponible est celui de la couche MAC (comme c'est le cas sur nos données extraites du Microsoft Band 2), la précision de synchronisation est principalement altérée par les retards d'interruption et de quantification. Ces retards peuvent conduire à des désynchronisations du système, et donc fausser nos résultats d'analyse. Une autre source d'inexactitude est la fluctuation de la fréquence d'horloge. Sachant que la dérive relative n'est pas constante, en suivant le schéma utilisé dans la plupart des estimations suivies pour estimer la dérive d'horloge basée sur une erreur déterministe, nous introduisons des erreurs de synchronisation supplémentaires.

Interruption de connexion BLE : Comme tout système avec une technologie de transmission radio (BLE dans notre cas), nous sommes régulièrement confrontés à des interruptions de connexion entre un ou plusieurs des objets connectés et le smartphone sur lequel est fait le traitement des données mesurées. Ceci fait qu'au moment de la reconnexion des appareils, nous avons un ou plusieurs signaux désynchronisés.

Contraintes matérielles : Notre système utilise une batterie avec une charge limitée et parfois des capteurs avec une qualité se détériorant au fil du temps. Ces limitations matérielles se traduisent par des désynchronisations de notre système.

3.6.3 Modélisation mathématique de la désynchronisation

Dans cette section, nous allons essayer de modéliser les erreurs de timestamp des données multi-capteurs ainsi que la désynchronisation des capteurs entre eux. L'équation modélisant le timestamp enregistré sur le smartphone en fonction de le timestamp de référence, du déphasage entre capteurs, des erreurs dues au délais transmission aléatoires et de la différence de fréquence d'échantillonnage entre les deux signaux est représentée ci-après :

$$C_i(t) = F_i x + \theta_i + \phi_i \quad (3.1)$$

Avec F_i le décalage en fréquence pour chaque capteur i , x le timestamp vrai, $C_i(t)$ le timestamp observé en sortie du smartphone, ϕ_i le déphasage dû à l'enclenchement aléatoire des différents capteurs et θ_i est la désynchronisation des différents capteurs due au retard de transmission. Dans le cas d'un timestamp idéal $C_i(t) = x$ avec $\phi_i = 0$, $\theta_i = 0$ et $F_i = 1$.

Le modèle décrit dans l'équation (3.1) prend en compte le caractère non linéaire de la désynchronisation des capteurs. Avec le protocole BLE, les délais de transmission sont beaucoup plus courts qu'avant, les données sont moins sensibles aux

interférences [Gomez et al. \(2012\)](#). De ce fait, nous avons supposé $\theta_i = 0$, le modèle représenté dans l'équation (3.1), devient :

$$C_i(t) = F_i x + \phi_i \quad (3.2)$$

Le modèle représenté dans (3.2) nous servira dans nos premiers algorithmes de synchronisation (voir section suivante).

3.6.4 Algorithmes de synchronisation dans la littérature

Plusieurs solutions existent dans la littérature pour la synchronisation multi-capteurs. Il est notamment possible de distinguer les méthodes de correction des timestamps en estimant les retards d'envoi et les méthodes d'estimation au moment de la réception des déphasages et des différences de fréquences entre capteurs.

Synchronisation par échange des timestamps

Trois approches pour les modèles de correction des délais sur les timestamps pour les réseaux sans fil sont identifiées : la synchronisation émetteur-récepteur [Ganerwal et al. \(2003\)](#), la synchronisation récepteur-récepteur [Elson et al. \(2002\)](#) et la diffusion de messages à sens unique [Maróti et al. \(2004\)](#) (récepteur seulement). Ces méthodes présentent plusieurs avantages notamment, un coût calculatoire très bas, élimination des retards d'envoi et estimation des déphasages entre capteurs.

Toutefois, ces méthodes ne sont pas applicables sur notre prototype car nos dispositifs sont fermés et nous ne pouvons pas intervenir sur leurs couches MAC. Pour cela, nous avons ignoré dans un premier temps les retards non-déterministes dus aux retards d'envoi, de réception et d'accès à la mesure.

Synchronisation des timestamps par estimation de la fréquence et du déphasage

Ci-après, nous citerons quelques méthodes connues dans la littérature pour l'estimation des décalages de fréquences entre capteurs et du déphasage temporel entre capteurs :

La méthode proposée par [Jeske \(2005\)](#) estime seulement le déphasage entre capteurs, ceci rend l'approche peu pratique pour une bonne synchronisation des capteurs.

[Leng et Wu \(2010\)](#) propose un estimateur à maximum de vraisemblance qui estime le délai de communication séparément, et l'utilise pour estimer le déphasage et le décalage en fréquence entre signaux.

Leng et Wu (2010) propose également un estimateur des moindres carrés à faible complexité qui élimine le déphasage et le décalage en fréquence entre signaux ainsi que les retards de la synchronisation.

Rajan et van der Veen (2011) propose une approche par la méthode des moindres carrés par paires pour estimer les paramètres de modélisant chaque capteur. Il propose également une approche par moindres carrés généralisés. La faible complexité de l'estimateur des moindres carrés par paires semble intéressante mais il est impossible de l'appliquer à nos capteurs à cause de la nature de nos dispositifs qui sont fermés.

3.6.5 Solution proposée

La synchronisation que nous proposons comporte deux étapes :

- Élimination du décalage de phase de démarrage. Différents capteurs ont leurs propres décalages horaires à partir du moment où ils sont allumés et du moment où ils commencent à transmettre des données. Nous utilisons une mémoire tampon qui ne commence à capturer des données que lorsque tous les capteurs ont commencé à transmettre des données. Pour l'ensemble des capteurs de notre appareil, la mémoire tampon devrait être d'environ $0,6 * 10^5$.
- Synchronisation des timestamps des différents capteurs. Pour ce faire, nous proposons de corriger, en temps réel, les valeurs de timestamps reçues par le smartphone en les comparant avec celles générées par le capteur.

Considérant deux signaux un timestamp réel $c_i(t)$ correspondant à l'heure à laquelle les données de chaque capteur i -th sont capturées et un signal horodaté par le smartphone $x(t)$. Nous recherchons le meilleur filtre à réponse impulsionnelle finie pour trouver le signal désiré. Pour ce faire, l'erreur entre les deux est minimisée, dans le sens de l'erreur quadratique moyenne. Son expression est donnée ci-dessous :

$$J(h) = E((c_i(t) - h_i(\tau)x(t))^2) \quad (3.3)$$

avec $c_i(t) = h(\tau)x(t)$, h - réponse du filtre et $x(t)$ représente les données de timestamps.

La méthode de minimisation que nous appliquons est la descente en gradient. L'algorithme du gradient a l'équation :

$$h_{n+1} = h_n - \mu \nabla J(h_n) \quad (3.4)$$

avec n - numéro d'itération. Nous utilisons la version stochastique de la descente du gradient en changeant avec les itérations le paquet de données dans le temps Haykin et al. (2003).

Résultats de la synchronisation

Les résultats observés varient en fonction de plusieurs paramètres : le pas de descente de gradient et le nombre d'échantillons sur lesquels le filtrage est fait ainsi que la taille du filtre. Les figures suivantes représentent la racine de l'erreur quadratique entre les timestamps d'envoi et les timestamps de réception des données avant et après filtrages et l'allure du timestamp de réception.

3.6.6 Influence de la taille du filtre

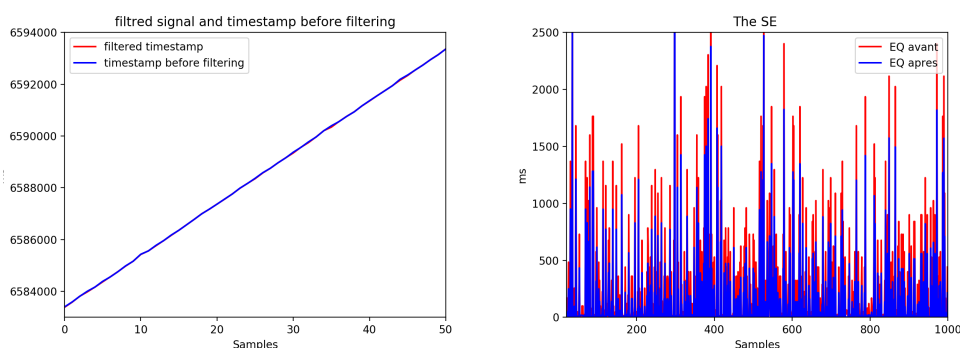


FIGURE 3.17 – À gauche la figure représente le signal timestamp de référence et le signal timestamp filtré. À droite l'erreur quadratique entre le signal timestamp de référence et le timestamp output avant et après filtrage. Taille du filtre = 1, nombre d'échantillons=1000, pas de descente = 0.0001.

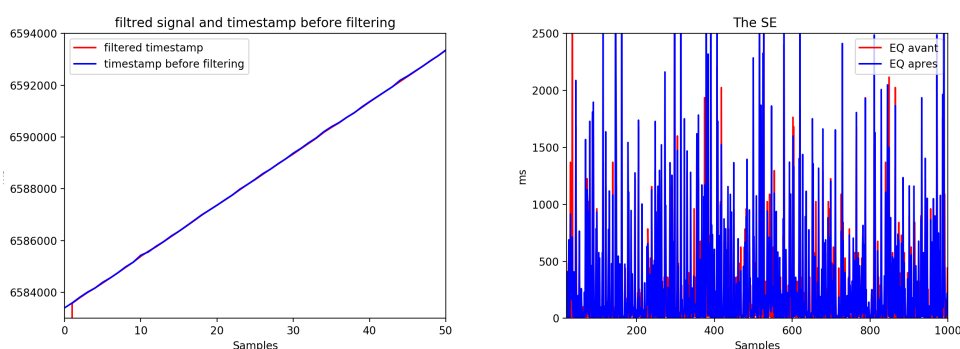


FIGURE 3.18 – À gauche la figure représente le signal timestamp de référence et le signal timestamp filtré. À droite l'erreur quadratique entre le signal timestamp de référence et le timestamp output avant et après filtrage. Taille du filtre = 2, nombre d'échantillons=1000, pas de descente = 0.0001.

Nous pouvons voir que le signal de référence et le signal filtré sont très proches. Néanmoins, nous pouvons constater à travers la courbe de l'erreur quadratique que

l'erreur entre les deux signaux augmente en augmentant la taille du filtre. En effet, l'erreur quadratique moyenne entre les deux signaux passe de 236 à 142 pour un filtre de taille $t = 1$ contre 310 pour un filtre $t = 2$. À partir de $t = 5$, le filtre diverge.

Influence du pas de descente de gradient

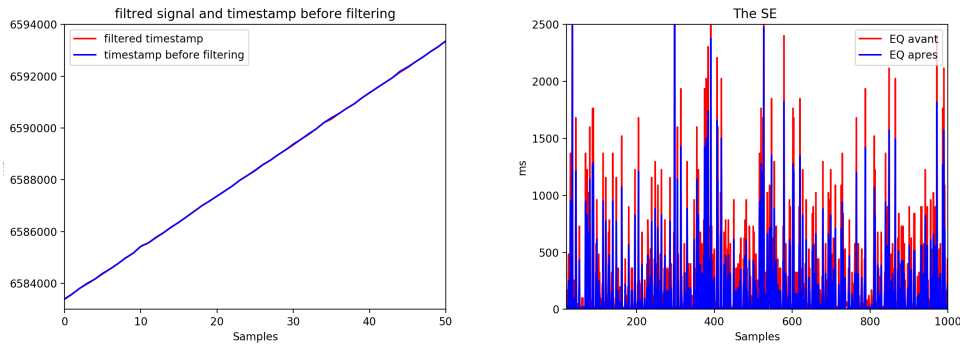


FIGURE 3.19 – À gauche la figure représente le signal timestamp de référence et le signal timestamp filtré. À droite l'erreur quadratique entre le signal timestamp de référence et le timestamp output avant et après filtrage. Taille du filtre = 1, nombre d'échantillons= 1000, pas de descente = 0.0001.

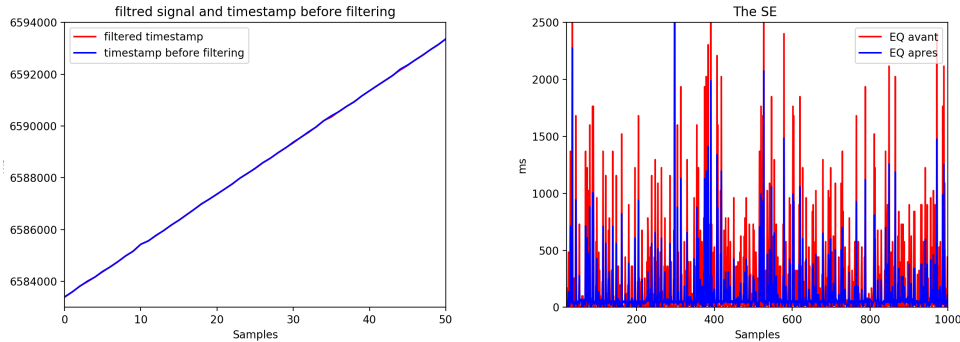


FIGURE 3.20 – À gauche la figure représente le signal timestamp de référence et le signal timestamp filtré. À droite l'erreur quadratique entre le signal timestamp de référence et le timestamp output avant et après filtrage. Taille du filtre = 1, nombre d'échantillons= 1000, pas de descente = 0.00001.

Nous pouvons voir que le signal de référence et le signal filtré sont très proches. Nous constatons à travers la courbe de l'erreur quadratique que l'erreur entre les deux signaux que plus le pas de descente est petit plus l'erreur diminue. L'erreur quadratique moyenne entre les deux signaux passe de 236 à 142 pour $\mu = 10^{-4}$ et de 236 à 124 pour $\mu = 10^{-5}$. De plus, lorsque nous augmentons la taille du filtre à

3. De la taxonomie à la mise en place du dispositif d'acquisition

des valeurs supérieures à 10^{-3} , le filtre diverge. Pour un bon compromis rapidité et bonne convergence, nous avons choisi de prendre un pas à 10^{-4} .

Influence de la taille de l'échantillon de données

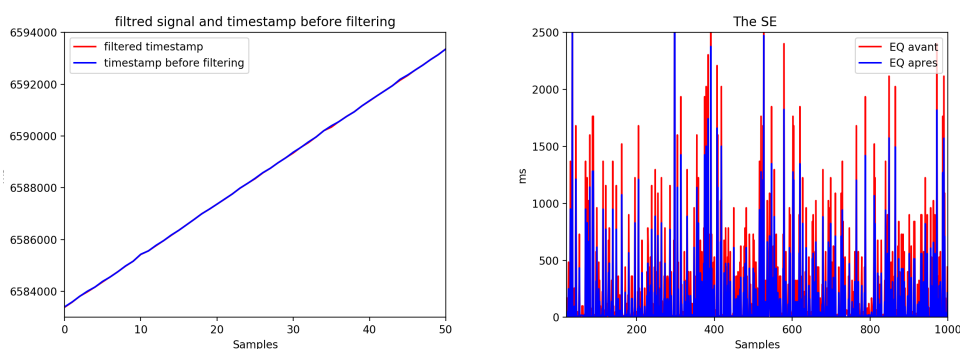


FIGURE 3.21 – À gauche la figure représente le signal timestamp de référence et le signal timestamp filtré. À droite l'erreur quadratique entre le signal timestamp de référence et le timestamp output avant et après filtrage. Taille du filtre = 1, nombre d'échantillons= 1000, pas de descente = 0.0001.

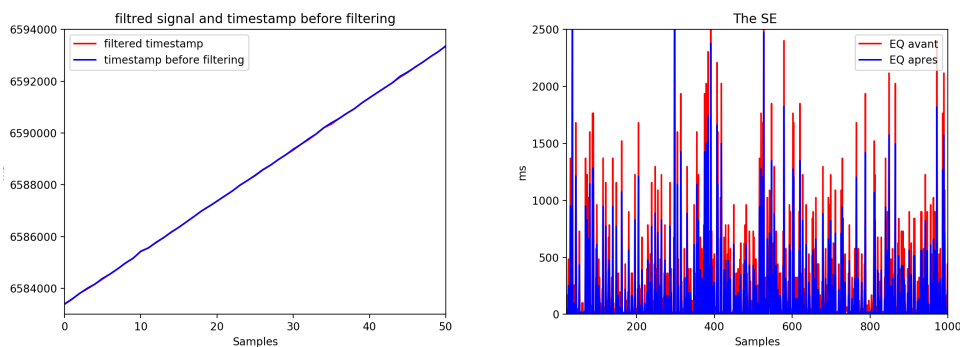


FIGURE 3.22 – À gauche la figure représente le signal timestamp de référence et le signal timestamp filtré. À droite l'erreur quadratique entre le signal timestamp de référence et le timestamp output avant et après filtrage. Taille du filtre = 1, nombre d'échantillons= 2000, pas de descente = 0.0001.

	Dataset 1		Dataset 2		Dataset 3		Dataset 4	
	MSE Before filtering	MSE After filtering	MSE Before filtering	MSE After filtering	MSE Before filtering	MSE After filtering	MSE Before filtering	MSE After filtering
GSR tmstp	236.57	142.92	396.26	271.81	2011.02	1370.43	280.51	128.54
Accelerat-ion tmstp	147.76	69.61	195.43	156.76	582.24	378.23	100.45	40.81
velocity tmstp	1251.96	345.76	877.96	287.61	3622.23	1200.23	1027.85	238.04
Skin temperatur e tmstp	1058.33	443.75	452.75	136.26	3225.05	784.09	666.73	343.98

Tableau 3.12 – Résultats obtenues par l’algorithme de synchronisation LMS en temps réel au sens de la MSE (mean square error) entre le timestamp du dispositif IoT et le timestamps lié à l’horloge interne du contrôleur principal au moment de la réception des données.

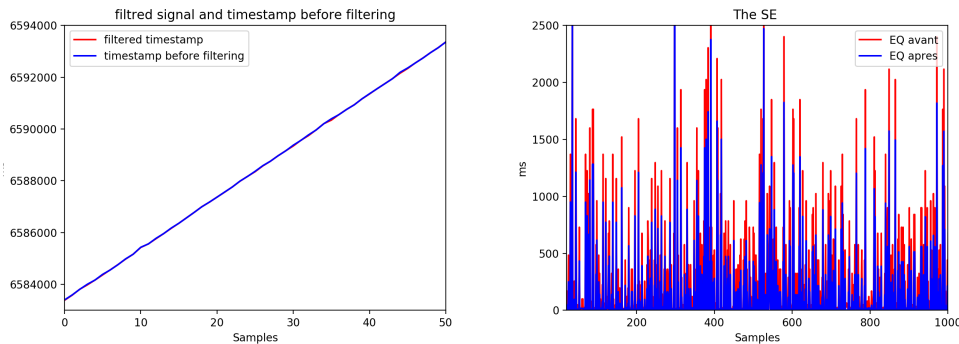


FIGURE 3.23 – À gauche la figure représente le signal timestamp de référence et le signal timestamp filtré. À droite l’erreur quadratique entre le signal timestamp de référence et le timestamp output avant et après filtrage. Taille du filtre = 1, nombre d’échantillons = 5000, pas de descente = 0.0001.

Pour la taille de l’échantillon de données, nous pouvons voir à travers les résultats obtenus que malgré des petites variances dans l’erreur quadratique moyenne dans les résultats des échantillons pris, le filtre marche aussi bien sur les trois tailles d’échantillons.

Le Tableau 3.12 présente les résultats de synchronisation des timestamps de différentes capteurs issus d’objets IoT distincts donc avec des horloges internes distinctes et non synchronisées entre elles. D’après nos expériences sur un ensemble de 1000

échantillons de couples d'horodatages provenant du smartphone concentrateur et du capteur EDA, le filtre converge dans le cas d'une configuration simple : sa longueur est de 1. Sur l'ensemble des données enregistrées des horodatages, l'erreur quadratique moyenne entre deux signaux $h(\tau) * x(t)$ et $y(t)$ diminue de 236 à 142 après le filtrage et elle correspond à une erreur moyenne de 11 ms par rapport à 15 ms avant le processus de filtrage. Ceci est compatible avec les mesures physiologiques - le temps d'apparition de la chute selon la littérature [J.Klenk et al. \(2011\)](#) est compris entre 150 et 600 ms. Le tableau 3.12, donne les résultats obtenus par l'algorithme de synchronisation LMS en temps réel au sens de la MSE (mean square error) entre le timestamp du dispositif IoT et le timestamp lié à l'horloge interne du contrôleur principal (le smartphone) au moment de la réception des données.

3.7 Conclusion

Ce chapitre rend compte de la conception et du développement d'un système multimodal de surveillance des personnes fragiles. Sur la base d'études psychologiques, nous avons élaboré une taxonomie des situations à risques et un scénario d'enregistrement. L'architecture du système a été décrite ainsi que ses principaux modules. Le concept clé présenté dans ce travail est l'approche multimodale qui nécessite une synchronisation appropriée des données des capteurs. Un modèle de synchronisation générique a donc été proposé, pouvant évoluer avec d'autres dispositifs.

Chapitre 4

Analyses de situations à risques sur des données Lifelog

4.1 Introduction

De nombreuses études ont été réalisées pour déterminer les situations à risques les plus fréquentes auxquelles sont confrontées les personnes fragiles [Vemulapalli et al. \(2017\)](#). Dans ce chapitre, nous cherchons à détecter des situations à risques citées dans notre taxonomie [3](#) et qui sont plus complexes. Nous pouvons qualifier ces situations à risques de situations "sémantiques" ; elles représentent des situations souvent rencontrées dans la vie quotidienne des personnes fragiles.

Nous présentons nos contributions sur la détection des risques dans les données du journal de vie pour la prévention des risques chez les personnes fragiles.

Nous proposons d'utiliser le dataset NTCIR13 [Gurrin et al. \(2019b\)](#) qui présente des données de journal de vie à usage général afin d'identifier des situations à risques potentielles. Nous appelons ces risques "sémantiques" en accentuant le contenu complexe que nous souhaitons détecter dans le flux de données. Des exemples en sont un risque d'accident domestique et un risque de fraude dans notre cas avec les données visuelles. Dans ce chapitre, nous présentons notre démarche d'identification des risques dans des données multimodales "Lifelog", en utilisant leur composante visuelle. Nous avons aussi fait un premier test avec un réseau LSTM sur les données de signaux 1D pour la détection des différentes situations à risques mentionnées [3](#).

Dans ce qui suit, nous présentons le dataset, notre méthodologie et nos contributions.

4.2 Les données Lifelog

Il n'existe pas de définition convenue de la journalisation de la vie. De nombreuses activités sont désignées sous le terme de data "Lifelog", chacune produisant une certaine forme d'archive de données de journalisation de la vie. Parmi les formats les plus populaires, nous pouvons citer le format journal de bord, les mémoires numériques personnelles (ou humaines), la boîte noire humaine, etc.

Pour choisir une définition appropriée, nous pouvons citer la définition de la journalisation de la vie quotidienne par [Dodge et Kitchin \(2007\)](#), où la journalisation de la vie est décrite comme "une forme d'informatique omniprésente, consistant en un enregistrement numérique unifié de la totalité des expériences d'un individu, capturées de manière multimodale par des capteurs numériques et stockées de manière permanente sous forme d'archives personnelles multimédias".

L'enregistrement numérique unifié utilise des données capturées de manière multimodale qui ont été rassemblées, stockées et traitées en informations sémantiquement significatives et récupérables et ont été rendues accessibles par une interface, qui peut potentiellement supporter une grande variété de cas d'utilisation.

Un aspect essentiel de cette définition est que les données Lifelog doivent s'efforcer d'enregistrer la totalité des expériences d'un individu. Dans les données présentées par le dataset NTCIR13 [Gurrin et al. \(2019b\)](#), la définition citée précédemment est adoptée. Il est supposé que la journalisation de la vie tente de capturer une trace détaillée des actions d'un individu. Trois principes sont à la base du lifelogging et sont respectés dans l'établissement de ce dataset :

La journalisation La journalisation de la vie est le processus de collecte, de traitement et de réflexion passifs sur les données de l'expérience de vie recueillies par une variété de capteurs, et elle est effectuée par un individu, le lifelogger. Les données de l'expérience de vie sont principalement basées sur des capteurs portables qui détectent les activités de la personne, bien que parfois des données provenant de capteurs environnementaux ou d'autres capteurs d'information puissent être intégrées au processus [Oh et al. \(2015\)](#) ;

Le journal de bord Le journal de bord est le recueil des données réelles. Il peut se trouver sur un disque dur personnel, dans le cloud ou dans un dispositif de stockage portable. Le journal de vie peut être aussi simple qu'une collection de photos, ou devenir aussi vaste et complexe qu'une vie entière de données sensorielles portables (par exemple, des journaux de localisation GPS ou des traces d'activité d'accéléromètre) ;

Une mémoire de substitution s'apparente à une bibliothèque numérique, c'est-à-dire les données du journal de bord et le logiciel associé pour organiser et gérer les données du journal de bord. C'est le principal défi à relever pour la recherche d'informations, à savoir développer une nouvelle génération de technologies de recherche capables de fonctionner sur des archives de données aussi énormes. Le terme "mémoire de substitution" n'implique aucune forme de processus cognitif, il s'agit simplement de la bibliothèque numérique des données du journal de vie, qui jusqu'à présent se concentrait sur la conservation d'une liste d'événements ou d'épisodes.

4.3 Présentation du dataset NTCIR 13

L'enregistrement et l'analyse de données Lifelog à des fins de remémoration des événements a été proposé, par exemple une recherche avec un historique important ayant déjà trouvé ses applications dans le domaine de la santé pour la remémoration des événements chez des patients atteints de démence par [Caprani et al. \(2014\)](#). L'ensemble de données NTCIR 13 [Gurrin et al. \(2014\)](#) a été mis à notre disposition par l'équipe de l'Université DCU à des fins de recherche. Une collaboration avec Cathal Gurrin et ses doctorants a été menée pour l'analyse de ce corpus.

L'ensemble de données LSC Lifelog [Gurrin et al. \(2014\)](#) comprend un large éventail de données multimodales sur l'expérience de vie de deux individus recueillies par des capteurs portables.

Les données ont été recueillies par deux lifeloggers qui portaient les dispositifs de suivi de la vie et recueillaient des données physiologiques de base pendant une bonne part des heures d'éveil de la journée. Un lifelogger a recueilli un mois de données et un autre a recueilli deux mois de données, soit un total de 90 jours de données pour le test.

Les lifeloggers portaient un appareil photo portable Narrative Clip accroché à un vêtement ou porté par une lanière autour du cou (voir figure 4.1, qui capturait les activités quotidiennes du porteur (du point de vue du porteur) figure 4.2 et fonctionnait pendant 10 à 12 heures avec une batterie chargée, recueillant environ 1 250 à 1 500 images par jour. Cet appareil prend des photos de manière passive (c'est-à-dire sans intervention explicite de l'utilisateur) à raison d'environ deux images par minute. En outre, les enregistreurs de vie comprennent des données provenant de l'application Moves (lieux et mouvements physiques) et un enregistrement de la musique écoutée, capturé à partir des comptes Last.fm. L'application Moves est une application pour smartphone qui enregistre automatiquement l'activité de l'utilisateur en termes de lieux sémantiques et d'activités physiques (veille, vélo, course, transport, etc.).

L'ensemble de données comprend des données sur la santé et le bien-être provenant de moniteurs de fréquence cardiaque continus, de moniteurs de glycémie quotidiens avec des lectures hebdomadaires supplémentaires du cholestérol et de l'acide urique, ainsi que des annotations manuelles de la consommation de nourriture et de boissons. Après le processus de collecte des données, un certain nombre de mesures ont été prises pour s'assurer que la collecte des tests était aussi réaliste que possible et tenait compte des sensibilités associées aux données personnelles :

Synchronisation temporelle Toutes les données ont été alignées temporellement sur l'heure UTC.

Filtrage des données Compte tenu de la nature personnelle des données du journal de bord, il est proposé aux lifeloggers de supprimer toutes les données du journal de bord qu'ils ne souhaitent pas partager. Il est aussi important de noter qu'une vérification a été effectuée sur l'ensemble du dataset afin de



FIGURE 4.1 – Exemple du modèle de caméra porté pour l'enregistrement du dataset LSC NTCIR 13 [Gurrin et al. \(2014\)](#).



FIGURE 4.2 – Exemple d'images enregistrées par le lifelogueur dans le cadre du corpus LSC NTCIR 13 [Gurrin et al. \(2014\)](#).

s'assurer qu'aucune donnée personnelle n'est apparente.

Protection de la vie privée Un floutage des visages apparents a été effectué manuellement sur l'ensemble du dataset.

Les données consistent donc en une grande collection de données multimodales de journaux de vie sur 90 jours par deux lifeloggers (cf figure 4.3). Dans la plupart des cas, les activités des lifeloggers étaient distinctes et ils ne se sont pas rencontrés.

Les données capturées pour ce dataset sont décrites ci-dessous :

Données physiologiques À l'aide de la smartwatch Basis, les chercheurs ont recueilli 24 heures sur 24 et 7 jours sur 7 la fréquence cardiaque, la réponse galvanique de la peau, les calories brûlées et les pas. En outre, la pression artérielle et la glycémie ont été enregistrées chaque matin avant le petit déjeuner et les taux de cholestérol et d'acide urique ont été relevés chaque semaine.

Données sur l'activité humaine Les activités quotidiennes des personnes ont été saisies en termes d'emplacements sémantiques visités, d'activités physiques (par exemple, marcher, courir, se tenir debout) ainsi qu'un journal de régime horodaté de tous les aliments consommés et des boissons prises.

Données contextuelles comme mentionné plus haut dans cette description, l'enregistrement de données contextuelles grâce à des images égocentrées prises par un appareil photo portable Narrative Clip.

Amélioration des données Les images de la caméra portable ont été annotées avec les sorties d'un détecteur de concepts sémantiques de Microsoft accessible via leur API de vision par ordinateur. En outre, afin d'aider les participants dans le développement de leur logiciel, un moteur de recherche de base a été développé pour prendre en charge les requêtes de base du système. Des requêtes ont été soumises pour générer des listes classées basées sur des requêtes à facettes, par ID utilisateur, par concept visuel de lieu et/ou par activité physique.

Nous avons utilisé une partie de ces données qui comporte 27 jours d'enregistrement d'un seul liflogueur. Le volume global de la base de données utilisée est constitué de 41664 images accompagnées des données physiologiques, des données informant sur l'activité humaine et des données contextuelles enregistrées. Ces données sont accompagnées de la sortie d'un détecteur de concepts qui identifie les concepts visuels dominants dans chaque image à partir d'un total de 548 concepts visuels tels que *ciel*, *intérieur*, *cuisine*, *ordinateur*, *personne*, *etc.* Ces données visuelles sont accompagnées aussi d'un historique d'écoute musicale, bien que nous n'ayons pas utilisé ces données musicales dans notre recherche. L'architecture en termes d'organisation de données du corpus NTCIR 13 est illustrée sur la figure 4.3. Les images donnant les différents épisodes du Lifelog sont placées le long de l'axe temporel d'une journée d'enregistrement.

4.4. Mise en correspondance des concepts avec la taxonomie de risques et réannotation des données NTCIR 13

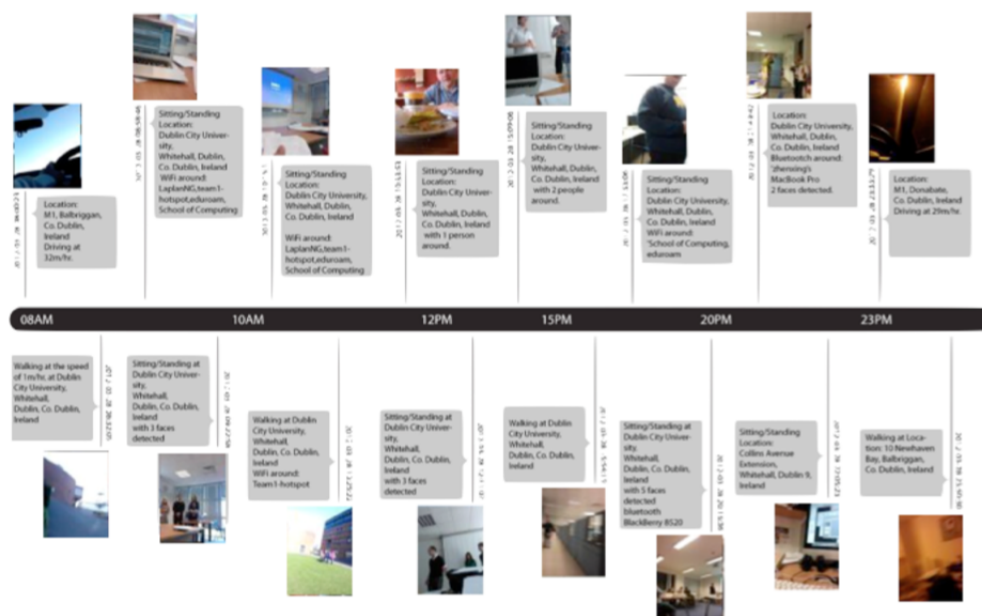


FIGURE 4.3 – Présentation de l'architecture du dataset NTCIR 13 Gurrin et al. (2014).

4.4 Mise en correspondance des concepts avec la taxonomie de risques et réannotation des données NTCIR 13

Les concepts du corpus NTCIR 13 représentent une classification fine des événements de la vie quotidienne. Par ailleurs, notre taxonomie de risques élaborée comporte des concepts plus complexes. Il est donc nécessaire de mettre en correspondance les concepts de NTCIR 13 détectés dans ce corpus automatiquement et les concepts de la taxonomie de risques. Ce travail a été réalisé en collaboration avec la doctorante en psychologie Marion Pech qui travaille sur le déploiement du dispositif auprès de sujets âgés. Dans un premier temps, nous avons utilisé les données d'images du corpus NTCIR 13 pour identifier les risques potentiels visuellement et avons élaboré la correspondance entre un ou plusieurs concepts NTCIR 13 annotés automatiquement avec une plateforme de reconnaissance de concepts Microsoft prévue à cet effet et chaque concept de notre taxonomie de risques. Le schéma sur la figure 4.4 illustre la mise en correspondance des concepts des deux taxonomies : celle du NTCIR 13 et la nôtre, présentée dans le chapitre 3.

Néanmoins, bien qu'il existe une intersection importante entre les concepts et les ensembles des concepts du NTCIR 13 et les risques sémantiques, il a été nécessaire de compléter la taxonomie des concepts NTCIR 13 pour répondre à notre taxonomie des risques. Par exemple, dans la situation "risque de chute", le concept "escalier"

4. Analyses de situations à risques sur des données Lifelog

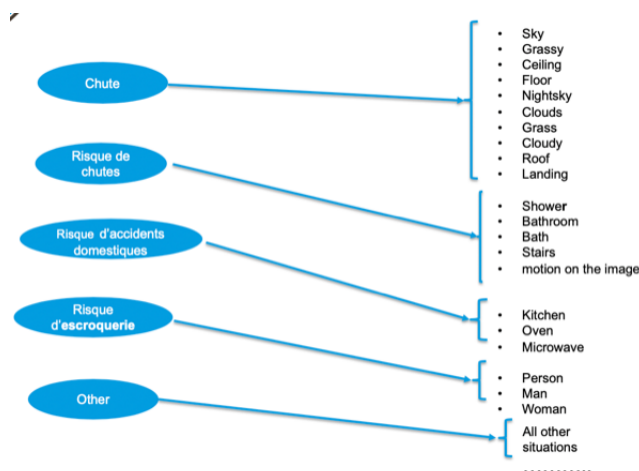


FIGURE 4.4 – Mise-en-correspondance des concepts du NTCIR 13 et des risques sémantiques.

Risk situations	Concepts
Fall	sky, grassy, ceiling, floor, night sky, clouds, roof, landing.
Risk of fall	shower, bathroom, bath, <i>staircase</i> .
Risk of fraud	person + door, woman + door, man + door.
Risk of domestic accident	kitchen, oven, microwave.

Tableau 4.1 – Concepts pour les situations à risques à partir du dataset Lifelog.

a été ajouté.

Le tableau 4.1) présente la correspondance entre les concepts NTCIR 13 complétés, représentés en italique dans la colonne de droite du tableau 4.1 et les risques sémantiques.

Avec la mise en correspondance des concepts, nous avons effectué une ré-annotation de la partie du corpus NTCIR 13 que nous avons utilisé suivant notre taxonomie des risques.

La première étape de l'annotation a été réalisée *automatiquement*. En analysant les métadonnées du corpus LSC NTCIR 13, c'est-à-dire le fichier des concepts liés aux images, les images ont été étiquetées avec la situation de risque correspondante. Par exemple, les images comportant le concept "cuisine" ont été affectées à la classe "risque d'accident domestique". Les images sans les concepts LSC NTCIR 13 de la table de correspondance ont été assignées à la classe de rejet "sans risque".

Les étiquettes des concepts LSC NTCIR 13 ont été générées dans ce corpus automatiquement avec une probabilité de présence dans chaque image. Les erreurs dans l'annotation NCIR 13 ont été donc possibles. Par ailleurs, il a été nécessaire d'in-

roduire les concepts manquants ("escalier" par exemple). Ainsi lors de la deuxième étape, toutes les annotations ont été vérifiées *manuellement*. Les images ont été ré-affectées à la classe de risque correspondante (colonne de gauche du tableau 4.1). Pour effectuer ce travail, une plateforme d'annotation a été développée à cet effet.

Les résultats d'annotations obtenus sur les données visuelles sont décrits dans le tableau 4.2. La distribution de ces annotations sur le reste des données du dataset LSC NTCIR 13 est exposée dans le tableau 4.4.

4.5 Reconnaissance des risques sémantiques sur les données du corpus NTCIR 13

Dans un premier temps, nous utilisons uniquement des données visuelles pour la reconnaissance des situations à risque. En considérant la correspondance entre les concepts de risque (colonne de gauche du tableau 4.1) et les concepts LSC NTCIR 13 (colonne de droite du tableau 4.1), nous pouvons remarquer que des situations à risques telles que "chute" ou "risque de chute" ont un ensemble hétérogène de concepts visuels dans le dataset LSC NTCIR 13. Il est illusoire d'essayer de les reconnaître uniquement sur des images. Les données physiologiques exprimant l'homéostasie de la personne doivent être prises en compte, comme la réponse cutanée galvanique, le rythme cardiaque... Néanmoins, pour le corpus LSC NTCIR 13, cela est discutable car il a été enregistré par une personne en bonne santé et très active. Un corpus réel sur des sujets fragiles est nécessaire. Par ailleurs, la fréquence d'échantillonnage des données capteurs toutes les minutes se prête difficilement à la détection des situations à risques qui ne durent que quelques secondes. Dans la suite de ce chapitre nous présentons la solution retenue pour la composante visuelle qui ne peut être considérée que comme preuve de concept : est-ce qu'une classe hétérogène visuellement et représentant le même concept de risque pourrait être reconnue avec une précision satisfaisante ? Nos résultats sur la reconnaissance sur des signaux des capteurs seront présentés par la suite.

Le tableau 4.4 présente la distribution des données LSC NTCIR 13 visuelles ré-annotées suivant notre taxonomie exposée dans le chapitre 3 pour la détection de situations à risques sémantiques pour les personnes fragiles.

Il est à noter que les données étant enregistrées toutes les minutes avec un intervalle fixe, l'absence de données est notée aussi, c'est la raison pour laquelle nous retrouvons une classe NaN (Not a Number) pour signifier l'absence de données.

4.5.1 Application des CNN sur la composante visuelle

Dans le traitement des données médicales et/ou de santé, nous avons observé le même phénomène que dans d'autres secteurs applicatifs concernant les problématiques liées à la détection d'anomalies. L'application des réseaux de neurones sur l'analyse des données s'est largement répandue pour les technologies eHealthcare.

4. Analyses de situations à risques sur des données Lifelog

Tableau 4.2 – Données visuelles annotées suivant la taxonomie BIRDS pour la détection de situations à risques sémantiques

Situations à risques	annotation automatique	Validation manuelle
Chute	6694	2208
Risque de chute	104	380
Risque d'accident domestique	331	494
risque de fraude	7715	205
Other	11946	23503

C'est également le cas pour les données du journal de vie (Lifelog). Récemment, il a été montré que différentes méthodes de Machine Learning notamment de Deep Learning ont une grande capacité à identifier des scènes complexes, même lorsque le dataset présente des données assez hétérogènes [Nait Aicha et al. \(2018a\)](#); [González-Díaz et al. \(2019\)](#); [Pandey et Janghel \(2019\)](#). Ces approches basées sur les réseaux de neurones profonds de classification de données nécessitent un effort important et coûteux pour étiqueter image par image lorsqu'on se base sur des modèles supervisés, même lorsque des outils d'annotation collaboratifs sont utilisés. Plusieurs recherches existent sur la détection d'activité sur le dataset [Gurrin et al. \(2019a\)](#). Alors que de nouvelles architectures profondes sont en cours de conception pour des tâches spécifiques dans les données Lifelog [Aimar et al. \(2019\)](#), les réseaux qui sont devenus standards du domaine de Deep Learning telles que ResNet [He et al. \(2016\)](#) conviennent à la reconnaissance de scènes égocentriques [González-Díaz et al. \(2019\)](#). La reconnaissance automatique de situations à risques qu'on qualifierait de sémantiques chez un patient fragile est un sujet peu abordé dans la littérature scientifique, probablement en raison de la difficulté à définir ces situations à risques mais aussi de la complexité que constitue l'enregistrement de ces situations avec un sujet fragile et plus spécialement dans un contexte écologique.

De ce fait, les données relatives aux situations à risques ne sont pas très fréquentes dans les données typiques du journal de vie ou dans les ensembles de données couramment utilisés [Gurrin et al. \(2019b\)](#). Par conséquent, un grand ensemble de données d'entraînement d'exemples positifs n'est pas facilement disponible pour entraîner un modèle robuste, il devient alors plus réaliste de recourir à des réseaux plus légers, tels que AlexNet [Krizhevsky et al. \(2012b\)](#). Dans ce travail, nous ne proposons pas une nouvelle architecture mais adaptons notre choix d'architecture, l'augmentation des données et le réglage des hyperparamètres à notre problème. Par ailleurs, nous nous intéressons à la classification binaire "un contre tous" ou le problème est posé comme la reconnaissance des événements de risque sémantique d'une seule classe.

Choix de l'architecture et réglage du CNN pour la reconnaissance des situations à risque

Dans ce travail, nous avons tout d'abord fait travailler avec l'architecture des standard ResNet26 et ResNet10. Néanmoins, nos tests préliminaires n'ont pas donné de résultats satisfaisants, nous avons donc appliqué AlexNet [Krizhevsky et al. \(2012c\)](#), et nos expérimentations montrent que des réseaux moins profonds, mais bien réglés permettent d'obtenir des précisions raisonnables. L'architecture d'AlexNet est illustrée dans la figure 4.5. C'est un réseau peu profond connu pour sa stabilité à l'entraînement.

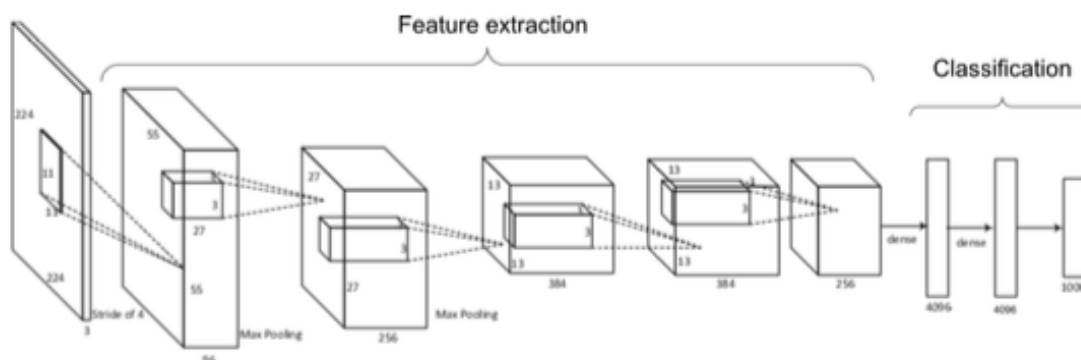


FIGURE 4.5 – Architecture AlexNet [Krizhevsky et al. \(2012a\)](#).

L'architecture AlexNet est proposée par Alex Krizhevsky et al. en 2012 [Krizhevsky et al. \(2012a\)](#). Il s'agit du premier modèle profond qui en 2015, à la compétition (ImageNet Large Scale Visual Recognition Competition) a surpassé les autres approches de machine learning plus classiques. À partir de là, il a donné naissance à la nouvelle ère de l'apprentissage à partir des données et a relancé l'intérêt sur les réseaux de neurones artificiels pour les problèmes de classification à grande échelle. L'architecture AlexNet proposée est un CNN à 8 couches qui se compose de couches convolutionnelles, de couches de Pooling, de couches d'activation et de couches dense. L'architecture est composée de 5 couches convolutionnelles, 3 couches de max pooling. Enfin, 3 couches denses entièrement connectées. Toutes les couches d'activation de cette architecture sont configurées avec des Rectified Linear Unit (ReLU). La fonction ReLU pour Unité Linéaire Rectifiée est définie par l'équation 4.1.

$$f(x) = \max(0, x) \quad (4.1)$$

Nous avons donc sélectionné l'architecture bien connue AlexNet pour entraîner nos sub-dataset sélectionnés depuis le dataset principal LSC NTCIR 13 avec une classification correspondant à notre taxonomie. Malgré le fait que l'architecture AlexNet date un peu, elle présente des avantages non négligeables notamment par sa simplicité et sa légèreté. AlexNet conserve principalement un faible coût de calcul, ce

qui est pratique pour nos expériences. Le réseau s'applique sur une base de données à faible échelle pour l'identification de situations à risques identifiées sur un dataset enregistré par un individu en bonne santé et ce, par la classification d'images. À ce stade, nous écartons d'autres réseaux plus profonds pour les raisons exposées plus haut dans la section. Dans la suite de cette section, nous évoquerons les datasets utilisés, l'augmentation des données effectuées pour éviter le sur-ajustement du modèle, la paramétrisation du modèle, avant de finir par présenter les résultats obtenus les plus probants.

Préparation des données visuelles

Nous avons donc mené des expériences pour reconnaître deux situations à risques sur la partie visuelle du jeu de données LSC NTCIR 13 [Gurrin et al. \(2019b\)](#). Il comprend 41664 images. Nous avons sélectionné quatre classes de risque sémantique, mais nous avons mené des expériences uniquement sur les classes plus sémantiques "risque de fraude" et "risque d'accident domestique" [4.4](#). Pour éviter les images trop similaires, nous avons limité notre sélection d'images du jeu de données NTCIR 13 à une image par minute, ce qui réduit le nombre total d'images à 21 152.

Nous avons procédé à une classification binaire sur les deux situations à risques choisies. Les images d'exemples positifs dans le jeu de données original LSC NTCIR 13 pour ces deux situations à risques sémantique ne sont pas nombreuses. Les exemples négatifs ont été sélectionnés proportionnellement au nombre de données. Les chiffres sont présentés dans le [tableau 4.3](#). La constitution de ces deux ensembles de données a été effectuée en prenant les échantillons positifs d'une classe et en sélectionnant aléatoirement $\times 1,5$ images du nombre de positifs dans le reste de l'ensemble de données.

Augmentation des données La méthode la plus simple et la plus courante pour réduire l'overfitting sur des données d'image est d'agrandir artificiellement l'ensemble de données en utilisant des transformations préservant les étiquettes. Les transformations préservant les étiquettes que nous appliquons à chaque image sont les suivantes : i) des rotations de 90° , ii) nous introduisons des retournements en miroir, dans toutes les directions, c'est-à-dire verticale, horizontale. Cette dernière transformation produit 2 variantes supplémentaires de l'image d'entrée, iii) nous mettons en place une translation à gauche et une translation à droite sur un total de 100 pixels et iiiii) nous rajoutons un bruit gaussien à l'image de moyenne = 0.1 et écart-type = 0.5. Totalisant ainsi, un total de 7 variantes par image. Les images augmentées sont seulement appliquées sur la phase d'entraînement du réseau.

La [figure 4.6](#) présente une image des images augmentées des d'images du LSC NTCIR 13

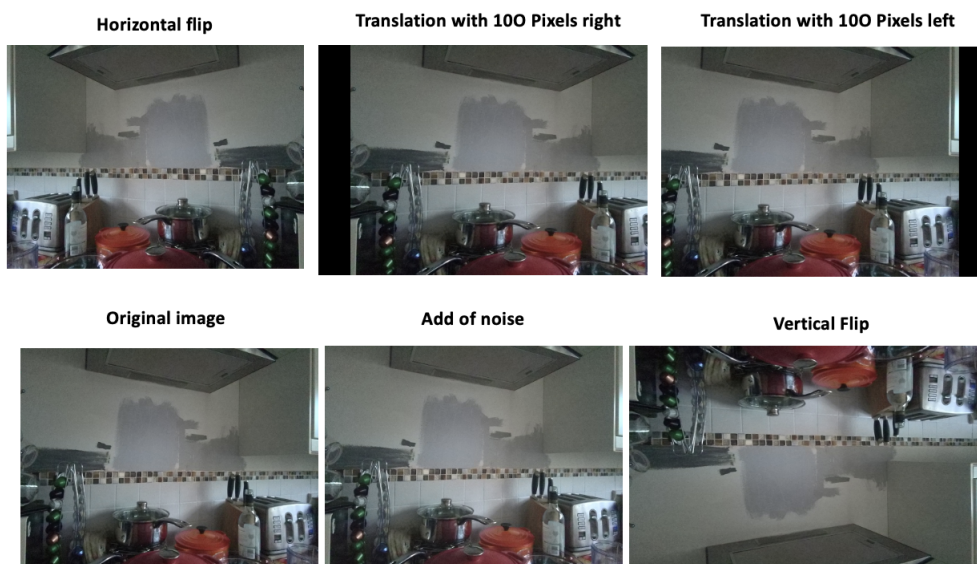


FIGURE 4.6 – Méthodes d’augmentation des données, rotation et retournement. Rangée du haut : remplissage flou. Rangée du bas : Remplissage de bruit gaussien.

Situation à risques	Échantillons positifs	Échantillons négatifs
Risque d’accident domestique	406	600
Risque de fraude	120	180

Tableau 4.3 – Ensembles de données utilisés pour la formation pour chaque situation de risque.

Le dataset

Enfin, pour le risque d’accident domestique, le jeu de données d’entraînement était composé de 4224 images, 201 images pour la validation et 101 pour le test. Pour le risque de fraude, le jeu de données d’entraînement se compose de 1254 images, 60 images pour la validation et 31 pour le test. Les poids du réseau AlexNet utilisé ont été initialisés aléatoirement.

4.5.2 Entraînement du modèle

Lors de l’entraînement du réseau, nous avons utilisé une taille de lot de 64 images pour la détection du risque d’accident domestique et une taille de lot de 10 images pour la détection du risque de fraude. Ce choix a été fait proportionnellement au nombre d’images originales dans le jeu de données d’entraînement. Comme méthode d’optimisation, nous avons utilisé la descente de gradient stochastique (SGD) avec

le momentum de Nesterov [Jia et al. \(2014\)](#). Afin de mettre à jour les paramètres W pour minimiser la fonction objective, $\mathcal{L}(W)$ la méthode adopte le schéma itératif,

$$\begin{aligned} V_{i+1} &= \mu V_i - \eta \nabla \mathcal{L}(W_i + \mu V_i) \\ W_{i+1} &= W_i + V_{i+1} \end{aligned} \tag{4.2}$$

où V_i désigne la vitesse, μ le coefficient de quantité de mouvement ($\mu = 0,9$), η le taux d'apprentissage, i l'itération et W_i les coefficients de filtrage.

La politique de décroissance exponentielle du taux d'apprentissage s'est avérée efficace dans les travaux où le nombre de données d'apprentissage disponibles n'est pas très élevé.

Nous l'utilisons donc : $\eta = \eta_0 \cdot \gamma^i$ où η_0 est le taux d'apprentissage initial, $\gamma \in [0, 1]$ donne la forme de la courbe de décroissance du taux d'apprentissage et i le nombre d'itérations d'apprentissage actuelles.

Avec une valeur plus faible de γ , le taux d'apprentissage diminue plus rapidement au début de l'apprentissage, et une valeur proche de 1, maintient une petite décroissance pour chaque itération.

Nous avons établi $\gamma = 0,97$ pour le risque d'accidents domestiques et des valeurs plus élevées entre 0,98 et 0,999 pour le cas du risque de fraude après plusieurs expériences, en raison de la plage de γ qui donne une décroissance appropriée du taux d'apprentissage est $[0,95, 0,99]$. Nous avons défini la valeur du taux d'apprentissage initial η_0 après plusieurs expériences en utilisant la méthode de bisection à $\eta_0 = 0.00125$ pour les deux situations à risques sur une barre de recherche entre $[0.0001, 0.1]$. Le nombre maximal d'époques était de 90.

L'apprentissage a été réalisé à l'aide du framework open-source MXNet [Chen et al. \(2015\)](#). L'entraînement a été effectué sur un seul serveur avec 2× Tesla P100-PCIE-16GB.

Résultats de la détection de deux situations de risque sémantique

Ici, nous présentons les résultats de la détection des deux risques sémantiques extraites du dataset LSC NTCIR 13 avec le modèle AlexNet.

Nous procédons à l'analyse des courbes d'apprentissage du modèle pour les deux situations de risque et les illustrons dans la figure 4.7 et la figure 4.8, en notant que les hyperparamètres du modèle AlexNet ont été optimisés pour chacune des situations de risque séparément. Comme l'illustre la figure 4.7, pour le risque d'accidents domestiques, nous obtenons plutôt de bons résultats et le modèle ne se surajuste pas sur l'ensemble des données d'entraînement, nous obtenons une précision de 81,33 % sur l'ensemble de test assez rapidement à l'époque 14. Pour le risque de fraude, nous obtenons une bonne précision de 96,88 % sur le jeu de test à l'époque 22, voir figure 4.8 après ce temps, le surajustement est observé et la perte de validation diverge.

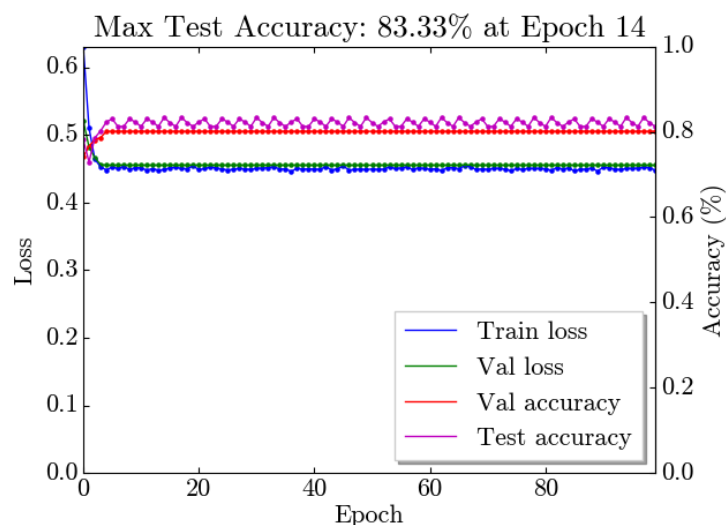


FIGURE 4.7 – Courbes d'AlexNet pour le risque d'accident domestique. La précision maximale (83,33 %) est atteinte à l'époque 14.

4.5.3 Application d'un LSTM sur des données (signal 1D)

Dans la section 4.5, une méthode d'annotation des situations à risques et de détection de celles-ci en utilisant uniquement la composante visuelle de l'ensemble de données NTCIR 13/LSC a été proposée. Conformément au "modèle de situations à risques" que nous avons présenté 3, nous avons besoin de détecter la situation à risques également dans les signaux sensoriels. Dans ce but, nous avons appliqué un réseau LSTM Hochreiter et Schmidhuber (1997).

Les réseaux neuronaux à mémoire à long et court terme (LSTM) sont largement utilisés pour la classification des séries temporelles car ils construisent des liens à très long terme pendant l'apprentissage Hochreiter et Schmidhuber (1997). Ils retiennent les informations pertinentes des entrées précédentes et utilisent ces informations pour modifier la sortie actuelle. Nous avons appliqué une architecture LSTM standard pour la détection de risques sur des signaux de capteurs.

4.5.4 Architecture LSTM

L'unité LSTM de base est composée de cellules avec une porte d'entrée, une porte de sortie et une porte d'oubli. La particularité des LSTMs consiste dans la manière dont l'état caché h est traité : dans le cas des RNNs simples, le traitement de la récurrence est assuré par une fonction \tanh . Pour les LSTMs, ce traitement est remplacé par une "cellule mémoire". La cellule LSTM est caractérisée par un noeud central, contenant l'état interne (ou mémoire) de la cellule, et un certain nombre de "portes" des trois catégories mentionnées. Ces portes sont utilisées pour gérer le stockage séquentiel des informations (portes d'entrée et d'oubli) et le rôle de l'état

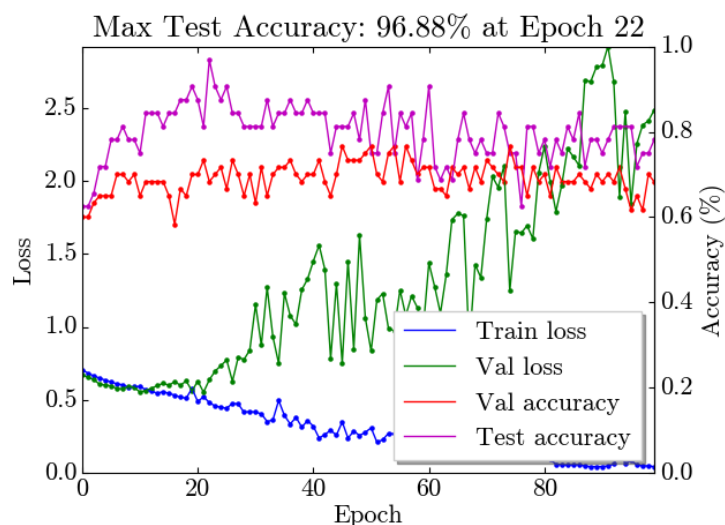


FIGURE 4.8 – Courbes d’AlexNet pour le risque de fraude. La précision maximale (96,88 %) est atteinte à l’époque 22.

interne dans la production de chaque sortie (porte de sortie). En fermant la porte d’entrée, par exemple, les nouveaux événements sont moins pris en compte. Les LSTM utilisent le concept de gating pour surmonter le problème de la disparition et de l’explosion du gradient : [Greff et al. \(2016\)](#).

L’architecture LSTM, avec le modèle appliqué sont présentés plus en détails dans le chapitre 6.

Ainsi, des expériences ont été menées pour reconnaître des situations de risque sémantique sur une partie des données biométriques et des données d’activité humaine à partir de signaux de l’ensemble de données NTCIR 13/LSC : [Gurrin et al. \(2019b\)](#). Nous avons utilisé les mesures suivantes :

- La fréquence cardiaque ;
- La réponse électrodermale ;
- La température de la peau ;
- Le nombre de pas.

Nous avons sélectionné quatre classes sémantiques de risque : *chute*, *risque de chute*, *risque de fraude* et *risque d’accident domestique*, ainsi qu’une classe sans risque - "*Autre*".

Les situations à risques ont été annotées en utilisant uniquement les données d’image, comme indiqué dans la section 4.4. Dans les signaux des capteurs, certaines données sont manquantes pendant l’enregistrement en raison d’une défaillance de certains capteurs du dispositif de mesure. Les données manquantes ont été annotées comme NaN (Not a number). Les statistiques des risques, de l’absence de risque et des NaN sont présentées dans le tableau 4.4 et la figure 4.9.

On observe sur le 4.4 et la figure 4.9 une majorité de données notées "Other" pour

4.5. Reconnaissance des risques sémantiques sur les données du corpus NTCIR 13

Tableau 4.4 – Distribution des données LSC NTCIR 13 obtenue avec une annotation correspondant à notre taxonomie BIRDS.

Situations à risques	Données selon les timestamps	Pourcentage
Risk of fraud	120	0.22 %
Risk of fall	347	0.63 %
Risk of domestic accident	406	0.73 %
Fall	2181	3.96 %
Other	21152	38.43 %
NaN	30826	56.01 %
Total	55032	100 %

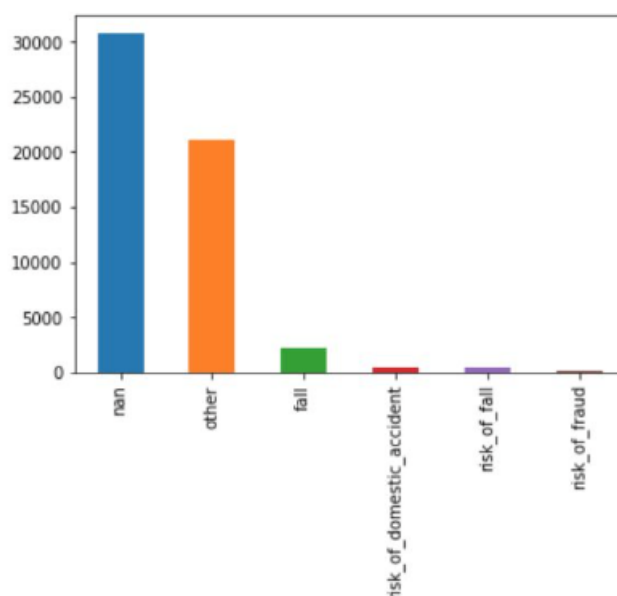


FIGURE 4.9 – Distribution des données LSC NTCIR 13 selon la taxonomie BIRDS.

"Sans risque". Le risque de fraude sur les données biométriques ne représente que 0.55 % de la situation dite "other", soit une proportion très faible, pouvant rendre l'apprentissage du réseau très difficile sur ce cas. La valeur NaN est majoritaire et représente 56 % de la distribution. Ces valeurs ne seront pas considérées dans les calculs de prédiction. Ce label est obtenu uniquement à partir des images. Le but du travail ici présenté est de pouvoir prédire le risque que la personne encourt, uniquement grâce aux données biométriques. Ces données sont également associées à une période, et donc à des risques grâce aux images. L'intérêt est donc de créer un réseau de neurones, uniquement via ces données biométriques, capable de déterminer avec une précision suffisante le risque encouru par une personne.

Experimentations et resultats

Ainsi, des expériences ont été menées pour reconnaître des situations de risque sémantique sur une partie des données biométriques et des données d'activité humaine à partir de signaux de l'ensemble de données LSC [Gurrin et al. \(2019b\)](#). La fréquence cardiaque, la réponse galvanique de la peau, la température de la peau et le nombre de pas ont été utilisés. Nous avons sélectionné quatre classes sémantiques de risque : chute, risque de chute, risque de fraude et risque d'accident domestique, ainsi qu'une classe sans risque - "Autre". Les situations à risques ont été annotées en utilisant uniquement les données d'image, comme indiqué dans la section 4.4. Dans les signaux des capteurs, certaines données sont manquantes pendant l'enregistrement en raison d'une défaillance du capteur. Les données manquantes ont été annotées comme NaN (Not a number). Les statistiques des risques, de l'absence de risque et des NaN sont présentées dans le tableau 4.4

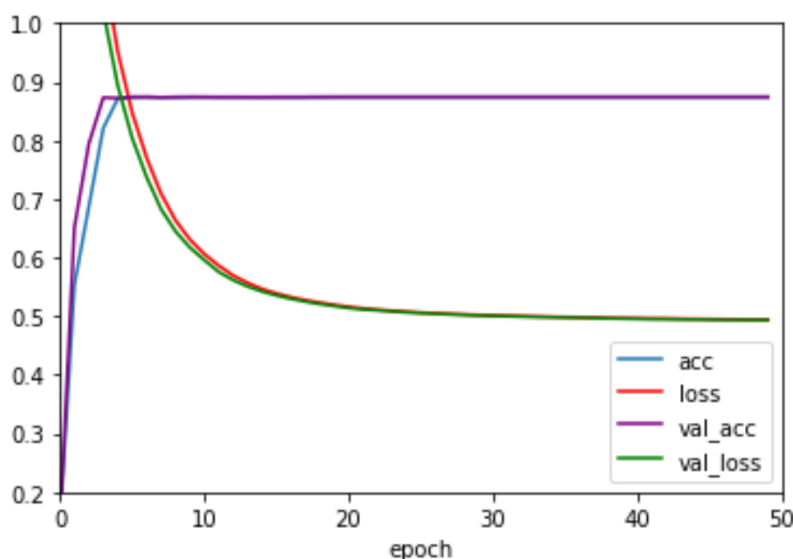


FIGURE 4.10 – Courbes LSTM obtenues sur le jeu de données LSC.

Le modèle LSTM a été entraîné sur ces données après que les données NaN aient été supprimées. L'algorithme d'optimisation employé est Adam [Kingma et Ba \(2014\)](#) avec un taux d'apprentissage fixe de 0,001 choisi expérimentalement. Le framework utilisé était Keras (<https://keras.io/>). Les courbes de précision et de perte pendant l'apprentissage sont illustrées dans la figure 3.10. Malgré une précision satisfaisante de 80 % l'analyse détaillée de ces résultats a montré que cette précision est totalement biaisée par la surreprésentation de la classe "Autre". En effet, pour toutes les classes sauf "Autre", les métriques de Rappel et de Précision sont proches de 0. Les classes ne sont pas équilibrées. En effet, en revenant au tableau 4.4, nous pouvons constater que, par exemple, le risque de chute ne représente que 0,22 % de toutes les données, alors que la présence de la classe "Autre" est de 38,43 %.

Avec une distribution aussi déséquilibrée, il est inutile d'appliquer une quelconque technique d'augmentation des données. En effet, dans nos expériences, l'ajout d'un bruit gaussien aux mesures des signaux dans les situations à risques pour augmenter les données n'a pas apporté d'amélioration. Même un rééquilibrage de données n'aurait pas pu résoudre le problème car la fréquence d'acquisition de ces données biométriques étant à maximum de 0.1 hz, il n'est pas possible d'avoir une détection fiable ni robuste. Ainsi, en revenant à notre modèle de situation à risques de premier type, voir chapitre 3, il apparaît clairement qu'un scénario général d'enregistrement d'un ensemble de données LSC n'est pas adapté à une collecte de données suffisante pour les situations à risque.

4.6 Conclusion

Dans ce chapitre, nous avons présenté la reconnaissance de situations à risques sémantiques, appliquée à des données Lifelog du dataset NTCIR 13 enregistré avec un sujet adulte "bien portant".

Nous avons proposé un schéma de correspondance de concepts du corpus avec nos situations à risques pour proposer une ré-annotation plus en conformité avec notre taxonomie. Nous avons testé des architectures de réseaux de neurones, connues pour la détection de situations à risques, d'abord sur les données visuelles nous avons ensuite testé une architecture LSTM sur les données signaux 1D.

Pour les données visuelles, nous avons étudié et choisi les hyperparamètres et les méthodes d'augmentation qui ont permis d'obtenir des scores satisfaisant de reconnaissance.

Après des résultats peu encourageants obtenus sur les données signal 1D, nous avons conclu que la faible fréquence d'échantillonnage de ces données et le fait que ces données ont été enregistrées par une personne saine et sans simulation des risques ne permettaient pas une bonne reconnaissance.

Avec l'avancement du projet, nous avons enregistré le corpus BIRDS qui contient les situations à risques simulées et nous nous sommes focalisés sur l'analyse de ses données. Dans le chapitre suivant, nous présentons le dataset BIRDS.

Chapitre 5

Dataset Birds

5.1 Introduction

Dans ce chapitre, nous présentons les données collectées avec le dispositif BIRDS. Les données ont été enregistrées suivant un scénario d’acquisition précis avec des situations à risques simulées dans un environnement écologique réel, à domicile. Ce chapitre présente le déroulement de l’acquisition des données suivant les scénarii représentant les différentes situations à risques, l’annotation de ces données ainsi que les pré-traitements que nous avons effectués sur ces données.

Dans un premier temps, nous décrivons le protocole d’enregistrement des données BIRDS en mettant l’accent sur les scénarios d’enregistrement pour chaque situation à risque. Ensuite, nous décrivons l’interface d’annotation des données BIRDS, ainsi que les évolutions qui ont été apportées au fil de l’évolution de la taxonomie.

Dans la continuité, nous explicitons les processus de pré-traitement des données mis en place afin de pouvoir les exploiter. Nous présentons les données qui en résultent et introduisons leur représentation plus formelle qui servira de support à nos contributions présentées dans la suite du manuscrit.

Enfin, nous parlerons des pré-traitements réalisés sur les données. Nous parlerons ainsi du problème des données manquantes et de leur imputation et nous décrirons aussi le rééquilibrage des classes réalisé afin de limiter la classe *Sans risque* par rapport aux classes décrivant les situations à risques enregistrées.

5.2 Dataset BIRDS

Dans cette section, nous présenterons le protocole d’enregistrement des données élaboré avec l’équipe multidisciplinaire présentée dans le chapitre 1, qui a participé au projet. Dans un second temps, nous présenterons le dataset BIRDS.

5.2.1 Protocole d’enregistrement des données

Nous avons mis en place une expérience miroir entre une université en Europe et une université en Amérique du Nord pour l’enregistrement du corpus, que nous avons appelé BIRDS (Bio Immersive Risk Detection System) : les mêmes dispositifs

sont utilisés, connectés à une application Android développée à cet effet et servant de concentrateur des données collectées par les capteurs.

Dans le chapitre 3, nous avons présenté le système d'acquisition conçu pour l'enregistrement des données motrices, physiologiques et contextuelles. Ici nous rappelons rapidement les fonctionnalités de ces capteurs.

- Tout d'abord, un bracelet développé par Empatica, le E4, qui est un dispositif médical portable, est utilisé. Il offre une acquisition de données physiologiques en temps réel, permettant aux chercheurs de réaliser une analyse approfondie et une visualisation des données. Ce bracelet est équipé d'un accéléromètre, d'un capteur d'activité électrodermale (EDA), d'un capteur PPG, d'un capteur thermopile infrarouge et d'un émetteur Bluetooth.
- Un dispositif portable porté sur la poitrine, MetaMotionR, permet de surveiller en temps réel et en continu les mouvements de la personne. Il offre les mesures de différents capteurs, tels que l'accéléromètre, le gyroscope, le magnétomètre.
- Une caméra vidéo portée sur l'épaule du sujet. Cette dernière est directement connectée à un téléphone considéré comme le contrôleur principal de l'ensemble du dispositif et vient compléter les données physiologiques et actimétriques. La caméra enregistre 3 secondes de vidéo toutes les 10 secondes avec une fréquence d'images de 10 fps. Nous la considérons comme un capteur contextuel.

Le dispositif a été porté par trois sujets différents.

Le premier sujet est une femme de 26 ans. Elle a enregistré des données avec le dispositif pendant 38 jours, les 12 premiers jours d'enregistrements ont servi pour tester le dispositif et améliorer son ergonomie. Ensuite pendant 26 jours sur deux périodes d'enregistrement différentes, le sujet bien familiarisé avec le dispositif a enregistré des données en essayant de produire les différentes situations à risques expliquées précédemment dans le chapitre 3. La collecte de données s'est faite au domicile de ce dernier.

Le deuxième sujet est une femme, âgée de 62 ans, en bonne santé. Nous avons effectué 10 jours d'enregistrement. Différentes situations à risques ont été réalisées dans son environnement de vie réel (son domicile). Sur ces dix jours, nous avons fait remonter les problèmes pouvant être rencontrés par des personnes âgées non habituées à l'utilisation des nouvelles technologies.

Enfin, *le un troisième sujet*, est une femme de 63 ans, présentant un certain niveau de fragilité en lien avec un antécédent récent de cancer. Elle a effectué 19 jours d'enregistrement. Sur ces 19 jours, le sujet a pu produire des situations à risques mentionnées plus haut avec des réponses physiologiques réelles et dans son environnement de vie réel (son domicile).

Néanmoins, l'intégration de ces données dans le corpus BIRDS en suivant les procédures du pré-traitement, de l'imputation et de l'équilibrage que nous présentons dans ce chapitre est en cours.

Le diagramme des enregistrements avec leurs durées est présenté dans la figure 5.1 ci-dessous et porte sur les 21 jours d'enregistrement par notre premier sujet. Ces données ont été utilisées pour l'analyse automatique de la détection des situations à risques. Quant aux enregistrements effectués pendant 9 jours par un second sujet, nous les avons considérés uniquement pour l'étude de l'acceptabilité, de l'ergonomie du dispositif et de la conception du protocole.

Scénario d'enregistrement. Avec l'avancement de l'application des objets IoT pour les études comportementales dans l'objectif de la reconnaissance automatique des événements et des actions dans les données enregistrées par les capteurs portés, deux stratégies sont possibles.

La première consiste à enregistrer les données sur un maximum de personnes et d'entraîner les modèles sur ces données multi-personnes. Telle est la démarche des auteurs du [Anguita et al. \(2013\)](#) qui a engendré le corpus UCIHAR de données accéléromètres par le téléphone mobile porté dans la poche des personnes. Ensuite, ces données sont utilisées pour la reconnaissance des actions. Ainsi les modèles peuvent être entraînés sur une grande quantité de données très hétérogènes.

Par ailleurs, notre problématique consiste en une détection des situations à risques beaucoup plus sémantiques, nous distinguons les risques contextuels, liés à l'environnement écologique propre au sujet, par exemple sa cuisine, sa cage d'escalier, son sol dans la pièce d'habitation (cf. chapitre 3.16). Il est illusoire à supposer que le système de la détection des risques avec l'analyse de la vidéo en provenance de la caméra portée puisse présenter des scores satisfaisants pour l'usage réel, si il n'est pas entraîné sur les données similaires à celles acquises à l'étape de généralisation. Donc l'enregistrement du corpus BIRDS est effectué selon le scénario que nous appelons "*une personne*" ou "*one person scenario*".

Dans ce scénario, une collecte de données préalable à toute possibilité de la détection des situations critiques est nécessaire avec une annotation de la vérité terrain. Pendant cette phase d'enregistrement, le sujet n'est pas "protégé" par le dispositif à proprement parler. Il est néanmoins envisageable de compléter le dispositif par des outils déjà disponibles sur le marché comme détecteurs de chute [Wang et al. \(2020\)](#) ainsi, apporter une sécurité aux sujets volontaires. Une tel enrichissement sort du cadre de nos travaux. Il s'agit de la faisabilité du scénario de l'enregistrement et du développement des outils de l'IA pour la détection des risques, tout le corpus BIRDS a été analysé postérieurement aux enregistrements. Un découpage ordinaire en l'ensemble d'apprentissage et l'ensemble de test a été effectué pour l'entraînement des réseaux profonds proposés.

Faciliter l'annotation de l'enregistrement par un scénario adapté. Dans notre démarche d'étude pluri-disciplinaire, nous avons mis en place un protocole qui permettant une annotation aisée postérieurement à l'acquisition de données dans l'objectif de la création de la vérité terrain. Ainsi, il a été demandé aux sujets participant à la collecte de données de noter dans un journal de bord à la minute près la situation à

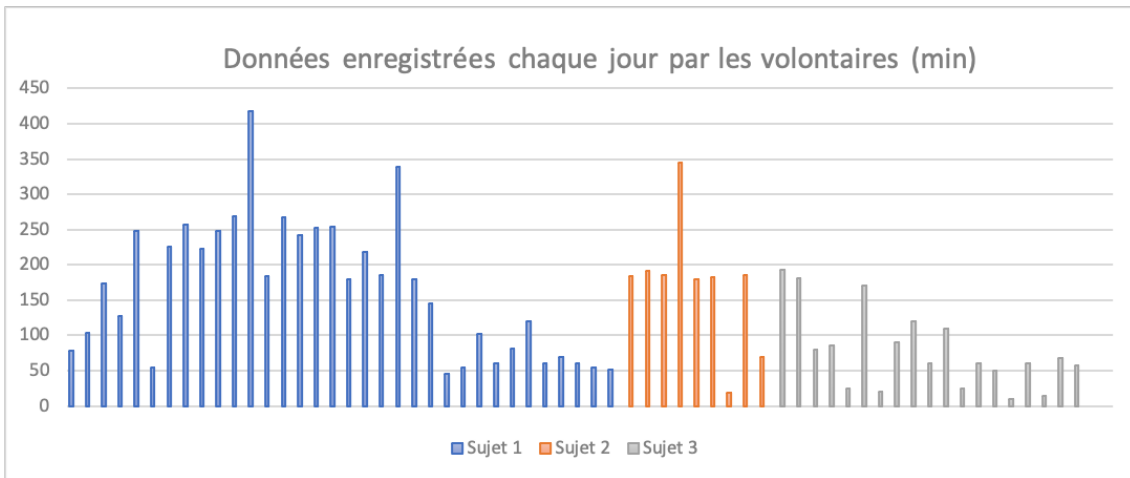


FIGURE 5.1 – Diagramme de la durée d'enregistrement du corpus BIRDS.

risques simulée provoquée.

Respect de la Norme RGPD Toute la démarche d'enregistrement sur des sujets volontaires a été conçue conformément à la réglementation européenne RGPD [Mattatia \(2019\)](#), et ce, après concertation avec le DPO de l'université de Bordeaux. La finalisation du dossier est en cours. Dans le même esprit, les données enregistrées sur des volontaires sont totalement anonymisées. Ceci comprend :

- L'indisponibilité des informations susceptibles d'indiquer la localisation des sujets (GPS) ;
- Floutage systématique des visages et des silhouettes d'autrui présents dans les enregistrements vidéo ;
- Floutage de tous les éléments visuels susceptibles faciliter la reconnaissance des personnes, comme par exemple les photos ou les reflets dans les miroirs.

5.2.2 Description du corpus BIRDS

Le dataset BIRDS a été enregistré avec le dispositif développé à cet effet et présenté dans le chapitre 3. Les données mesurées par les différents capteurs que compose ce dispositif sont collectées à travers l'application BIRDS. Ces données sont dans un premier temps enregistrées sur le téléphone avant d'être envoyées sur les serveurs de traitement du Laboratoire LaBRI. Nous avons constitué une nomenclature avec une architecture bien spécifique pour ces données. Le format des données est le suivant : i) un répertoire principal pour chaque jour d'enregistrement avec le nom "dataset-JJ-MM-AAAA", ii) dans ce répertoire nous avons un fichier regroupant les données et ce, pour chaque capteur et iii) dans le répertoire principal nous avons un répertoire "VIDEO" regroupant les vidéos enregistrées à ce moment là et enfin un répertoire "SOUND" regroupant les fichiers audio enregistrés.

Dans son état actuel, le corpus BIRDS contient des enregistrements de 67 jours

5. Dataset Birds

Tableau 5.1 – Description du corpus BIRDS.

	Somme totale	Moyenne	Écart-type
Recording time (Min)	8197	141.33	94
Global data volume (MB)	70806	1220.79	3238.62
Video data volume (Mo)	39224	676.27	460.468



FIGURE 5.2 – Exemples d’images provenant du jeu de données LSC Gurrin et al. (2019b).

pour un volume global de 8197 min. La durée d’enregistrement par jour varie de près de 1 heure à cinq heures, la durée dépend aussi du sujet comme nous pouvons le voir sur la figure 5.1. Les données audio ainsi que les données GPS ne sont pas utilisées pour des raisons de confidentialité et ont été retirées du dataset final. Le tableau 5.1 décrit les informations globales du dataset BIRDS enregistré à ce jour.

Les figures 5.2 et 5.3 illustrent le contenu des deux ensembles de données LSC NTCIR 13 et BIRDS. Dans 5.2, les images du jeu de données LSC Gurrin et al. (2019b) sont présentées et dans 5.3, les images vidéo de différentes scènes de notre corpus BIRDS sont affichées. Les deux ensembles de données ont été enregistrés avec des positions de caméra différentes. Le jeu de données LSC est enregistré avec une caméra portée sur la poitrine, tandis que le jeu de données BIRDS - avec une caméra portée sur l’épaule. Cette position a été définie comme l’ont proposé les ergothérapeutes dans des recherches antérieures, spécifiquement adaptées aux sujets âgés. Les images vidéo de la figure 5.3 représentent diverses situations à risque. Nous pouvons remarquer que le corpus BIRDS est plus réaliste par rapport à l’environnement domestique d’une personne âgée.

Comme nous pouvons le voir dans le tableau 5.2, la quantité de situations à risques n’est pas encore suffisante pour effectuer un entraînement significatif du réseau neuronal et devra être étendue lors des prochaines sessions d’enregistrement, cela est aussi dû au fait qu’une anomalie est par définition un événement rare. Le scénario de reconnaissance des risques doit être adapté à l’environnement de chaque sujet. Nous avons décidé de séparer notre corpus en quatre parties en fonction de l’évolution des situations à risques ajoutées à notre taxonomie de base (à savoir l’affinage des risques de chutes).

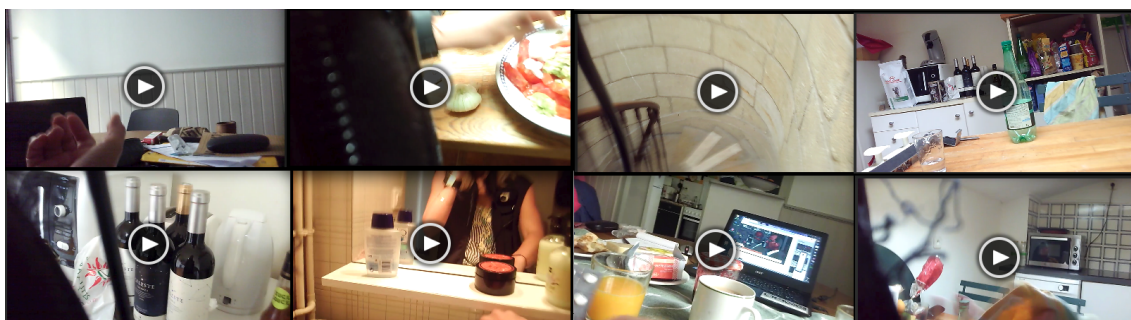


FIGURE 5.3 – Données du Corpus BIRDS.

5.3 Annotation des données

L'annotation des données est une étape importante pour entraîner des réseaux de neurones supervisés. Nous pouvons distinguer trois manières d'annoter un ensemble de données : (i) automatique, (ii) manuelle et (iii) semi-automatique. Il est habituel d'avoir d'abord un outil automatique avant de compléter cette annotation par un étiquetage manuel. Nous allons présenter, dans ce qui suit notre interface d'annotation et le processus d'annotation mis en place pour annoter le dataset BIRDS.

Le processus d'annotation a été conçu comme une méthode d'annotation manuelle en prenant en compte le journal de bord où sont notées les activités menées au cours des enregistrements à la minute près.

Les sessions d'annotation ont été réalisées par des neuropsychologues spécialisés sur les thématiques de la vieillesse et de la fragilité suivant la taxonomie mise en place. Une plateforme d'annotation permettant de présenter les données du dataset BIRDS a été développée (figure 5.4).

La plateforme d'annotation récupère le répertoire à annoter avec les données des capteurs dans chaque fichier réservé à chaque capteur et un répertoire pour les données vidéos. Une opération de concaténation de toutes les données 1D nous permet d'obtenir un seul fichier sous format "CSV" avec une colonne "label" ajoutée à la fin et initialisée à "0" qui représente la situation "*sans risque*" (voir figure 5.5. Cette opération d'initialisation de toutes les données à "*sans risque*" nous permet de nous concentrer lors de l'annotation des activités seulement sur les situations simulées et dites à risque puisque les autres données sont déjà taguées "sans risque".

A l'aide de cette interface, l'annotateur visualise les données des capteurs physiologiques, actimétriques et la vidéo, grâce au journal de bord, il repère la plage horaire où une situation à risques a été simulée, actionne la vidéo pour confirmer la situation à risques et l'étiquette suivant la taxonomie affichée sur la plateforme.

On observe que les données sont donc annotées par inférence visuelle et suivant le journal de bord dans lequel les volontaires qui enregistrent le dataset mentionnent leurs activités. Ceci donne alors une fenêtre plus précise pour confirmer l'annotation

5.3. Annotation des données

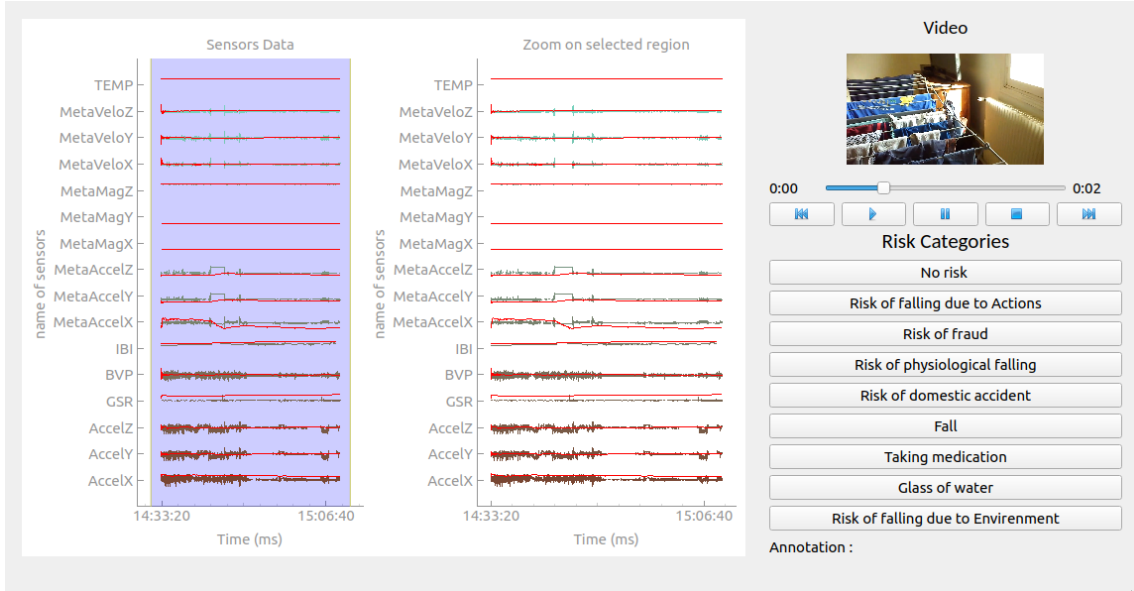


FIGURE 5.4 – Interface d’annotation BIRDS.

TIMESTAMP	ACCELERATION	GSR	BVP	IBI	META_ACCELERATION	META_MAGNETICFIELD	META_VELOCITY	TEMP	LABEL
1623907249529	[x:-33.0,y:18.0,z:61.0]	{0.06692029}	{-146.62592}	{0.7187829}	{x:0.37615967,y:-0.7651367,z:-0.52819824}	{x:-0.002048,y:-0.002048,z:0.0013313125}	{x:0.06097561,y:0.30487806,z:-0.24390244}	{33.97}	0
1623949649055	NaN	NaN	NaN	{0.7187829}	NaN	NaN	NaN	NaN	0
1623949649056	NaN	NaN	{22.595856}	NaN	NaN	NaN	NaN	NaN	0
1623949649058	NaN	NaN	{13.318359}	NaN	NaN	NaN	NaN	NaN	0
1623949649064	NaN	NaN	NaN	NaN	{x:-0.50671387,y:0.83825684,z:0.025817871}	NaN	NaN	NaN	0
1623949649065	[x:-36.0,y:-6.0,z:-51.0]	NaN	NaN	NaN	{x:-0.5054321,y:0.83807373,z:0.030700684}	NaN	NaN	NaN	0
1623949649068	[x:-36.0,y:-6.0,z:-52.0]	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0
1623949649074	NaN	NaN	{36.543945}	NaN	NaN	NaN	NaN	NaN	0
1623949649077	NaN	NaN	{37.645386}	NaN	{x:-0.5288086,y:0.83825684,z:0.042114258}	NaN	NaN	NaN	0
1623949649080	NaN	NaN	{-12.711975}	NaN	NaN	NaN	NaN	NaN	0
1623949649082	NaN	NaN	NaN	NaN	NaN	NaN	{x:1.402439,y:1.6463414,z:0.79268295}	NaN	0
1623949649107	NaN	NaN	NaN	NaN	NaN	NaN	{x:1.402439,y:1.6463414,z:-0.06097561}	NaN	0
1623949649110	NaN	NaN	NaN	NaN	NaN	NaN	{x:0.73170733,y:1.2195122,z:0.12195122}	NaN	0
1623949649114	NaN	NaN	{28.954895}	NaN	NaN	NaN	NaN	NaN	0
1623949649115	NaN	NaN	{34.77777}	NaN	NaN	NaN	NaN	NaN	2
1623949649118	NaN	NaN	{26.12677}	NaN	NaN	NaN	NaN	NaN	2
1623949649119	NaN	NaN	{1.4452167}	NaN	NaN	NaN	NaN	NaN	2
1623949649122	NaN	NaN	{1.4439348}	NaN	NaN	NaN	NaN	NaN	2
1623949649172	NaN	NaN	NaN	NaN	NaN	NaN	NaN	{30.15}	2
1623949649174	NaN	NaN	NaN	NaN	NaN	NaN	{x:1.4634147,y:1.2804878,z:0.30487806}	NaN	2
1623949649181	NaN	NaN	{29.052307}	NaN	NaN	NaN	NaN	NaN	2
1623949649186	NaN	NaN	{1.4439348}	NaN	NaN	NaN	NaN	NaN	2
1623949649199	NaN	NaN	NaN	NaN	{x:-0.5036621,y:0.8390503,z:0.022216797}	NaN	NaN	NaN	2
1623949649248	NaN	NaN	NaN	NaN	NaN	NaN	NaN	{30.15}	2
1623949649257	NaN	NaN	NaN	NaN	NaN	NaN	{x:0.9756098,y:3.0487804,z:1.8902439}	NaN	2
1623949649260	NaN	NaN	NaN	NaN	NaN	NaN	NaN	{30.21}	0
1623949649263	NaN	NaN	NaN	NaN	NaN	NaN	NaN	{30.21}	0
1623949649266	[x:-36.0,y:-6.0,z:-52.0]	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0
1623949649267	[x:-36.0,y:-6.0,z:-51.0]	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0

FIGURE 5.5 – Données du Corpus BIRDS.

des activités sur les signaux pour enfin la labellisation de ces données. Dans le fichier obtenu à la sortie de l'interface d'annotation, la labellisation se fait suivant les timestamps des données des signaux 1D sur le fichier récapitulatif 5.5. On précise que malgré notre appui sur le journal de bord que les sujets nous transmettent, nous visualisons toutes les données avec toutes les vidéos pour confirmer qu'il n'y a pas d'oubli et que les données sont bien annotées.

La taxonomie a été mise en place en collaboration avec les experts travaillant sur le sujet et a évolué au cours du projet, le tableau 3.1 résume la taxonomie utilisée pour enregistrer ce corpus. La taxonomie comprend en définitive 9 classes de situations à risques identifiées chez les personnes fragiles, à savoir :

Sans risque : scènes de vie ne présentant pas de situations à risques ;

Chute : chute accidentelle qui est survenue ;

Risque de chute environnemental : marcher sur un sol glissant, monter des escaliers, trébucher sur des obstacles ;

Risque de chute : due à des actions trébucher, soulever un objet, se précipiter, balayer le sol ;

Risque de chute physiologique : inclut l'accroupissement, la position assis debout ;

Risque d'intrusion d'une tierce personne : un évènement incluant la présence d'une personnes étrangère devant la porte d'entrée pendant quelques minutes.

Risque d'accidents domestique : inclut les brûlures, l'utilisation de couteaux, la manipulation d'ustensiles "dangereux" et le repassage ;

Oubli de prise de médicament : nous détectons la prise de médicament pour ensuite pouvoir dire s'il y a un oubli de prise de médicament ;

Risque de déshydratation : nous détectons la prise de boissons non alcoolisés.

5.4 Pré-traitement des données

Avant de pouvoir utiliser les données réelles BIRDS enregistrées dans des conditions écologiques pour l'entraînement des réseaux profonds, nous avons conçu une série de pré-traitements permettant de synchroniser et de "nettoyer" ces données.

5.4.1 La prise en compte du déphasage initial des capteurs

Comme nous avons pu le voir dans le chapitre 3, le dispositif BIRDS est constitué de différents capteurs avec des horloges internes propres à chaque système IoT. Il est à noter que chaque capteur possède des propriétés physiques qui lui sont propres, et donc avec des temps de démarrage différents. Ceci crée un déphasage

entre les différents capteurs au début de l'enregistrement y compris la caméra vidéo. À travers les différents enregistrements effectués pour tester le dispositif, nous nous sommes rendus compte qu'il arrive parfois d'avoir des déphasages allant de quelques millisecondes à plusieurs minutes entre le démarrage des différents capteurs. Le tableau 5.3 nous montre que sur un même dispositif IoT, nous pouvons avoir un temps de démarrage propre à chaque capteur.

Capteurs	Déphasage "ms" ϕ
EDA - HR	911
HR - Temp	1484
Temp - EDA	573
Acc - EDA	422
Acc - Band-Acc	870

Tableau 5.3 – Différence de timestamp entre le début de collecte des données.

Pour résoudre ce problème de déphasage, nous avons mis en place un algorithme qui compare les timestamps de chaque première donnée reçue par smartphone y compris les données vidéos et il ne garde que les données reçues à partir du moment où nous avons reçu au moins une seule donnée par capteur.

5.4.2 Anonymisation du corpus par image

Compte tenu des préoccupations liées à la protection de la vie privée, la question de l'anonymisation des données des journaux de vie est importante.

La norme ISO/TS 25237:2008 définit l'anonymisation comme «un processus qui supprime l'association entre l'ensemble de données identifiant et le sujet des données». Il est sous-entendu, que suite à un processus d'anonymisation, les données sont présentées sous une forme qui ne permet pas d'identifier les individus et dont la combinaison avec d'autres données ne devrait pas permettre de les identifier.

Le processus d'anonymisation des données est une activité complexe dans le sens où elle est soumise à deux risques :

- Le risque de divulgation si des données confidentielles et / ou sensibles sont accessibles par des utilisateurs non autorisés ;
- Le risque de fournir des données de faible valeur si le processus d'anonymisation dissocie les données et ne permet pas d'apporter des réponses aux problématiques posées.

Dans le cadre de mise en libre-service de notre dataset BIRDS, il est essentiel d'aboutir à une anonymisation du corpus tout en fournissant un dataset suffisamment riche pour pouvoir identifier les situations à risques présentes dans notre taxonomie. Pour cela nous avons identifié les données pouvant aider à remonter aux volontaires qui ont réalisé ces enregistrements. Nous avons alors identifié :

- Les données audio ;



FIGURE 5.6 – Exemples du floutage.

- Les données GPS ;
- Les données vidéo.

comme étant, les données donnant suffisamment d'informations pour pouvoir remonter jusqu'au sujet. De ce fait nous avons décidé de retirer les données audio et GPS. Quant aux données vidéo, qui présentent une part importante de notre dataset, au lieu de les retirer, nous avons décidé de les rendre anonymes. L'anonymisation à laquelle nous avons procédé est une anonymisation ordinaire pratiquée notamment dans le corpus NTCIR décrit dans le chapitre 4 et qui consiste à anonymiser les images vidéo par le floutage des visages et des objets pouvant aider à remonter au sujet.

Pour une grande quantité de données, il est nécessaire d'automatiser le processus. Ainsi le floutage des visages "automatique" consiste en deux étapes :

- Détecter les visages présents dans chaque image vidéo ;
- Appliquer un filtre pass-bas fort sur la zone de visage détectée.

Pour ce faire, nous avons utilisé des outils OpenCV [Bradski \(2000\)](#) avec un détecteur de visage par apprentissage profond, basé sur le cadre Single Shot Detector (SSD) avec un réseau de base ResNet, pour chaque image de la vidéo originale. Le visage détecté est flouté et les images sont ré-encodées dans une vidéo reprenant les mêmes dimensions. Malgré les performances rapportées dans la littérature [Martin et al. \(2020a\)](#), le corpus BIRDS présente des difficultés du fait que les visages dans les vidéos égocentrées sont souvent prises sous des angles de vue variés et que la prise n'est pas frontale. Ainsi une étape de vérification "manuelle" avec un traitement manuel des "ratés" s'est avéré nécessaire.

Par ailleurs, d'autres éléments dans les images qui auraient permis l'identification des sujets devaient être floutés. Ainsi, nous avons établi le protocole suivant :

- Floutage automatique des visages avec les outils d'OpenCV ;
- Vérification de floutage et identification d'autres éléments à masquer ;
- Floutage manuel des visages non-masqués et des éléments sensibles.

Une illustration des résultats de cette procédure est présentée sur la figure 5.6. Sur l'image de gauche, c'est un visage qui est flouté, sur l'image de droite les éléments sensibles identifiés par l'annotateur sont floutés.

5.4.3 Les données imparfaites

Les données Lifelog, les données de santé et plus spécialement les données Healthcare, sont des données souffrant de bruit lié au mouvement ou aux capteurs mêmes, mais aussi des données incomplètes. Nous parlons ici des données incomplètes pour parler de données manquantes.

La gestion du bruit Comme nous l'avons mentionné précédemment, la qualité des résultats d'algorithmes dépendant de données est principalement liée à la qualité des données. La plupart des algorithmes considèrent qu'il n'y a pas de perturbations dans les données.

Néanmoins, il est vrai que les données collectées par les capteurs des objets connectés sont rarement entièrement propres et qu'elles souffrent régulièrement des perturbations.

Bizzego et al. (2020) montrent que la qualité des signaux provenant des dispositifs IoT est fiable et parfois même comparable à celle provenant de données de laboratoires.

Il a aussi été montré que les réseaux de neurones profonds sont suffisamment robustes pour pallier les bruits présents sur les données Hu et al. (2019).

Imputation des données Par définition, une valeur manquante est un attribut qui n'a pas été échantillonné dans l'ensemble de données, ou qui n'a jamais été enregistré, pour quelle que raison que ce soit. La précision des applications de détection usant d'algorithmes d'apprentissage par les données dépend fortement d'un ensemble de données supposé complet.

Cependant, pour diverses raisons telles que des problèmes matériels, des mesures incorrectes, des limitations dans le processus d'acquisition des données ou un échantillonnage différent, les données manquantes sont un incident courant.

La présence de valeurs manquantes complique la conduite de l'analyse des données. Il est donc important de bien gérer le problème des données manquantes.

Pour pallier ce problème des données manquantes, nombreuses solutions existent :

- L'application du maximum de vraisemblance où les paramètres d'un modèle pour les données complètes sont estimés, tandis que les fonctions de probabilité sont échantillonnées pour imputer la valeur manquante ;
- Une des méthodes les plus utilisées, consiste à supprimer les échantillons qui contiennent les données manquantes. La méthode présente l'avantage d'être simple mais n'est efficace que lorsque la quantité de données manquantes est faible ;
- Imputation des valeurs manquantes par les compléter par des valeurs estimées. Dans la plupart des cas, les attributs d'un ensemble de données ne sont pas indépendants les uns des autres. Par conséquent, les valeurs manquantes peuvent être déterminées par l'identification des relations entre les valeurs.

Pour notre dataset, les valeurs manquantes sont imputées, afin qu’elles puissent être utilisées à des fins d’apprentissage des modèles.

Données du signal du capteur : Le pré-traitement des données se compose de deux étapes : l’imputation et la normalisation des données. Nous utilisons une imputation classique des données par la valeur moyenne, calculée sur les données brutes disponibles de chaque capteur sur l’ensemble des données. Cette méthode présente l’avantage de conserver la même moyenne et la même taille d’échantillon. Les données enregistrées ont des échelles différentes. Ainsi pour les données des capteurs nous avons introduit l’étape de normalisation.

Dans ce qui suit, nous désignons par x_i le vecteur de dimension S dont les entrées sont les valeurs observées des S capteurs au temps i et par x_{mean} le vecteur des moyennes de toutes les *available* valeurs des capteurs. Nous désignons également par x l’ensemble des variables $\{x_1, x_2, \dots, x_T\}$. Il s’agit de notre ensemble de mesures dans l’ensemble du corpus. En supposant l’indépendance des coordonnées de x la matrice de covariance Σ de x sera diagonale. La normalisation que nous proposons consiste à blanchir les données, voir l’équation 5.1 sous une forme vectorielle.

$$\zeta(x_i) = \frac{x_i - x_{mean}}{\sqrt{\Sigma(x)}}. \quad (5.1)$$

Ensuite, nous mettons à l’échelle de façon linéaire l’ensemble des mesures pour qu’elles correspondent à l’intervalle $[0, 1]$ pour chaque coordonnée de x_i ,

$$x_{norm_i} = \frac{\zeta(x_i) - \min(\zeta(x_i))}{\max(\zeta(x_i)) - \min(\zeta(x_i))}. \quad (5.2)$$

Données vidéo : En ce qui concerne les données vidéos, le remplacement des images manquantes du flux video par des valeurs de substitution risque d’engendrer du biais à l’entraînement des réseaux profonds. Ainsi nous les avons gardés bruts. Nous avons observé les mêmes phénomènes de disparition des images dans le corpus NTCIR (cf. chapitre 4). Dans ce corpus, nous avons identifié les situations quand les plages d’enregistrement ne contenaient aucune image prise par l’appareil photo porté. Dans le cas du corpus BIRDS, il s’agit de l’enregistrement vidéo enregistré toutes les 3 secondes pendant chaque période de 10 secondes, ce choix étant fait pour limiter le volume de stockage des vidéos sur le concentrateur - téléphone portable. Ainsi avec la fréquence d’acquisition à 25 fps la probabilité du manquement total des images est faible.

Synchronisation des données au sens du rééchantillonnage : La synchronisation des données est introduite pour faciliter l’architecture de réseau de bout en bout ainsi que pour relever le défi de la fréquence d’échantillonnage variable entre le dispositif d’enregistrement vidéo et les dispositifs IoT, comme nous le montrons dans le chapitre 3. Les données visuelles sont échantillonnées par rapport aux données multi-variables des capteurs en utilisant une clé, c’est-à-dire que pour ces données,

il s'agit du timestamp. Ainsi, les timestamps enregistrés par les capteurs pendant l'enregistrement correspondent aux vidéos et les images correspondantes sont générées pour le timestamp particulier. Ainsi, pour une vidéo particulière, les données correspondantes du capteur sont établies. La synchronisation se fait donc au niveau de l'image.

5.4.4 Équilibrage des données

Le fait d'avoir des classes déséquilibrées dans les ensembles de données de classification est un problème courant mais néanmoins important. Cela est dû à une distribution asymétrique des données entre les classes. Les échantillons d'une classe dépassent largement les échantillons de l'autre classe. Malheureusement, les classes dites critiques qu'on cherche à prévenir ou à détecter sont souvent les classes minoritaires.

En outre, il y a généralement une association entre le problème des classes déséquilibrées et les problèmes de classifications biaisées et lorsque l'on a plusieurs classes minoritaires, le problème devient de plus en plus difficile. En conséquence, la plupart des techniques standard supposent une distribution équilibrée des classes. Par conséquent, ces algorithmes ne parviennent pas à représenter avec précision les caractéristiques de distribution des données lorsqu'ils sont présentés avec des ensembles de données déséquilibrées. Afin de surmonter le déséquilibre des données, deux approches peuvent être appliquées au niveau des données. On parle alors des techniques de rééchantillonnage pour obtenir une distribution équilibrée. Elles ont été appliquées dans la plupart des études. Le sous-échantillonnage de la classe majoritaire, le sur-échantillonnage de la classe minoritaire ou les deux, peuvent être utilisés en même temps. La technique de suréchantillonnage synthétique des minorités (SMOTE) est la méthode la plus courante au niveau des données [Chawla et al. \(2002\)](#).

Dans l'ensemble des enregistrements BIRDS, les situations à risques ne représentent qu'un ensemble très restreint par rapport à l'ensemble des données collectées, ce qui signifie que le nombre de situations à risques est très faible par rapport aux situations "*sans risques*", conformément aux scénarios de la vie réelle.

Ainsi, nous optons pour un rééchantillonnage afin d'obtenir un dataset plus équilibré et ce, par la suppression de trames de données des situations "*sans risques*" représentant la classe majoritaire. Après le rejet d'une partie des données de cette classe. Pour l'équilibrer, nous n'avons retenu qu'une proportion de 0,85 % d'échantillons "*sans risques*" de l'ensemble du jeu de données. Nous avons adopté une approche de fenêtre glissante lors d'une étape de pré-traitement pour la création des instances. Une fenêtre de longueur fixe de 20 échantillons a été considérée avec un pas de glissement de 4 échantillons. Notez que les intervalles de temps entre les différents échantillons ne sont pas stables, en raison de l'imprécision des capteurs.

La fenêtre fixe de 20 échantillons peut donc avoir une durée différente dans le temps (de l'ordre de quelques millisecondes). Pour attribuer une étiquette unique à une fenêtre, nous avons utilisé le mode maximal de l'histogramme de l'étiquette de la vérité terrain calculée sur celle-ci.

5.5 Conclusion

Dans ce chapitre, nous avons présenté le dataset BIRDS enregistré par le dispositif développé et décrit dans le chapitre 3.

Les données du corpus BIRDS sont principalement mesurées par des objets IoT avec des WS. Les données se caractérisent par leur volumétrie, la temporalité, les différentes caractéristiques et variables d'études, etc. Elles présentent aussi certains problèmes, au rang desquels figurent les fréquences d'échantillonnage différentes, le problème des données manquantes et le déséquilibre des classes enregistrées notamment la classe dite *Sans risque* fortement représentée.

Nous avons essayé d'apporter des solutions par les pré-traitements réalisés afin de pouvoir les exploiter dans le prochain chapitre pour l'analyse de ces situations à risques et leur classification.

Chapitre 6

Analyse des données BIRDS

Ce chapitre décrit un système de détection de situations à risques basé sur la combinaison des données multicapteurs et de modèles prédictifs pour une détection multi-classes. Le dispositif de détection qui est un système multicapteurs destiné à être porté par un sujet âgé fragile. Il a servi à enregistrer des données multimodales liées aux activités et à l'état physiologique du sujet. Les données captées sont pré-traitées comme décrit dans le chapitre 5 avant une analyse avec des réseaux de neurones récurrents pour la détection des situations à risques décrites dans notre taxonomie (voir chapitre 3). Les méthodes de traitement utilisées sur le dataset BIRDS sont présentées dans ce chapitre, nous rendons compte des résultats obtenus et nous mettons en lumière les perspectives envisagées.

6.1 Introduction

Le but ultime des systèmes IoT pour le suivi des sujets fragiles consiste en une prise de décision en temps réel, pour la détection et la prévention des risques. Ces systèmes doivent pouvoir répondre à un large éventail de situations à risques liées à la fois au contexte écologique, c'est-à-dire l'environnement domestique non contraint, et aux mouvements indépendants de la personne âgée/fragile.

L'objectif des nouveaux dispositifs est d'offrir une certaine forme d'indépendance aux personnes âgées grâce à des systèmes qui les suivent quotidiennement et qui ont une réponse rapide à l'aide d'algorithmes efficaces. Aujourd'hui, un progrès rapide des dispositifs IoT rend possible des réponses rapides à des situations quotidiennes complexes. La tendance générale pour une telle détection consiste à développer des approches d'apprentissage automatique supervisé [Boll et al. \(2019\)](#) s'alliant aux données collectées par ces objets IoT.

Ces données multiples enregistrées par des capteurs hétérogènes ont une composante commune à toutes, la temporalité. Ce sont donc des séries temporelles, avec une variable commune t , ce qui nous permet d'envisager une fusion de ces données. En effet, décrivant les mêmes scènes de vie dans un même intervalle de temps, elles peuvent être complémentaires. Cela nous permet donc d'améliorer les taux de détection.

Le problème de détection des situations à risques dans les données des capteurs du corpus BIRDS s'apparente à la détection des anomalies dans des séries temporelles relevant de contextes différents. Une vaste littérature existe à ce sujet.

[Chalapathy et Chawla \(2019\)](#) présente une vue d'ensemble structurée et complète des méthodes de recherche en matière de détection d'anomalies basées sur l'apprentissage profond.

[Habeeb et al. \(2019\)](#) fait état des dernières méthodes de traitement des données en temps réel liées à la détection des anomalies et les caractéristiques essentielles des algorithmes d'apprentissage automatique associés. Cette étude a présenté un diagramme en couches qui aide à comprendre le processus à travers les méthodes proposées dans l'état de l'art pour la détection des anomalies, ce qui donne un cadre pour le traitement des big data en temps réel afin de prédire l'avènement de ces anomalies à l'avenir.

Les modèles de détection d'anomalies basés sur l'apprentissage profond sont divisés en fonction des types de réseaux de neurones et des données utilisées, impliquant des réseaux neuronaux non supervisés, semi-supervisés et hybrides. L'idée des modèles hybrides profonds est d'utiliser des réseaux neuronaux profonds, principalement des auto-encodeurs, comme extracteurs de caractéristiques. Après l'apprentissage dans les représentations cachées des auto-encodeurs, ces caractéristiques sont transmises aux algorithmes traditionnels de détection d'anomalies tels que les SVM ou encore les réseaux neuronaux profonds. Ce type de modèles d'apprentissage profond a été appliqué à plusieurs situations de détection d'anomalies et a obtenu des scores de prédiction assez élevés.

[Nguyen et al. \(2021\)](#) propose deux approches basées sur les données qui permettent de prendre de meilleures décisions dans la gestion de la chaîne d'approvisionnement dans un contexte industriel. Ils suggèrent une méthode basée sur un réseau de mémoire à long terme (LSTM) pour la prédiction de séries chronologiques multivariées et une méthode basée sur un réseau LSTM auto-encodeur combiné à un algorithme SVM à une classe pour la détection d'anomalies dans un contexte de ventes. D'après la littérature sur la détection d'anomalies à travers des ensembles de données multivariées, il apparaît clairement que des architectures complexes alliant les réseaux auto-encodeurs avec parfois des réseaux hybrides donnent de meilleurs scores de détection et de prédiction.

Les réseaux neuronaux récurrents (RNN), tels que le réseau LSTM [Hochreiter et Schmidhuber \(1997\)](#) et le réseau [Chung et al. \(2014a\)](#), ont montré qu'ils obtenaient des résultats de pointe dans de nombreuses applications avec des séries temporelles ou des données séquentielles, notamment la traduction automatique et la reconnaissance vocale. Les RNN bénéficient de plusieurs propriétés intéressantes, telles que de bonnes performances de prédiction, ainsi que la capacité de capturer des dépendances temporelles à long terme et des observations de longueur variable.

Dans ce chapitre, nous proposons la solution retenue pour le traitement des séries temporelles pour la reconnaissance des situations à risques des personnes fragiles. Après une revue de la littérature, nous avons proposé la solution à base de réseaux LSTM et d'auto-encodeurs. Nous évoquerons aussi les mécanismes d'attention, pour une classification multi-classes à entrées multi-variables, lors de l'entraînement des DNN, plus précisément des auto-encodeurs LSTM. Après cette introduction, ce chapitre est organisé comme suit. Dans la section 6.2 nous parlerons des réseaux neuronaux profonds pour l'analyse des séries temporelles. Ensuite, dans la section 6.3, nous exposerons les réseaux de neurones auto-encodeurs et nous aborderons les mécanismes d'attention appliqués à ces derniers. Dans la section 6.4, nous présentons les expérimentations mises en place avec les réseaux de neurones développés à cet effet. Ainsi, nous exposons les résultats obtenus, et mettons en perspective les analyses faites sur ces derniers. Enfin, dans la section 6.5, nous concluons ce chapitre.

6.2 Méthodes de traitement des données

Dans cette section, nous passons brièvement en revue certains algorithmes d'IA nécessaires à la construction des modèles proposés pour la prédiction et la détection des anomalies, notamment le réseau LSTM, le réseau auto-encodeur et l'intégration de l'attention dans le réseau auto-encodeur.

6.2.1 Les séries temporelles multi-entrées à classifications multiples avec les réseaux de neurones récurrents

Les réseaux neuronaux LSTM sont largement utilisés pour la classification des séries temporelles. C'est un type de réseaux de neurones récurrents palliant le problème de l'explosion du gradient. Ils permettent d'établir des liens à très long terme pendant l'apprentissage [Hochreiter et Schmidhuber \(1997\)](#), c'est-à-dire de conserver les dépendances à long terme entre les données à un moment donné et celles de nombreux pas de temps antérieurs. Le réseau LSTM se présente sous la forme d'une chaîne de modules répétés de réseaux neuronaux, où chaque module comprend trois portes de contrôle, à savoir la porte d'oubli, la porte d'entrée et la porte de sortie. La particularité des LSTMs consiste dans la manière dont l'état caché h est traité : dans le cas des RNNs simples, le traitement de la récurrence, est assuré par une fonction \tanh . Pour les LSTMs, ce traitement est remplacé par une "cellule mémoire". Chaque porte est composée d'une couche de réseau neuronal sigmoïde et d'une opération de multiplication ponctuelle représentant une partie de l'information d'entrée que l'on doit laisser passer. Comme l'utilisation d'un RNN pour les données de séries temporelles, le LSTM lit une séquence de vecteurs d'entrée $x = x_1, x_2, \dots, x_t, \dots$, où $x_t \in \mathbb{R}^m$ représente un vecteur m -dimensionnel de lectures pour m variables à l'instant t . Ils retiennent les informations pertinentes des entrées

précédentes et utilisent ces informations pour modifier la sortie actuelle. Nous avons appliqué une architecture LSTM standard dans un premier temps pour la détection de risques sur des signaux de capteurs du dataset Birds. Rappelons que les couches sigmoïdes produisent des nombres dans l'intervalle $[0, 1]$.

Architecture LSTM

L'unité LSTM de base est composée de cellules avec une porte d'entrée, une porte de sortie et une porte d'oubli. La cellule LSTM est caractérisée par un noeud central, contenant l'état interne (ou mémoire) de la cellule, et un certain nombre de "portes" des trois catégories mentionnées. Ces portes sont utilisées pour gérer le stockage séquentiel des informations (portes d'entrée et d'oubli) et le rôle de l'état interne dans la production de chaque sortie (porte de sortie). En fermant la porte d'entrée, par exemple, les nouveaux événements sont moins pris en compte. Les LSTM utilisent le concept de gating pour surmonter le problème du gradient qui disparaît ou explose : [Greff et al. \(2016\)](#).

Les principales équations expliquant le fonctionnement d'une cellule LSTM sont données dans ce qui suit : x_t sont les données courantes, l'état caché h_{t-1} , w_x sont les poids associés à chaque multiplication dans la porte x , b_x sont les biais associés à chaque porte x , σ la fonction Sigmoidale logistique, et \tanh sont les options pour la fonction d'activation définie par :

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \quad (6.1)$$

$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f) \quad (6.2)$$

$$\tilde{c}_t = \tanh(w_c[h_{t-1}, x_t] + b_c). \quad (6.3)$$

$$c_t = f_t \circ c_t + i_t \circ \tilde{c}_t \quad (6.4)$$

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o). \quad (6.5)$$

$$h_t = o_t \circ \tanh(c_t) \quad (6.6)$$

Ici \circ désigne la multiplication, f_t la porte d'oubli de la cellule LSTM, c_t son vecteur d'états, o_t sa porte de sortie, h_t son vecteur d'états cachés, \tilde{c}_t la cellule d'entrée d'activation du vecteur. La figure 6.1 présente l'architecture d'une cellule LSTM.

Comme les données sont des séries temporelles séquentielles, nous nous concentrons principalement sur les variantes de *RNN*. Nous utiliserons dans nos expérimentations le réseau LSTM qui convient bien aux traitements de données multidimensionnelles. En outre, tous les capteurs n'ont pas la même importance pour la tâche de classification multi-classes. Par conséquent, au lieu d'utiliser des données brutes comme entrée du réseau de neurones LSTM, nous les encodons d'abord. Comme entrée de l'encodeur, nous prenons tous les échantillons normalisés des données x_{norm_i}

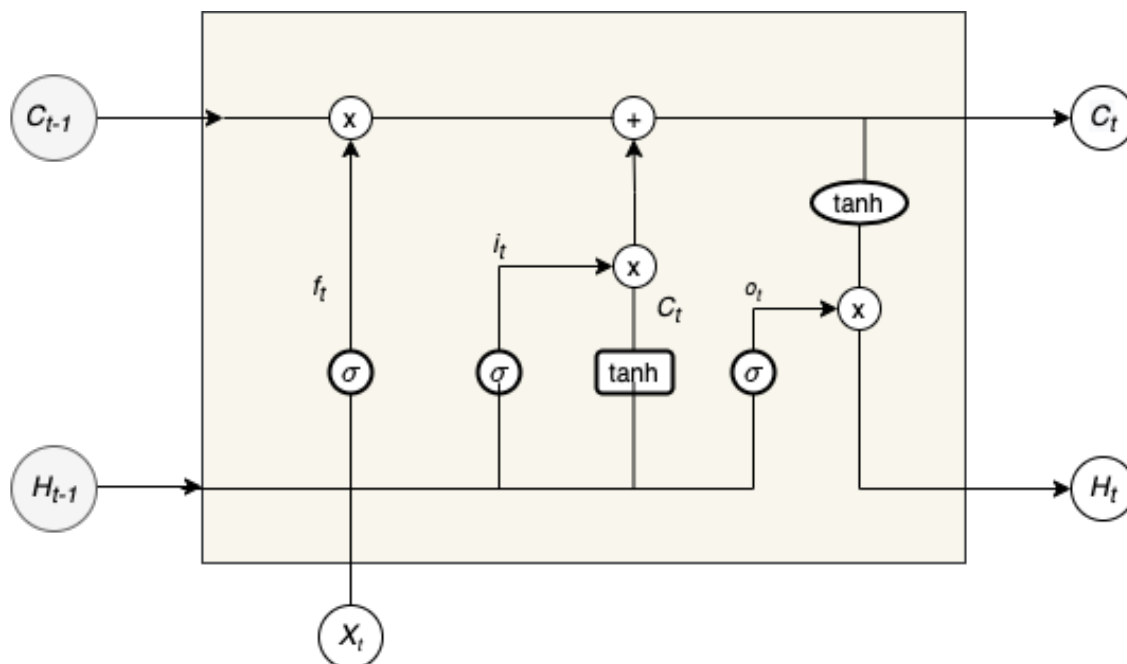


FIGURE 6.1 – Architecture d'une cellule de mémoire à long et court terme (LSTM).

dans une fenêtre temporelle de longueur N . Ainsi, notre échantillon d'entrée au temps t est $x_t \in \mathbb{R}^{S \times N}$.

6.2.2 L'auto-encodeur LSTM avec ajout de l'attention

Les auto-encodeurs L'auto-encodeur est un réseau neuronal auto-supervisé qui vise à apprendre le meilleur schéma d'encodage-décodage à partir de ses données. La structure générale d'un auto-encodeur se compose d'une couche d'entrée, d'une couche de sortie. Un réseau neuronal encodeur, un réseau neuronal décodeur et un espace latent. Les données présentées en entrée du réseau y sont introduites par la couche d'entrée, l'encodeur les compresse dans l'espace latent, tandis que le décodeur décompresse la représentation codée dans la couche de sortie. La sortie codée-décodée est ensuite comparée aux données initiales et l'erreur est rétropropagée à travers le réseau présenté pour mettre à jour les poids du réseau à travers les algorithmes de rétropropagation. Étant donné l'entrée $X \in \mathbb{R}^m$, le codeur compresse X composé de m entrées (x_1, x_2, \dots, x_m) pour obtenir une représentation codée de l'entrée. Le décodeur reconstruit cette représentation pour retrouver une estimation de l'entrée en sortie et donc la sortie du réseau. L'auto-encodeur est entraîné en minimisant l'erreur de reconstruction

Le principe de l'auto-encodeur est qu'à travers l'estimation de la sortie du réseau par son entrée, on contraint l'espace latent à avoir une dimension plus petite que l'entrée, c'est à dire compresser l'entrée pour passer d'une dimension m à une

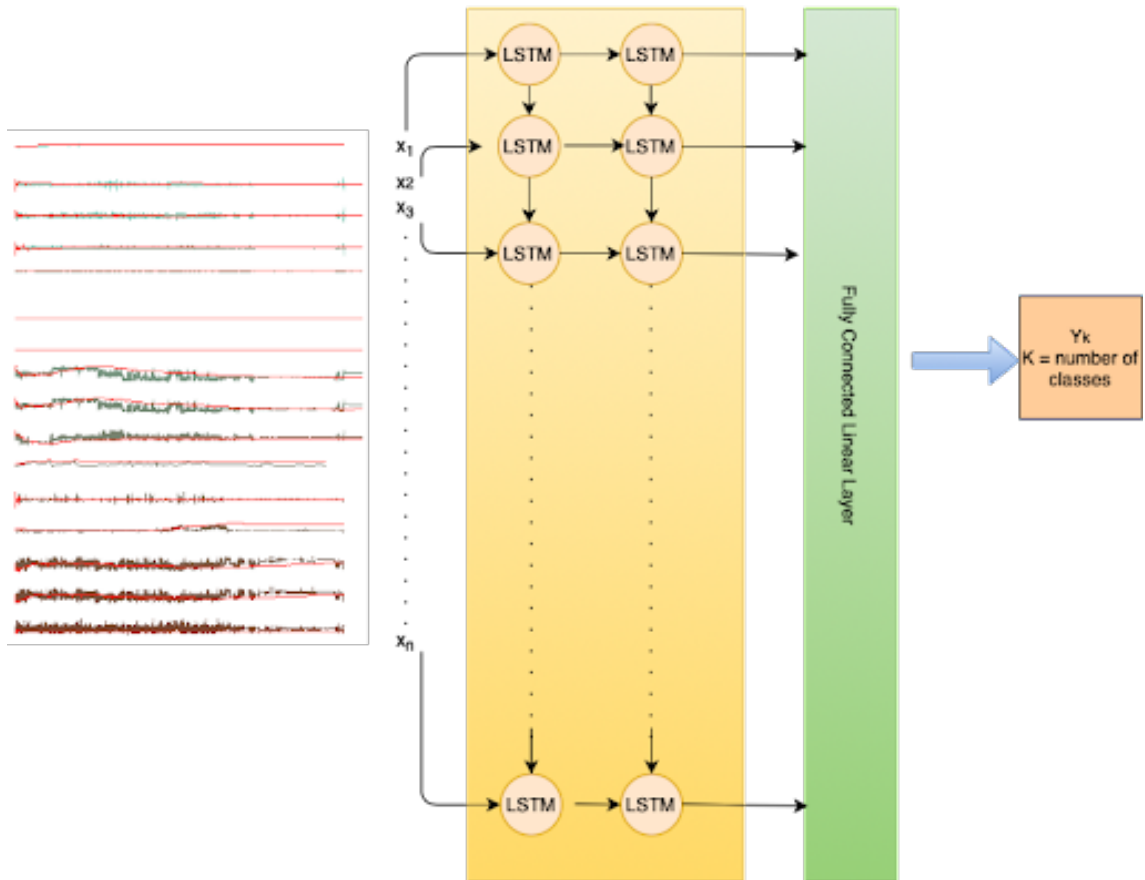


FIGURE 6.2 – Architecture du réseau de neurones LSTM entrées multiples et sorties multiples.

dimension n , sachant $m > n$. L'auto-encodeur est donc forcé à apprendre les caractéristiques les plus saillantes des données d'entraînement. En d'autres termes, une caractéristique importante dans la conception d'un auto-encodeur est qu'il réduit les dimensions des données tout en conservant les informations majeures de la structure des données.

Les auto-encodeurs LSTM Plusieurs types d'auto-encodeurs ont été proposés dans la littérature, tels que l'auto-encodeur vanille, l'auto-encodeur convolutif, l'auto-encodeur régularisé et l'auto-encodeur LSTM. Parmi ces types, l'auto-encodeur LSTM fait référence à l'auto-encodeur dont l'encodeur et le décodeur sont tous deux des réseaux LSTM.

La capacité des LSTM à apprendre des modèles à travers les données sur de longues séquences les rend appropriés pour la prédiction des séries temporelles ou à savoir reconnaître une anomalie survenue sur une séquence d'une série temporelle. L'avantage de l'utilisation de la cellule LSTM est de capturer les dépendances temporelles dans les données multivariées. L'utilisation d'un auto-encodeur LSTM pour la détection d'anomalies sur des données de séries temporelles multivariées est visible dans plusieurs études. Longari et al. (2020), propose une méthode pour un réseau de contrôleurs pour un système de détection d'intrusion basé sur des auto-encodeurs à mémoire à long terme. Pendant une phase d'apprentissage, le système analyse automatiquement les flux de données et construit un modèle des séquences de données légitimes. Ensuite, il détecte les anomalies en calculant la différence entre les séquences reconstruites et les séquences réelles respectives. Il a été montré que l'ajout des auto-encodeurs améliorerait le taux de précision par rapport à l'état de l'art.

L'attention dans les réseaux de neurones auto-encodeurs Le principe de base d'un RNN classique est d'analyser chaque élément d'une série d'entrées, l'un après l'autre, et de mettre à jour son vecteur "état caché" à chaque étape du processus.

Les auto-encodeurs estiment les sorties du réseau par les données injectées en entrée. Le système permet alors de compresser l'entrée pour ne retenir que les caractéristiques représentant au mieux l'entrée, et ce, afin de minimiser l'erreur entre l'entrée x et celle estimée en sortie \hat{x} . L'auto-encodeur est donc forcé à apprendre les caractéristiques les plus saillantes des données par soucis d'optimisation.

Une mise en œuvre naïve d'un encodeur-décodeur comme dans le modèle simple d'encodeur-décodeur exige que l'encodeur comprime l'entrée en un vecteur unique de dimensionnalité prédéfinie, quelle que soit la taille de l'entrée ou la quantité d'informations qu'elle contient. Même lorsque la taille de l'entrée est fixe, comme dans

le cas d'une image à résolution fixe, la quantité d'information contenue dans chaque image peut varier de manière significative. En outre, l'interprétabilité du codeur-décodeur simple est extrêmement faible. Comme toutes les informations nécessaires au décodeur pour générer la sortie sont comprimées dans un vecteur de contexte sans aucune structure présupposée.

Afin de sélectionner les données et des caractéristiques les plus importantes, les mécanismes d'attention ont été introduits. Une des premières méthodes a été proposée par [Bahdanau et al. \(2014\)](#) en 2014 pour la traduction automatique. L'hypothèse principale dans les DNNs de modélisation de séquence, tels que les GRUs, les LSTMs et les RNNs est que l'état actuel détient des informations pour l'entrée entière vue auparavant. Un mécanisme d'attention examine les états cachés pour l'ensemble de la séquence d'entrée et apprend à sélectionner les états qui sont les meilleurs pour faire une prédiction. La méthode proposée est composée de quatre étapes : i) codage, ii) calcul des poids d'attention, iii) création d'un vecteur de contexte et iv) décodage. L'idée centrale est d'apprendre à pondérer les états cachés. D'abord, l'attention est calculée avec une opération de softmax intermédiaire sur les états cachés. Ensuite, le vecteur de contexte est donné par la combinaison linéaire des poids cachés (états actuels) et des poids d'attention (opération softmax). Enfin, le décodage est calculé à partir de la sortie et des états précédents, ainsi que du vecteur de contexte.

Cette approche a inspiré les réseaux Transformers, proposés par [Vaswani et al. \(2017\)](#), qui n'impliquent aucun RNN, comme les GRU ou les LSTM. Un réseau Transformer est composé d'une pile d'encodeurs et de décodeurs fonctionnant en parallèle, chacun d'entre eux étant constitué de plusieurs blocs d'"attention multi-têtes" et de couches directes.

Les transformers sont dans le focus de recherche pour l'analyse et prédiction des séries temporelles multivariées [Zerveas et al. \(2021\)](#) néanmoins les réseaux RNN munis des blocs d'attention restent un outil performant pour la classification des séries temporelles. Telle est notre solution que nous exposons dans la suite de ce chapitre.

6.3 Les réseaux de neurones auto-encodeurs avec attention

Dans cette section, nous présentons le réseau LSTM auto-encodeur avec une couche d'attention. La partie encodeur a comme entrée les données de séries temporelles multimodales. Le réseau encodeur fait correspondre les données normalisées des capteurs à une séquence de caractéristiques. Dans la partie décodeur une couche d'attention est introduite. Nous avons considéré deux types d'attention : "molle" et "dure". La figure 6.3 présente l'architecture du réseau de neurones LSTM entrées

multiples et sorties multiples intégrant un auto-encodeur avec un bloc d'attention (attention dure ou attention molle).

6.3.1 Le LSTM encodeur

Les équations décrivant l'encodeur LSTM de notre réseau reprennent les équations décrivant les cellules LSTM de base 6.1, 6.2, 6.3, 6.4, 6.5, 6.6 du réseau LSTM de base décrites dans la section 6.2.

Dans les équations (6.1), (6.2), (6.5), (6.6), (6.4), i_t, f_t, o_t, h_t, c_t font référence respectivement à la porte d'entrée, la porte d'oubli, la porte de sortie, l'état caché et l'état de la cellule. σ désigne la fonction sigmoïde, h_{t-1} sortie du dernier bloc LSTM précédent, (w_x) poids des neurones respectifs de la porte (x) et (b_x) biais respectifs à chaque porte (x) .

Notre réseau de codage est constitué de deux couches, comme le montre la partie supérieure de la figure 6.3. La sortie multidimensionnelle est $h_t \in R^{M \times N}$ avec M le nombre de caractéristiques sur l'état caché dans le bloc LSTM. La sortie h_t de l'encodeur basé sur des réseaux LSTM est ensuite transmise au décodeur avec le bloc d'attention intégré. Par souci de clarté, nous appellerons la sortie de l'encodeur o_t dans la figure elle est dénotée par "sortie encodée". Nous présentons deux types de décodeurs : i) le décodeur avec attention molle, ii) le décodeur avec attention dure.

6.3.2 Le décodeur LSTM

Pour pousser la sortie o_t vers le décodeur, nous commençons par initialiser le LSTM du décodeur avec le même état initial h_0 que pour l'initialisation de l'encodeur. Ensuite, les sorties o_t sont passées au décodeur, dont la première couche est le bloc d'attention, qui est une couche linéaire, voir équation 6.7.

$\hat{\mathbf{g}}_t$ est une représentation dynamique de la partie pertinente de l'entrée des données à l'instant t .

$$a_t = e_t \odot W_t, \tag{6.7}$$

où \odot signifie produit scalaire.

La sortie de la couche est passée au softmax pour la normalisation comme dans l'équation 6.8

$$\tilde{a}_t = \text{softmax}(a_t) \tag{6.8}$$

Enfin, les poids d'attention ainsi obtenus vont multiplier la sortie de l'encodeur comme le montre l'équation 6.9 :

$$g_t = \tilde{a}_t \otimes o_t, \tag{6.9}$$

où \otimes est une multiplication matricielle. Ainsi définie l'attention dans le réseau est molle, dans le sens où la sortie du bloc d'attention représente une somme pondérée des sorties de l'encodeur.

Ensuite, le résultat g_t sera soumis au classificateur - décodeur LSTM, voir figure 6.3, partie inférieure. Nous avons également conçu ce classificateur LSTM comme un réseau à deux couches car l'ajout de la deuxième couche améliore les performances selon nos expériences préliminaires.

Avant de transmettre les données au réseau, nous utilisons une approche de fenêtre glissante sur les données. Pour cette approche, nous définissons une longueur de séquence qui est la longueur de la fenêtre glissante.

6.3.3 Les blocs d'attention

Le mécanisme d'attention contrôle l'entrée réellement vue par le décodeur et nécessite un autre réseau neuronal, auquel nous nous référons en tant que modèle d'attention. La tâche principale du modèle d'attention est de noter chaque vecteur de contexte par rapport à l'état caché actuel h_{t-1} du décodeur. L'attention molle et l'attention dure ont été présentées par Xu Show, Xu et al. (2015). Dans le papier Xu et al. (2015), il présente deux mécanismes d'attention appliqués à la génération de légende d'images. Un des deux mécanismes est un mécanisme d'attention déterministe dit "soft" ou "attention molle" pouvant être entraîné par des méthodes de rétropropagation standard et un mécanisme d'attention stochastique dit "dur" pouvant être entraîné en maximisant une limite inférieure variationnelle approximative ou de manière équivalente.

L'attention déterministe "molle"

L'attention déterministe place les poids d'attention sur tous les patches d'entrée à savoir des séquences de données (données multivariés, image, des phrases, etc.), c'est-à-dire qu'elle utilise le mécanisme de moyenne pondérée. Il mesure l'attention concernant les différents morceaux de l'entrée, et produit les caractéristiques pondérées de l'entrée. Il exclut les zones qui ne sont pas pertinentes pour la tâche à accomplir en leur attribuant un poids faible. De cette façon, l'attention "molle" ne se limite pas à des parties spécifiques de l'image ou de la phrase ; au contraire, elle apprend en permanence.

L'attention "molle" est un mécanisme d'attention entièrement différentiable, où les gradients peuvent être propagés automatiquement pendant la rétropropagation. L'attention molle se présente sous l'équation 6.10 :

$$\mathbb{E}_{p(s_t|a)}[\hat{\mathbf{g}}_t] = \sum_{i=1}^L \alpha_{t,i} \mathbf{a}_i \quad (6.10)$$

\mathbf{a} représente les états cachés de l'encodeur/entrée.
 α représente les scores d'attention.

$s_{t,i}$ est une variable **à un coup** avec "1" si le "i-ième" emplacement doit être sélectionné.

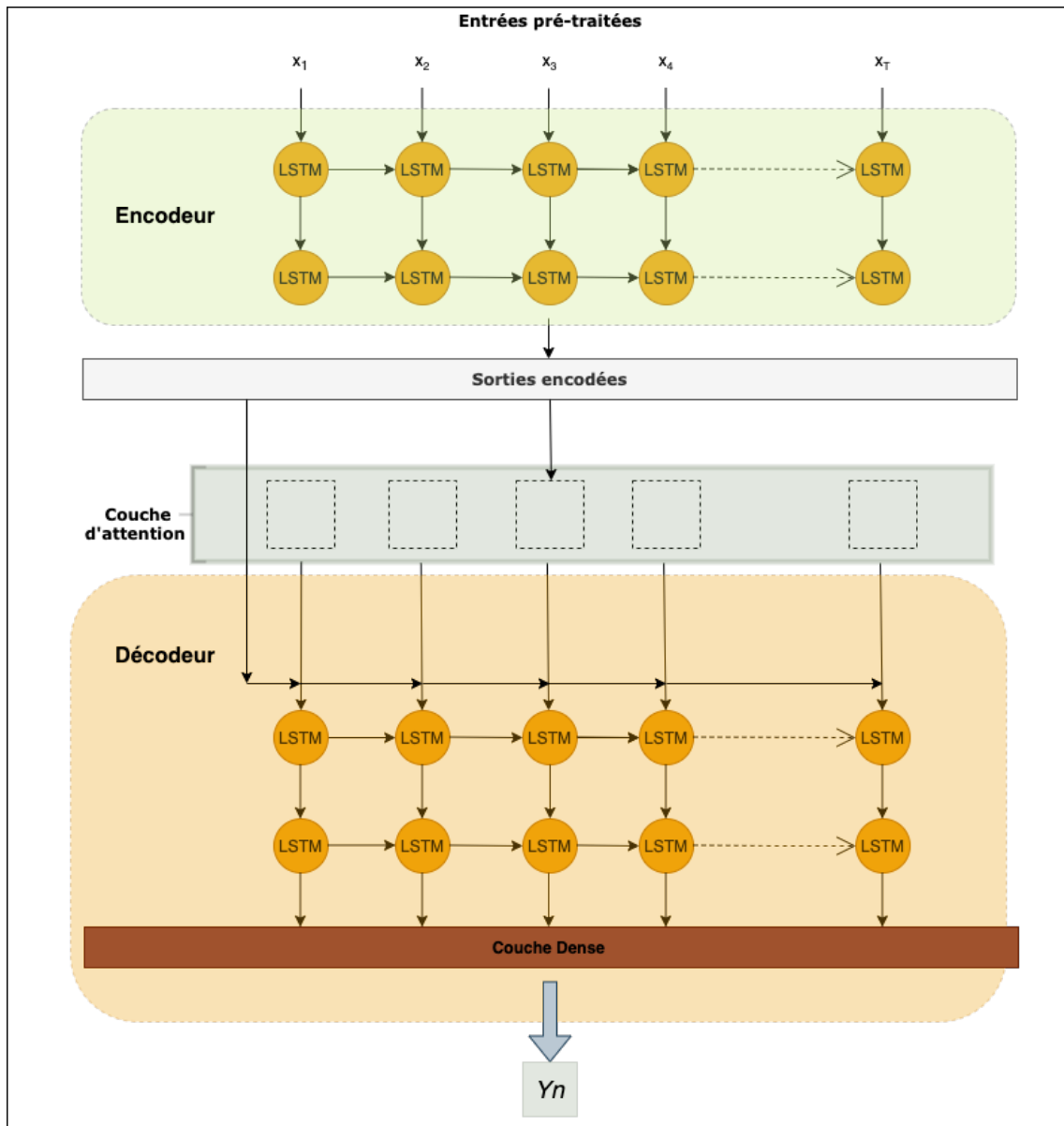


FIGURE 6.3 – Architecture du réseau de neurones LSTM entrées multiples et sorties multiples intégrant un auto-encodeur avec un bloc d'attention.

L'attention stochastique "dure"

L'attention "dure", comme son nom l'indique, se concentre uniquement sur une partie spécifique du patch de données ou de l'image/de la phrase. Le score d'attention est utilisé comme la probabilité que le i -ème emplacement soit sélectionné. Nous pourrions utiliser un simple argmax pour faire la sélection, mais il n'est pas différentiable et des techniques complexes sont donc employées. Pendant la rétropropagation, pour estimer les gradients pour tous les autres états, nous devons effectuer un échantillonnage et faire la moyenne des résultats en utilisant la méthode de Monte Carlo. L'attention dure a été introduite par les équations 6.11, 6.12, 6.13 comme suit :

$$\hat{\mathbf{g}}_t = \sum_i s_{t,i} \mathbf{a}_i \quad (6.11)$$

$$p(s_{t,i} = 1 | s_{j < t}, \mathbf{a}) = \alpha_{t,i} \quad (6.12)$$

$$\tilde{s}_t^n \sim \text{Multinoulli}_L(\{\alpha_i^n\}) \quad (6.13)$$

Multinoulli est une loi de probabilité également appelée distribution catégorielle, car elle peut être utilisée pour modéliser des événements ayant K issues possibles. La distribution de Bernoulli peut être considérée comme un cas spécifique de Multinoulli, où le nombre d'issues possibles K est de 2.

\mathbf{a} représente les états cachés de l'encodeur/entrée.

α représente les scores d'attention.

$s_{t,i}$ est une variable **à un coup** avec "1" si le "i-ème" emplacement doit être sélectionné.

6.4 Expérimentations et résultats

Ci-après, nous présentons les détails de réglage des réseaux profonds que nous avons utilisés, leur entraînement et les résultats d'expérimentation.

6.4.1 Réseaux de neurones utilisés

Le réseau LSTM Afin d'analyser les séries temporelles du dataset BIRDS, nous avons opté pour les réseaux de neurones récurrents et plus spécialement les LSTM Hochreiter et Schmidhuber (1997) dans un premier temps cette architecture nous a servi de baseline. Comme, nous le montrons dans le 6.4.2, la partie du dataset BIRDS incluant les données 1D, a une longueur de 16 correspondant aux 16 capteurs utilisés dans le dispositif BIRDS en plus de la caméra, la largeur de ces données correspond à N (le nombre de timestamps présentant des données). La longueur de la séquence d'entrée que nous utilisons est de 20 avec un échantillonnage aléatoire de l'ensemble de données du dataset sélectionné. La longueur de la taille de l'entrée du réseau LSTM est donc de 16 par rapport et deux couches cachées de taille 128 suivie d'une

couche dense. Le nombre de classes de sortie est de 4 et correspond aux situations à risques que nous choisissons de reconnaître.

Le réseau auto-encodeur LSTM L'auto-encodeur LSTM utilisé inclut une partie encodeur et une partie décodeur. L'encodeur présente une couche d'entrée d'une taille de 16 correspondant aux 16 entrées du dataset qui est de taille $(16 \times N)$, suivie de deux couches cachées LSTM de taille 128. La sortie encodée est ensuite soit transmise au bloc d'attention dure ou au bloc d'attention molle si l'attention est intégrée. Sinon les données encodées sont directement transmises au décodeur composé aussi de deux couches LSTM cachées et enfin une couche dense. Le nombre de classes en sortie est de 4 et correspond aux situations à risques que nous choisissons de reconnaître.

L'intégration des blocs d'attention Tous les mécanismes d'attention décrits dans la section 6.2 : les blocs d'attention molle et d'attention dure sont intégrés sur deux expérimentations différentes entre l'encodeur et le décodeur dans notre modèle d'auto-encodeur LSTM mis en place. Différentes configurations sont adoptées pour l'analyse des données BIRDS. La figure 6.3 présente l'architecture de notre auto-encodeur pour la prédiction multi-classes avec des entrées multivariées. Sur cette figure, nous pouvons voir l'emplacement des blocs d'attention lorsqu'ils sont intégrés à l'architecture.

6.4.2 Dataset

Les données analysées dans nos expérimentations appartiennent à notre dataset BIRDS présenté dans le chapitre 5. Le jeu de données est enregistré à l'aide des dispositifs IoT qui envoient les données sur l'application BIRDS Android développée à cet effet (chapitre 3). Les enregistrements vidéo ont été utilisés pour l'inférence visuelle des annotations effectuées par les psychologues sur l'ensemble des mesures. Sans utiliser la vidéo, ces dispositifs nous fournissent un total de 16 mesures provenant des différents capteurs.

Dans le scénario d'enregistrement d'un système adaptable à chaque individu, nous avons utilisé 12 jours d'enregistrement effectué par un sujet sain (voir chapitre 5) simulant des situations à risques dans sa vie quotidienne.

L'une des difficultés est le nombre très limité de situations à risques parmi les situations *sans risque*, ce qui est conforme au scénario de la vie réelle. Ainsi, pour équilibrer les données, nous avons procédé à une réduction du nombre d'échantillons dans la classe "*Sans risque*". Le volume total contient des enregistrements de 2601 minutes sur une période de 12 jours. La durée des enregistrements varie d'une période d'environ 2,5 heures à 5,5 heures par jour. Quatre des situations à risques apparaissent sur cette période de 12 jours. Les données utilisées sont donc une partie du dataset BIRDS. Les situations à risques de l'ensemble de la taxonomie que

nous abordons dans ce travail sont : *le risque de chute*, *le risque d'accident domestique*, *le risque de déshydratation* qui est un risque qu'on estime en détectant une situation "opposée au risque" (en l'occurrence voir si la situation s'est bien produite *Hydratation par ingestion d'eau*). La classe de rejet représente les situations *Sans risque*. Le tableau 6.1 présente la distribution des échantillons par timestamps pour les classes de taxonomie de l'ensemble de données.

Situations à risques	Nombre d'échantillons	Pourcentage
<i>Sans risque</i>	17039653	98.31 %
<i>Risque de chute</i>	77800	0.45 %
<i>Risque d'accident domestique</i>	147500	0.85 %
<i>Hydratation par prise d'eau</i>	52799	0.30 %

Tableau 6.1 – Distribution des situations à risques dans le dataset sélectionné.

Les situations à risques de l'ensemble de la taxonomie que nous abordons dans ce travail sont : *Risque de chute*, *risque d'accident domestique* et une situation "opposée au risque", telle que *Hydratation par prise d'eau*. Une classe de rejet représente les situations *Aucun risque*. Le tableau 6.1 présente la distribution des échantillons d'horodatage pour les classes de taxonomie de l'ensemble de données.

Comme indiqué précédemment, les échantillons *Sans risque* constituent la majorité des données. Par conséquent, l'ensemble de données est fortement déséquilibré et ne peut pas servir à l'entraînement d'un DNN tel quel. Pour l'équilibrer, nous n'avons retenu que 0,85 % de la proportion d'échantillons *sans risque* de l'ensemble du jeu de données (voir sous-section 5.4.4). Nous avons adopté une approche de fenêtre glissante lors d'une étape de pré-traitement pour la création des instances. Une fenêtre de longueur fixe de 20 échantillons a été considérée avec un pas de glissement de 4 échantillons. Notez que les intervalles de temps entre les différents échantillons ne sont pas stables, en raison de des différentes fréquences d'acquisition des capteurs. La fenêtre fixe de 20 échantillons peut donc avoir une durée différente dans le temps (de l'ordre de quelques millisecondes).

6.4.3 Paramètres et détails des entraînements

Dans cette sous-section, nous présentons les hyper-paramètres optimaux sélectionnés pour chaque réseau de neurones présenté dans la section 6.2

LSTM baseline Dans le réseau LSTM, nous utilisons une séquence de longueur 20 avec un échantillonnage aléatoire de l'ensemble de données. La taille de l'entrée est de 16 et deux couches cachées de taille 128. Le nombre de classes de sortie est de 4 et correspond aux situations de risques que nous choisissons de reconnaître. Le

6. Analyse des données BIRDS

Tableau 6.2 – Tableau d’optimisation du taux d’apprentissage.

Taux d’apprentissage	10^{-3}	10^{-4}	10^{-5}
LSTM			
Précision de validation	63.4 %	58 %	43.3 %
Précision de test	46.92 %	56.54 %	45.37 %
Epoch	7	15	30

Tableau 6.3 – Tableau d’optimisation de la taille des lots.

Taux d’apprentissage	128	256	512
LSTM			
Précision de validation	63 %	58 %	58 %
Précision de test	54.3 %	56.54 %	53 %
Epoch	13	15	8

réseau a été entraîné sur 30 époques avec une taille de lot "batch size" de 256. Le taux d’apprentissage choisi est fixé à une valeur de 0.0001. Ces hyper paramètres ont été choisis après une expérimentation sur plusieurs valeurs (voir tableaux 6.2 et 6.3). L’algorithme d’optimisation choisi est Adam [Kingma et Ba \(2014\)](#) avec un taux d’apprentissage fixe de 0,0001. L’optimisation des hyper-paramètres s’est faite quivant la méthode de dichotomie (voir tableau 6.3 et tableau 6.2). L’entraînement a été réalisé à l’aide du framework open-source Pytorch. L’entraînement a été effectué sur un seul serveur avec 2× Tesla P100-PCIE-16GB.

Auto-encodeur LSTM sans attention Dans le réseau auto-encodeur LSTM sans attention, nous utilisons une séquence de longueur 20 avec un échantillonnage aléatoire de l’ensemble de données et l’hyperparamètre "seed= 4". L’encodeur est composé d’une couche d’entrée de taille de 16 correspondant aux 16 mesures et deux couches cachées LSTM de taille 128. Le décodeur comporte une couche d’entrée de taille 128 et de deux couches cachées LSTM de taille 128. Le réseau a été entraîné sur 50 époques avec une taille de lot "batch size" de 32. Le nombre de classes de sortie est de 4 et correspond aux situations de risques que nous choisissons de reconnaître. L’algorithme d’optimisation choisi est également Adam [Kingma et Ba \(2014\)](#) avec un taux d’apprentissage initial 0,0001. Nous avons appliqué la politique de décroissance du taux d’apprentissage qui s’est avérée efficace "reduce Learning rate decay" avec un "Decay Rate = 0.5" et un nombre d’époque de passage = 5. Nous avons appliqué la méthode Dropout pour prévenir l’overffiting avec $P = 0.2$.

L’entraînement a été réalisé à l’aide du framework open-source Pytorch. L’entraînement a été effectué sur un seul serveur avec 2× Tesla P100-PCIE-16GB.

LSTM auto-encodeur avec attention "dure" et avec attention "molle"

Pour l'auto-encodeur LSTM avec mécanismes d'attention l'architecture reste la même sauf concernant l'ajout des blocs d'attention. Un bloc d'attention stochastique ("dure") pour l'auto-encodeur LSTM avec attention dure et un bloc d'attention déterministe ("molle") pour l'auto-encodeur LSTM avec attention "molle" ont été rajoutés dans deux expériences distinctes. Nous appliquons la même optimisation des hyperparamètres et la même méthode d'optimisation Adam. Pour le dropout, nous l'avons aussi appliqué avec $P = 0.2$ pour pallier le problème d'overfitting.

6.4.4 Résultats de la classification

Dans cette section, nous comparons différentes méthodes de traitement de données sur le dataset BIRDS. Nous évaluons les performances des auto-encodeurs LSTM appliqués aux données multivariées pour la détection d'anomalie à savoir dans notre cas la détection de situations à risques pour les personnes fragiles. Nous avons aussi appliqué deux mécanismes d'attention sur les auto-encodeurs LSTM suivants : l'attention déterministe "molle" et l'attention stochastiques dite "dure" sur des données de santé multivariées et une classification multiple. En outre, nous analysons les résultats suivant les méthodes proposées et la performance des réseaux sur chaque classe qu'on a choisi d'étudier. Les architectures proposées ont été testés en utilisant l'ensemble des données BIRDS.

Le tableau 6.4 présente les scores obtenus selon les architectures proposées.

Les résultats obtenus avec les méthodes d'analyse (voir figure 6.4 et le tableau 6.4) montrent clairement une disparité dans les taux de détection des différentes classes analysées. Nous pouvons également voir que le réseau de neurones auto-encodeur avec attention "molle" produit de meilleurs résultats dans la détection des différentes situations à risques. Dans le tableau 6.4, nous présentons la précision par classe (Précision) pour toutes les classes du problème de classification multi-classes. Les valeurs de précision pour la classe *hydratation par apport hydrique* sont plus faibles par rapport aux autres classes, cela peut être dû au nombre plus faible d'échantillons par rapport aux autres classes. Mais d'après la matrice de confusion 6.4 et les valeurs de rappel qu'on peut lire dans le tableau 6.4, la valeur de rappel de cette classe est très faible, ce qui peut être dû au fait que les faux négatifs sont élevés pour cette classe. Comme le montre la matrice de confusion 6.4, aucun risque n'est prédit. Nous rappelons que l'hydratation est un risque "opposé", il s'agit de détecter ici l'action de boire. Il est normal qu'une telle action ne peut pas être bien reconnue en utilisant les signaux des capteurs seuls. L'analyse de la composante vidéo est nécessaire.

On peut aussi, voir des matrices de confusion et du tableau 6.4 que la classe risque d'accident domestiques est très bien détectée. Ceci est dû à la quantité de données assez conséquente par rapport aux autres classes, mais aussi au fait qu'il y ait une réponse physiologique spécifique au moment de l'exécution des tâches domestiques.

Nous remarquons que par rapport à la baseline - LSTM basique et sans les

F-Scores				
Algorithmes	cl(0)(%)	cl(1)(%)	cl(2)(%)	cl(3)(%)
LSTM auto-encodeur Chung et al. (2014b)	63.54	59.24	70.81	27.74
LSTM auto-encodeur + Hard Attn Xu et al. (2015)	61.11	57.84	69.87	25.83
LSTM auto-encodeur + Soft Attn Xu et al. (2015)	63.91	61.39	72.50	29.57
Precision				
Algorithmes	cl(0)(%)	cl(1)(%)	cl(2)(%)	cl(3)(%)
LSTM auto-encodeur Chung et al. (2014b)	61.51	60.23	66.84	41.07
LSTM auto-encodeur + Hard Attn Xu et al. (2015)	62.92	59.82	62.18	41.11
LSTM auto-encodeur + Soft Attn Xu et al. (2015)	63.04	61.98	68.02	41.27
Recall				
Algorithmes	cl(0)(%)	cl(1)(%)	cl(2)(%)	cl(3)(%)
LSTM auto-encodeur Chung et al. (2014b)	65.70	58.27	75.29	20.94
LSTM auto-encodeur + Hard Attn Xu et al. (2015)	59.39	55.98	79.73	18.83
LSTM auto-encodeur + Soft Attn Xu et al. (2015)	64.80	60.81	77.61	23.04

Tableau 6.4 – Comparaison des résultats obtenus avec les méthodes proposées sur le dataset BIRDS sélectionné avec les scores de précision et les F-Scores pour la classification multi-classes. Les classes sont représentées par Sans risque - cl(0), Risque de chute - cl(1), Risque d'accident domestique - cl(2), Hydratation par apport d'eau - cl(3).

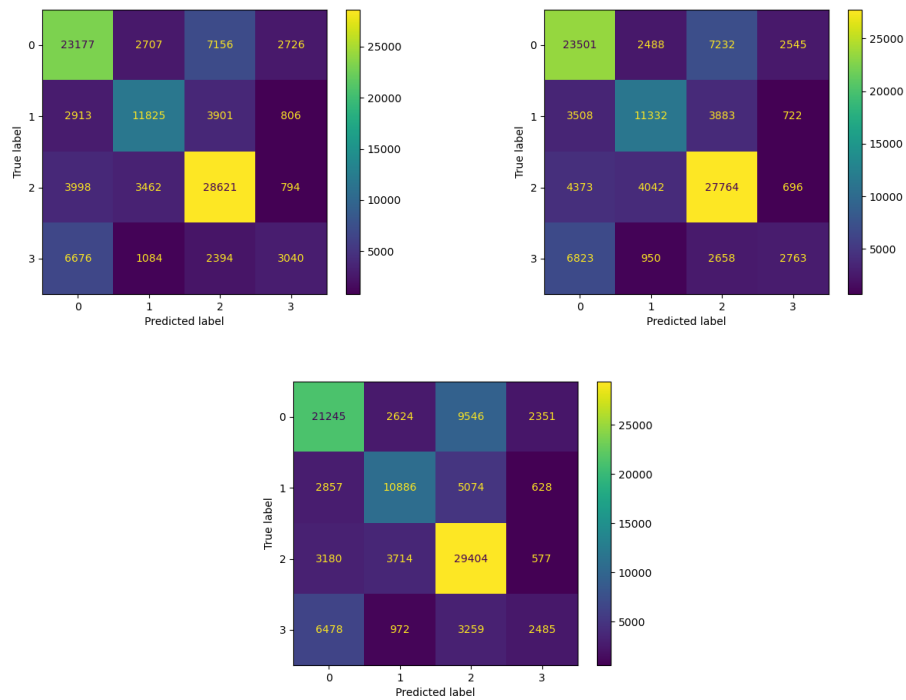


FIGURE 6.4 – Matrices de confusion obtenues avec les méthodes d’analyse de données. Dans l’ordre LSTM auto-encodeur avec attention molle et LSTM auto-encodeur sans attention, LSTM auto-encodeur avec attention dure proposées sur le dataset BIRDS avec la classification multi-classes. Les classes sont représentées par Sans risque - cl(0), Risque de chute - cl(1), Risque d’accident domestique - cl(2), Hydratation par apport d’eau - cl(3).

mécanismes d'attention, la *test accuracy* ne dépasse pas les 56,5 %. Avec les auto-encodeurs LSTM, nous arrivons à une *test accuracy* sur toutes les classes de 62,4 %. Quant à l'introduction des mécanismes d'attention "dure" et "molle" dans les réseaux auto-encodeurs LSTM, nous obtenons respectivement, une *test accuracy* 60.8 % avec l'attention "dure" et une *test accuracy* 63.82 % avec l'attention "molle". Il est important de mentionner que ces précisions concernent celles de toutes les classes réunies. Nous pouvons constater que nous obtenons les meilleurs résultats avec le réseau auto-encodeur LSTM avec attention "molle".

La contribution théorique de cette étude comprend deux parties. Tout d'abord, nous développons un modèle de prédiction basé de séries temporelles multivariées et multiclassées basé sur les auto-encodeurs LSTM avec une application des mécanismes d'attention. Les modèles développés sont testés dans le contexte de détection de situations à risques pour des personnes fragiles. Des situations que l'on pourrait qualifier de sémantiques.

Pour comparer nos modèles, nous avons pris le réseau LSTM simple en baseline et nous avons comparé les résultats des différentes architectures proposées. Les modèles que nous proposons, qui sont simplement basés des auto-encodeurs LSTM, sont évidemment assez simples même si l'ajout de l'attention nous permet d'explorer les mécanismes d'attention sur ce genre de modèle. Cependant, en considérant l'optimisation des paramètres (c'est-à-dire le taux d'apprentissage, le nombre de cellules et l'abandon) plutôt que de choisir une valeur prédéterminée, il a apporté une performance significativement plus élevée par rapport à la performance des réseaux de base. Ce résultat pourrait être très utile pour le traitement des séries temporelles multivariées avec des sorties multiclassées. L'application de ce genre de modèle pour la classification de situations de vie des personnes fragiles dites sémantiques est une nouveauté.

6.5 Conclusion

Dans ce chapitre, nous avons proposé notre approche à la détection des situations à risques sur les signaux des capteurs. Nous avons montré que cette détection s'apparente à la détection d'anomalies dans des séries temporelles. Ainsi, nous avons basé notre solution sur les réseaux LSTM pour leur efficacité sur les problèmes de classification ou de prédiction sur les séries temporelles.

Nous avons également fait appel aux auto-encodeurs LSTM qui permettent d'obtenir les représentations de données plus compactes et "saillantes" et ainsi avoir une détection plus ciblée.

Finalement, nous avons introduit les blocs d'attention en ayant expérimenté deux mécanismes : l'attention "molle" et l'attention "dure" et ce, afin d'améliorer la détection de ces situations à risques.

Les résultats obtenus montrent que les auto-encodeurs LSTM donnent de meilleurs résultats que les architectures LSTM basiques. Nous pouvons également voir que

le mécanisme d'attention "molle" montre les performances plus intéressantes, car il utilise le contexte général par le calcul de la moyenne pondérée des caractéristiques en sortie de l'encodeur. Néanmoins, les performances sont assez variables en fonction des classes, car certaines classes requièrent l'utilisation de la composante visuelle, notamment la classe *Hydratation par l'eau*. La fusion des mesures 1D avec des données visuelles dans un réseau hybride est aujourd'hui une des perspectives les plus prometteuses pour notre système de détection à multivariées en entrée et une multiclassés en sortie.

Chapitre 7

Conclusion générale et perspectives

Dans ce manuscrit de thèse, nous avons présenté nos contributions pour la détection de situations à risques pour les personnes âgées fragiles.

Nous avons développé le système de détection des situations à risques *ex nihili* de la conception du dispositif d'acquisition de données, aux solutions de détection par l'apprentissage profond.

Contrairement, aux différents systèmes de détection déjà connus dans la littérature, comme par exemple, les détecteurs de chute, il s'agit d'une approche nouvelle quand les risques encourus par la personne vivant dans la situation écologique à son domicile sont détectés pour prévenir l'événement dangereux, comme par exemple la chute, ou un accident domestique.

Les travaux de thèse se sont déroulés dans un contexte pluridisciplinaire, quand la taxonomie des situations à risques a été développée par des psychologues-gériatres et des spécialistes du mouvement humain. Il s'agit des risques que nous avons qualifiés de "sémantiques", c'est à dire correspondant aux situations de la vie quotidienne beaucoup plus complexes que la chute ou encore la baisse de tension artérielle.

Dans la première partie de ce travail, les chapitres 1 et 2 ont présenté le contexte et l'état de l'art des systèmes de détection de situations à risques pour les personnes âgées. Nous avons également fait l'état de l'art des méthodes d'apprentissage pour la détection d'actions ou d'événements par des méthodes de classification. Nous avons constaté alors l'hétérogénéité des méthodes et des systèmes de détection existant dans la littérature.

Dans le chapitre 3, nous avons présenté notre analyse des situations à risques et la mise en place d'un système de détection de ces situations que rencontrent les personnes fragiles au quotidien. Nous avons fait état des mesures existantes sur les systèmes portables et nous avons expliqué notre choix quant aux mesures adaptées à notre taxonomie.

Après le choix des mesures basé sur la taxonomie adoptée, nous avons expliqué la phase de conception du dispositif pour la collecte de données nécessaire à la mise en place de la solution BIRDS. La solution d'enregistrement des données a été basée

7. Conclusion générale et perspectives

sur les technologies BLE et a abouti à une application de collecte de données sur l'objet - concentrateur, le smartphone dédié.

Avec la société Inflexsys, nous avons conçu et développé l'application BIRDS pour un système Android. Les différents délais de transmission des capteurs ont nécessité le développement d'un algorithme de synchronisation des timestamps liés aux horloges internes des IoT concentrant les différents capteurs du système de collecte. La solution de synchronisation repose sur un estimateur au sens des moindres carrées (LMS) fonctionnant en temps réel. L'algorithme vient corriger le déphasage des timestamps des différents dispositifs de captation de données. Il a été intégré dans l'application BIRDS avec la participation de l'équipe du développement de la société Inflexsys.

Au vu de la nature complexe et sémantique de la taxonomie des situations de risques à détecter, l'utilisation de la composante visuelle a été nécessaire. Le corpus des données étant dans le processus d'enregistrement, nous avons développé une solution de détection des situations à risques sur un corpus Lifelog, enregistré par DCU (Dublin City University, corpus LSC NTCIR 13). Ces travaux ont été présentés dans le chapitre 4 du manuscrit. Généralement, les corpus Lifelogs sont enregistrés par des personnes en bonne santé et comportent les enregistrements continus de la journée à l'intérieur et à l'extérieur. Nous avons, avec la doctorante en psychologie Marion Pech, élaboré la correspondance des concepts entre les concepts primaires (548 concepts MS) du corpus et les situations à risques potentiels. Cette mise en correspondance est générique et est applicable à tous les corpus Lifelogs possibles pour le problème d'identification des risques potentiels. La mise en correspondance a été effectuée d'abord de façon automatique par l'analyse des méta-données du corpus NTCIR 13 et ensuite par une vérification visuelle avec l'interface d'annotation développée à cet effet. Cette vérification manuelle a également été nécessaire pour étiqueter les concepts absents dans le corpus et faisant partie de notre taxonomie, comme par exemple "l'escalier" qui correspond au risque de chute.

Ensuite, nous avons proposé la solution pour la détection automatique des risques à la base de réseaux de neurone profonds. Pour cela, nous avons étudié les architectures existantes et avons opté pour un réseau peu profond - AlexNet qui a montré une meilleure convergence lors de l'entraînement que le réseaux plus profond ResNet26 et ResNet10. Nous nous sommes intéressés à la classification binaire pour les deux situations à risques sémantiques : "risque d'accident domestique" et "risque de fraude". Les taux de reconnaissance (précision) obtenus ont été assez élevés (83,6 % pour le risque d'accident domestique et 96,8 % pour le risque de fraude).

Cette expérience nous a permis de voir la faisabilité de reconnaissance de risques sur la composante visuelle seule. Par ailleurs, sur ce corpus nous avons pu étudier le risque de fraude, ce qui n'était pas possible ensuite dans notre corpus BIRDS en raison de la protection de la vie privée.

La disponibilité du corpus Lifelog assez tôt dans notre étude quand le corpus réel BIRDS n'avait pas été encore enregistré, nous a permis de tester les solutions de l'état de l'art pour la détection des situations à risques sur les données des capteurs du corpus NTCIR 13. Cette détection a été réalisée à l'aide des réseaux RNN et nous a montré la nécessité de l'équilibrage de données préalable dans des corpus où les événements à détecter sont rares.

Dans le chapitre 5 du manuscrit, nous avons présenté le travail effectué pour :

- Enregistrer le corpus BIRDS ;
- Anonymiser les données pour leur usage dans des classifieurs profonds ;
- Pré-traiter les données en les synchronisant après l'enregistrement, effectuer leur imputation et normalisation ;
- Équilibrer les données pour leur utilisation dans des classifieurs profonds à la base de notre expérience sur le corpus NTCIR 13.

Plusieurs développements technologiques se sont avérés nécessaires. Nous avons mis en place des scripts de la synchronisation et de synthèse des données, deux interfaces d'annotation ergonomiques et utilisables par des non-informaticiens pour deux datasets différents et enfin de l'équilibrage des données.

Une fois le corpus écologique BIRDS enregistré, nous avons élaboré les méthodes de détection des risques sur les données en provenance des capteurs dans le chapitre 6. L'approche s'est basée sur les réseaux LSTM auto-encodeurs avec les blocs d'attention. Nous avons constaté que la solution proposée est plus performante qu'un LSTM de base. Néanmoins, une utilisation conjointe des données des capteurs et des données visuelles s'avère nécessaire ce qui reste en cours des travaux de notre projet.

En concluant ce travail de thèse et en traçant ses perspectives, nous constatons que nous sommes encore loin de pouvoir proposer une solution "clé en main" pour la détection des situations à risques chez les personnes fragiles vivant seules à leur domicile. Cette étude a avancé l'état de l'art aussi bien en conception des dispositifs d'acquisition, de préparation de données, de l'étude du comportement des réseaux profonds sur des données complexes bruitées et multimodales.

Si nous écartons les questions d'adaptabilité ergonomique du dispositif qui sont du ressort de nos collègues psychologues et gériatres, il reste plusieurs développements méthodologiques à poursuivre :

- Un travail est en cours de réalisation par mon collègue Rupayan MALLICK permettant l'analyse des données signaux et vidéos avec un réseau hybride prenant en compte les spécificités des différentes composantes ;
- La mise en libre service du dataset BIRDS sur les serveurs NAKALA du CNRS, et continuer à l'enrichir par des données provenant des expérimentations miroir menées par nos partenaires de l'université de Waterloo ;

7. Conclusion générale et perspectives

- Utilisation de l'apprentissage par transfert ;
- L'allègement des architectures Deep Learning pour qu'elles puissent tourner en temps réel sur les smartphones ;
- Enfin, notre travail sur le modèle basé sur l'apprentissage profond a suggéré des perspectives évidentes telles que la détection d'autres situations à risques survenant dans la vie quotidienne des personnes fragiles.

Annexe A

Publications liées à la thèse

- [Mallick et al. 2022a] MALLICK, Rupayan; BENOIS-PINEAU, Jenny; ZEMMARI, Akka; PECH, Marion; YEBDA, Thinhinane; AMIEVA, Hélène; MIDDLETON, Laura : A Hybrid Transformer Network for Detection of Risk Situations on Multimodal Life-Log Health Data. Dans : *Proceedings of the 3rd ACM Workshop on Intelligent Cross-Data Analysis and Retrieval*, 2022, pp. 22–26
- [Mallick et al. 2021] MALLICK, Rupayan; YEBDA, Thinhinane; BENOIS-PINEAU, Jenny; ZEMMARI, Akka; PECH, Marion; AMIEVA, Hélène : A GRU Neural Network with attention mechanism for detection of risk situations on multimodal lifelog data. Dans : *2021 International Conference on Content-Based Multimedia Indexing (CBMI) IEEE (Veranst.)*, 2021, pp. 1–6
- [Mallick et al. 2022b] MALLICK, Rupayan; YEBDA, Thinhinane; BENOIS-PINEAU, Jenny; ZEMMARI, Akka; PECH, Marion; AMIEVA, Hélène : Detection of Risky Situations for Frail Adults With Hybrid Neural Networks on Multimodal Health Data. Dans : *IEEE MultiMedia* 29 (2022), n. 1, pp. 7–17
- [Pech et al. 2021] PECH, Marion; SAUZEON, Helene; YEBDA, Thinhinane; BENOIT-PINEAU, Jenny; AMIEVA, Helene : Fall detection and prevention systems of homecare for the elderly: myth or reality? (2021)
- [Yebda et al. 2019] YEBDA, Thinhinane; BENOIS-PINEAU, Jenny; AMIEVA, Helene; FROLICHER, Benjamin : Multi-sensing of fragile persons for risk situation detection: devices, methods, challenges. Dans : *2019 International Conference on Content-Based Multimedia Indexing (CBMI) IEEE (Veranst.)*, 2019, pp. 1–6
- [Yebda et al. 2020] YEBDA, Thinhinane; BENOIS-PINEAU, Jenny; PECH, Marion; AMIÈVA, Hélène; GURRIN, Cathal : Detection of semantic risk situations in lifelog data for improving life of frail people. Dans : *Proceedings of the 2020 International Conference on Multimedia Retrieval*, 2020, pp. 402–406
- [Yebda et al. 2021] YEBDA, Thinhinane; BENOIS-PINEAU, Jenny; PECH, Marion; AMIEVA, Hélène; MIDDLETON, Laura; BERGELT, Max : Multimodal Sensor

Data Analysis for Detection of Risk Situations of Fragile People in@ home Environments. Dans : *International Conference on Multimedia Modeling* Springer (Veranst.), 2021, pp. 342–353

Bibliographie

- [Aimar et al. 2019] AIMAR, Emanuel S. ; RADEVA, Petia ; DIMICCOLI, Mariella : Social Relation Recognition in Egocentric Photostreams. Dans : *ICIP, IEEE*, 2019, pp. 3227–3231 99
- [Andreu-Perez et al. 2015] ANDREU-PEREZ, Javier ; LEFF, Daniel R. ; IP, Henry M. ; YANG, Guang-Zhong : From wearable sensors to smart implants—toward pervasive and personalized healthcare. Dans : *IEEE Transactions on Biomedical Engineering* 62 (2015), n. 12, pp. 2750–2762 36
- [Anguita et al. 2013] ANGUIITA, Davide ; GHIO, Alessandro ; ONETO, Luca ; PARRA, Xavier ; REYES-ORTIZ, Jorge L. : A Public Domain Dataset for Human Activity Recognition using Smartphones. Dans : *21st European Symposium on Artificial Neural Networks, ESANN 2013, Bruges, Belgium, April 24-26, 2013*, URL <http://www.eleucl.ac.be/Proceedings/esann/esannpdf/es2013-84.pdf>, 2013 111
- [Araújo Neto et al. 2017] ARAÚJO NETO, Antonio Herculano d. ; PATRÍCIO, Anna Cláudia Freire de A. ; FERREIRA, Milenna Azevedo M. ; RODRIGUES, Brenda Feitosa L. ; SANTOS, Thayná Dias dos ; RODRIGUES, Thays Domingos de B. ; SILVA, Richardson Augusto Rosendo d. : Falls in institutionalized older adults: risks, consequences and antecedents. Dans : *Revista brasileira de enfermagem* 70 (2017), n. 4, pp. 719–725 6
- [Assistance] ASSISTANCE, Arkea : https://www.arkeaassistance.fr/arkeaassistance/teleassistance/web/c_17661/fr/montre-connectee. – Accessed: 2018-08-31 7, 66
- [Aubry et al. 2018] AUBRY, Emilie ; FRIEDLI, Natalie ; SCHUETZ, Philipp ; STANGA, Zeno : Refeeding syndrome in the frail elderly population: prevention, diagnosis and management. Dans : *Clinical and experimental gastroenterology* 11 (2018), pp. 255 26
- [Avila-Funes et al. 2008] AVILA-FUNES, José A. ; HELMER, Catherine ; AMIEVA, Hélène ; BARBERGER-GATEAU, Pascale ; GOFF, Mélanie Le ; RITCHIE, Karen ; PORTET, Florence ; CARRIÈRE, Isabelle ; TAVERNIER, Béatrice ; GUTIÉRREZ-ROBLEDO, Luis M. et al. : Frailty among community-dwelling elderly people in France: the three-city study. Dans : *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 63 (2008), n. 10, pp. 1089–1096 16, 45
- [Bahdanau et al. 2014] BAHDANAU, Dzmitry ; CHO, Kyunghyun ; BENGIO, Yoshua : Neural machine translation by jointly learning to align and translate. Dans : *International Conference on Learning Representations* (2014) 132

-
- [Baig et al. 2019] BAIG, Mirza M. ; AFIFI, Shereen ; GHOLAMHOSSEINI, Hamid ; MIRZA, Farhaan : A Systematic Review of Wearable Sensors and IoT-Based Monitoring Applications for Older Adults—a Focus on Ageing Population and Independent Living. Dans : *Journal of medical systems* 43 (2019), n. 8, pp. 233 44
- [Banaee et al. 2013] BANAEI, Hadi ; AHMED, Mobyen U. ; LOUTFI, Amy : Data Mining for Wearable Sensors in Health Monitoring Systems: A Review of Recent Trends and Challenges. Dans : *Sensors* 13 (2013), n. 12, pp. 17472–17500. – URL <http://www.mdpi.com/1424-8220/13/12/17472>. – ISSN 1424-8220 80
- [Beard et al. 2009] BEARD, John R. ; BLANEY, Shannon ; CERDA, Magda ; FRYE, Victoria ; LOVASI, Gina S. ; OMPAD, Danielle ; RUNDLE, Andrew ; VLAHOV, David : Neighborhood characteristics and disability in older adults. Dans : *Journals of Gerontology: Series B* 64 (2009), n. 2, pp. 252–257 49
- [Beard et al. 2016] BEARD, John R. ; OFFICER, Alana ; DE CARVALHO, Islene A. ; SADANA, Ritu ; POT, Anne M. ; MICHEL, Jean-Pierre ; LLOYD-SHERLOCK, Peter ; EPPING-JORDAN, JoAnne E. ; PEETERS, GMEER G. ; MAHANANI, Wahyu R. et al. : The World report on ageing and health: a policy framework for healthy ageing. Dans : *The lancet* 387 (2016), n. 10033, pp. 2145–2154 16
- [Bell et al. 2020] BELL, Brooke M. ; ALAM, Ridwan ; ALSHURAF, Nabil ; THOMAZ, Edison ; MONDOL, Abu S. ; HAYE, Kayla de la ; STANKOVIC, John A. ; LACH, John ; SPRUIJT-METZ, Donna : Automatic, wearable-based, in-field eating detection approaches for public health research: a scoping review. Dans : *NPJ digital medicine* 3 (2020), n. 1, pp. 1–14 13
- [Bellagente et al. 2018] BELLAGENTE, Paolo ; CREMA, Claudio ; DEPARI, Alessandro ; FERRARI, Paolo ; FLAMMINI, Alessandra ; LANFRANCHI, Giovanni ; LENZI, Giovanni ; MADDIONA, Marco ; RINALDI, Stefano ; SISINNI, Emiliano et al. : Remote and non-invasive monitoring of elderly in a smart city context. Dans : *2018 IEEE Sensors Applications Symposium (SAS) IEEE (Veranst.)*, 2018, pp. 1–6 44
- [Bennett 2000] BENNETT, Jill A. : Dehydration: hazards and benefits. Dans : *Geriatric nursing* 21 (2000), n. 2, pp. 84–88 25
- [Bian et al. 2015] BIAN, Z. ; HOU, J. ; CHAU, L. ; MAGNENAT-THALMANN, N. : Fall Detection Based on Body Part Tracking Using a Depth Camera. Dans : *IEEE Journal of Biomedical and Health Informatics* 19 (2015), March, n. 2, pp. 430–439. – ISSN 2168-2194 23
- [Bizzego et al. 2020] BIZZEGO, Andrea ; GABRIELI, Giulio ; FURLANELLO, Cesare ; ESPOSITO, Gianluca : Comparison of wearable and clinical devices for acquisition of peripheral nervous system signals. Dans : *Sensors* 20 (2020), n. 23, pp. 6778 120
- [Blanpain et Chardon 2010a] BLANPAIN, N. ; CHARDON, O. : Projections de population à l’horizon 2060 : Un tiers de la population âgé de plus de 60 ans. Dans : *Insee Première* (2010) v, 15

- [Blanpain et Chardon 2010b] BLANPAIN, N.; CHARDON, O : Projections de population à l’horizon 2060 : Un tiers de la population âgé de plus de 60 ans. Dans : *Insee Première* (2010) 4, 14
- [Bolaños et al. 2017] BOLAÑOS, M.; DIMICCOLI, M.; RADEVA, P. : Toward Storytelling From Visual Lifelogging: An Overview. Dans : *IEEE Transactions on Human-Machine Systems* 47 (2017), Feb, n. 1, pp. 77–90. – ISSN 2168-2291 57
- [Boll et al. 2019] BOLL, Susanne; LEE, Jeannie S.; MEYER, Jochen; NAG, Nitish; O’CONNOR, Noel E. : HealthMedia’19: 4th International Workshop on Multimedia for Personal Health and Health Care. Dans : *ACM Multimedia*, ACM, 2019, pp. 2720–2721 36, 125
- [Boonnuddar et Wuttidittachotti 2017] BOONNUDDAR, Nichapat; WUTTIDITTACHOTTI, Pongpisit : Mobile Application: Patients’ Adherence to Medicine In-take Schedules. Dans : *Proceedings of the International Conference on Big Data and Internet of Thing*, 2017, pp. 237–241 33
- [Borzi et al. 2021] BORZÌ, Luigi; MAZZETTA, Ivan; ZAMPOGNA, Alessandro; SUPPA, Antonio; OLMO, Gabriella; IRRERA, Fernanda : Prediction of freezing of gait in Parkinson’s disease using wearables and machine learning. Dans : *Sensors* 21 (2021), n. 2, pp. 614 v, 30, 31
- [Bourke et Lyons 2008a] BOURKE, A. K.; LYONS, G. M. : A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor. Dans : *Elsevier BV* 26 (2008), n. 12. – URL <http://www.mdpi.com/1424-8220/13/12/17472> 19, 57
- [Bourke et Lyons 2008b] BOURKE, A. K.; LYONS, G. M. : A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor. Dans : *Elsevier BV* 30 (2008), n. 12. – URL <http://www.mdpi.com/1424-8220/13/12/17472> 20, 57
- [Bourke et Lyons 2008c] BOURKE, A.K.; LYONS, G.M. : A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor. Dans : *Medical Engineering & Physics* 30 (2008), n. 1, pp. 84 – 90. – URL <http://www.sciencedirect.com/science/article/pii/S1350453306002657>. – ISSN 1350-4533 58
- [Bradski 2000] BRADSKI, Gary : The openCV library. Dans : *Dr. Dobb’s Journal: Software Tools for the Professional Programmer* 25 (2000), n. 11, pp. 120–123 119
- [Bronzi et al. 2016] BRONZI, Walter; FRANK, Raphael; CASTIGNANI, German; ENGEL, Thomas : Bluetooth Low Energy Performance and Robustness Analysis for Inter-Vehicular Communications. Dans : *Ad Hoc Netw.* 37 (2016), February, n. P1, pp. 76–86. – URL <https://doi.org/10.1016/j.adhoc.2015.08.007>. – ISSN 1570-8705 63
- [Brown et Bussell 2011] BROWN, Marie T.; BUSSELL, Jennifer K. : Medication adherence: WHO cares? Dans : *Mayo clinic proceedings* 86 Elsevier (Veranst.), 2011, pp. 304–314 31

-
- [Buckinx et al. 2016] BUCKINX, Fanny; REGINSTER, Jean-Yves; PETERMANS, Jean; CROISIER, Jean-Louis; BEAUDART, Charlotte; BRUNOIS, Théo; BRUYÈRE, Olivier : Relationship between frailty, physical performance and quality of life among nursing home residents: the SENIOR cohort. Dans : *Aging clinical and experimental research* 28 (2016), n. 6, pp. 1149–1157 [16](#)
- [Cambois et al. 2013] CAMBOIS, Emmanuelle; BLACHIER, Audrey; ROBINE, Jean-Marie : Aging and health in France: an unexpected expansion of disability in mid-adulthood over recent years. Dans : *The European Journal of public health* 23 (2013), n. 4, pp. 575–581 [14](#)
- [Cao et al. 2016] CAO, H.; WU, S.; ZHOU, Z.; LIN, C.; YANG, C.; LEE, S.; WU, C. : A fall detection method based on acceleration data and hidden Markov model. Dans : *2016 IEEE International Conference on Signal and Image Processing (ICSIP)*, Aug 2016, pp. 684–689 [19](#), [51](#)
- [Caprani et al. 2014] CAPRANI, Niamh; PIASEK, Paulina; GURRIN, Cathal; O’CONNOR, Noel E.; IRVING, Kate; SMEATON, Alan F. : Life-Long Collections: Motivations and the Implications for Lifelogging with Mobile Devices. Dans : *Int. J. Mob. Hum. Comput. Interact.* 6 (2014), n. 1, pp. 15–36 [93](#)
- [Carreira et Zisserman 2017] CARREIRA, Joao; ZISSERMAN, Andrew : Quo vadis, action recognition? a new model and the kinetics dataset. Dans : *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6299–6308 [40](#)
- [Chaabouni et al. 2016] CHAABOUNI, S.; BENOIS-PINEAU, J.; AMAR, C. B. : Transfer learning with deep networks for saliency prediction in natural video. Dans : *2016 IEEE International Conference on Image Processing (ICIP)*, Sept 2016, pp. 1604–1608. – ISSN 2381-8549 [57](#)
- [Chalapathy et Chawla 2019] CHALAPATHY, Raghavendra; CHAWLA, Sanjay : Deep learning for anomaly detection: A survey. Dans : *arXiv preprint arXiv:1901.03407* (2019) [126](#)
- [Chang et al. 2019] CHANG, Wan-Jung; CHEN, Liang-Bi; HSU, Chia-Hao; LIN, Cheng-Pei; YANG, Tzu-Chin : A deep learning-based intelligent medicine recognition system for chronic patients. Dans : *IEEE Access* 7 (2019), pp. 44441–44458 [32](#)
- [Chawla et al. 2002] CHAWLA, Nitesh V.; BOWYER, Kevin W.; HALL, Lawrence O.; KEGELMEYER, W P. : SMOTE: synthetic minority over-sampling technique. Dans : *Journal of artificial intelligence research* 16 (2002), pp. 321–357 [122](#)
- [Che et al. 2018] CHE, Zhengping; PURUSHOTHAM, Sanjay; CHO, Kyunghyun; SONTAG, David; LIU, Yan : Recurrent neural networks for multivariate time series with missing values. Dans : *Scientific reports* 8 (2018), n. 1, pp. 1–12 [v](#), [37](#), [38](#)
- [Chen et al. 2015] CHEN, Tianqi; LI, Mu; LI, Yutian; LIN, Min; WANG, Naiyan; WANG, Minjie; XIAO, Tianjun; XU, Bing; ZHANG, Chiyuan; ZHANG, Zheng : Mxnet: A flexible

- and efficient machine learning library for heterogeneous distributed systems. Dans : *arXiv preprint arXiv:1512.01274* (2015) 103
- [Cheon et al. 2020] CHEON, Andy ; JUNG, Stephanie Y. ; PRATHER, Collin ; SARMIENTO, Matthew ; WONG, Kevin ; WOODBRIDGE, Diane Myung-kyung : A Machine Learning Approach to Detecting Low Medication State with Wearable Technologies. Dans : *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* IEEE (Veranst.), 2020, pp. 4252–4255 33
- [Chua et al. 2014] CHUA, Jia-Luen ; CHANG, Yoong C. ; JAWARD, Mohamed H. ; PARKINEN, Jussi ; WONG, Kok-Sheik : Vision-based hand grasping posture recognition in drinking activity. Dans : *2014 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)* IEEE (Veranst.), 2014, pp. 185–190 27
- [Chun et al. 2019] CHUN, Keum S. ; SANDERS, Ashley B. ; ADAIMI, Rebecca ; STREEPER, Necole ; CONROY, David E. ; THOMAZ, Edison : Towards a generalizable method for detecting fluid intake with wrist-mounted sensors and adaptive segmentation. Dans : *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 2019, pp. 80–85 vi, 27, 50, 51
- [Chung et al. 2014a] CHUNG, Junyoung ; GÜLÇEHRE, Çağlar ; CHO, KyungHyun ; BENGIO, Yoshua : Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. Dans : *CoRR* abs/1412.3555 (2014). – URL <http://arxiv.org/abs/1412.3555> 126
- [Chung et al. 2014b] CHUNG, Junyoung ; GULCEHRE, Caglar ; CHO, Kyunghyun ; BENGIO, Yoshua : Empirical evaluation of gated recurrent neural networks on sequence modeling. Dans : *NIPS 2014 Workshop on Deep Learning, December 2014*, 2014 141
- [Cohen et al. 2021] COHEN, Rachel ; FERNIE, Geoff ; ROSHAN FEKR, Atena : Fluid Intake Monitoring Systems for the Elderly: A Review of the Literature. Dans : *Nutrients* 13 (2021), n. 6, pp. 2092 v, 6, 13, 26
- [Corinna et Vapnik 1995] CORINNA, Cortes ; VAPNIK, V : Support-Vector Networks, Machine Learning. Dans : *Available from World Wide Web:< http://www. springerlink.com/content/k238jx04hm87j80g* (1995) 36
- [Coskun et al. 2013] COSKUN, Vedat ; OZDENIZCI, Busra ; OK, Kerem : A Survey on Near Field Communication (NFC) Technology. Dans : *Wireless Personal Communications* 71 (2013), Aug, n. 3, pp. 2259–2294. – URL <https://doi.org/10.1007/s11277-012-0935-5>. – ISSN 1572-834X 63
- [Cuevas-Trisan 2017] CUEVAS-TRISAN, Ramon : Balance problems and fall risks in the elderly. Dans : *Physical Medicine and Rehabilitation Clinics* 28 (2017), n. 4, pp. 727–737 vi, 46, 47
- [Das et al. 2020] DAS, Srijan ; SHARMA, Saurav ; DAI, Rui ; BREMOND, Francois ; THONNAT, Monique : Vpn: Learning video-pose embedding for activities of daily living. Dans : *European Conference on Computer Vision* Springer (Veranst.), 2020, pp. 72–90 41

-
- [Delgado-Escañó et al. 2020] DELGADO-ESCAÑO, Rubén ; CASTRO, Francisco M. ; CÓZAR, Julián R ; MARÍN-JIMÉNEZ, Manuel J. ; GUIL, Nicolás ; CASILARI, Eduardo : A cross-dataset deep learning-based classifier for people fall detection and identification. Dans : *Computer methods and programs in biomedicine* 184 (2020), pp. 105265 24
- [Dodge et Kitchin 2007] DODGE, Martin ; KITCHIN, Rob : ‘Outlines of a world coming into existence’: pervasive computing and the ethics of forgetting. Dans : *Environment and planning B: planning and design* 34 (2007), n. 3, pp. 431–445 92
- [Drover et al. 2017] DROVER, Dylan ; HOWCROFT, Jennifer ; KOFMAN, Jonathan ; LEMAIRE, Edward D. : Faller Classification in Older Adults Using Wearable Sensors Based on Turn and Straight-Walking Accelerometer-Based Features. Dans : *Sensors* 17 (2017), n. 6. – URL <http://www.mdpi.com/1424-8220/17/6/1321>. – ISSN 1424-8220 vi, 61
- [Du et al. 2019] DU, Bowen ; LU, Chris X. ; KAN, Xuan ; WU, Kai ; LUO, Man ; HOU, Jianfeng ; LI, Kai ; KANHERE, Salil ; SHEN, Yiran ; WEN, Hongkai : HydraDoctor: real-time liquids intake monitoring by collaborative sensing. Dans : *Proceedings of the 20th International Conference on Distributed Computing and Networking*, 2019, pp. 213–217 29
- [Elson et al. 2002] ELSON, Jeremy ; GIROD, Lewis ; ESTRIN, Deborah : Fine-grained Network Time Synchronization Using Reference Broadcasts. Dans : *SIGOPS Oper. Syst. Rev.* 36 (2002), December, n. SI, pp. 147–163. – URL <http://doi.acm.org/10.1145/844128.844143>. – ISSN 0163-5980 83
- [Empatica 2021] EMPATICA : *E4 Wristband User’s Manual*. <https://www.empatica.com/en-eu/>. 2021 vi, 73, 74
- [Fakhrulddin et al. 2017] FAKHRULDDIN, A. H. ; FEI, X. ; LI, H. : Convolutional neural networks (CNN) based human fall detection on Body Sensor Networks (BSN) sensor data. Dans : *2017 4th International Conference on Systems and Informatics (ICSAI)*, Nov 2017, pp. 1461–1465 20
- [Fedarko 2011] FEDARKO, Neal S : The Biology of Aging and Frailty. Dans : *Clinics in Geriatric Medicine* 27 (2011), 2, n. 1, pp. 27–37. – ISSN 0749-0690 16
- [Feichtenhofer et al. 2016] FEICHTENHOFER, C. ; PINZ, A. ; ZISSERMAN, A. : Convolutional Two-Stream Network Fusion for Video Action Recognition. Dans : *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1933–1941 vi, 40, 41
- [Fozoonmayeh et al. 2020] FOZOOMMAYEH, Donya ; LE, Hai V. ; WITTFOTH, Ekaterina ; GENG, Chong ; HA, Natalie ; WANG, Jingjue ; VASILENKO, Maria ; AHN, Yewon ; WOODBRIDGE, Diane Myung-kyung : A scalable smartwatch-based medication intake detection system using distributed machine learning. Dans : *Journal of medical systems* 44 (2020), n. 4, pp. 1–14 vi, 51, 53

- [Fried et al. 2001] FRIED, Linda P. ; TANGEN, Catherine M. ; WALSTON, Jeremy ; NEWMAN, Anne B. ; HIRSCH, Calvin ; GOTTDIENER, John ; SEEMAN, Teresa ; TRACY, Russell ; KOP, Willem J. ; BURKE, Gregory et al. : Frailty in older adults: evidence for a phenotype. Dans : *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 56 (2001), n. 3, pp. M146–M157 5, 6
- [Friedrich et al. 2021] FRIEDRICH, Björn ; LAU, Sandra ; ELGERT, Lena ; BAUER, Jürgen M ; HEIN, Andreas : A Deep Learning Approach for TUG and SPPB Score Prediction of (Pre-) Frail Older Adults on Real-Life IMU Data. Dans : *Healthcare* 9 Multidisciplinary Digital Publishing Institute (Veranst.), 2021, pp. 149 v, 6, 37, 38
- [Fukushima et Miyake 1982] FUKUSHIMA, Kunihiro ; MIYAKE, Sei : Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. Dans : *Competition and cooperation in neural nets*. Springer, 1982, pp. 267–285 35
- [Ganeriwal et al. 2003] GANERIWAL, Saurabh ; KUMAR, Ram ; SRIVASTAVA, Mani B. : Timing-sync Protocol for Sensor Networks. Dans : *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems*. New York, NY, USA : ACM, 2003 (SenSys '03), pp. 138–149. – URL <http://doi.acm.org/10.1145/958491.958508>. – ISBN 1-58113-707-9 83
- [Garbarino et al. 2014] GARBARINO, M. ; LAI, M. ; BENDER, D. ; PICARD, R. W. ; Tognetti, S. : Empatica E3 — A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. Dans : *2014 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*, Nov 2014, pp. 39–42 vi, 61, 67
- [Garcia-Constantino et al. 2020] GARCIA-CONSTANTINO, Matias ; KONIOS, Alexandros ; MUSTAFA, Mustafa A. ; NUGENT, Christopher ; MORRISON, Gareth : Ambient and Wearable Sensor Fusion for Abnormal Behaviour Detection in Activities of Daily Living. Dans : *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)* IEEE (Veranst.), 2020, pp. 1–6 29
- [Gasparriani et al. 2014] GASPARRINI, Samuele ; CIPPITELLI, Enea ; SPINSANTE, Susanna ; GAMBI, Ennio : A Depth-Based Fall Detection System Using a Kinect® Sensor. Dans : *Sensors* 14 (2014), n. 2, pp. 2756–2775. – URL <http://www.mdpi.com/1424-8220/14/2/2756>. – ISSN 1424-8220 23
- [Giacovelli et al. 2009] GIACOVELLI, A1. L. ; ELLI, M. ; GARIBOLDI, S. ; PELUCCHI, R. ; BONDI, G. ; D, H. B. : Modifications of vital signs during hygiene care in intensive care patients: an explorative study. Dans : *PMC* (2009) 54
- [Gislason 2008] GISLASON, Drew : *Zigbee Wireless Networking*. Pap/Onl. Newton, MA, USA : Newnes, 2008. – ISBN 0750685972, 9780750685979 62
- [Gomez et Paradells 2010a] GOMEZ, C. ; PARADELLS, J. : Wireless home automation networks: A survey of architectures and technologies. Dans : *IEEE Communications Magazine* 48 (2010), June, n. 6, pp. 92–101. – ISSN 0163-6804 63

-
- [Gomez et Paradells 2010b] GOMEZ, C.; PARADELLS, J. : Wireless home automation networks: A survey of architectures and technologies. Dans : *IEEE Communications Magazine* 48 (2010), June, n. 6, pp. 92–101. – ISSN 0163-6804 63
- [Gomez et al. 2012] GOMEZ, Carles; OLLER, Joaquim; PARADELLS, Josep : Overview and Evaluation of Bluetooth Low Energy: An Emerging Low-Power Wireless Technology. Dans : *Sensors* 12 (2012), n. 9, pp. 11734–11753. – URL <http://www.mdpi.com/1424-8220/12/9/11734>. – ISSN 1424-8220 83
- [González-Díaz et al. 2019] GONZÁLEZ-DÍAZ, Iván; BENOIS-PINEAU, Jenny; DOMENGER, Jean-Philippe; CATTART, Daniel; RUGY, Aymar de : Perceptually-guided deep neural networks for ego-action prediction: Object grasping. Dans : *Pattern Recognit.* 88 (2019), pp. 223–235 99
- [Goulding et al. 2003] GOULDING, Margie R.; ROGERS, ME; SMITH, SM : Public health and aging: trends in aging-United States and worldwide. Dans : *Jama* 289 (2003), n. 11, pp. 1371–1371 31
- [Greff et al. 2016] GREFF, Klaus; SRIVASTAVA, Rupesh K.; KOUTNÍK, Jan; STEUNBRINK, Bas R.; SCHMIDHUBER, Jürgen : LSTM: A search space odyssey. Dans : *IEEE transactions on neural networks and learning systems* 28 (2016), n. 10, pp. 2222–2232 105, 128
- [Gurrin et al. 2019a] GURRIN, Cathal; SCHOEFFMANN, Klaus; JOHO, Hideo; LEIBETSEDER, Andreas; ZHOU, Liting; DUANE, Aaron; DANG-NGUYEN, Duc-Tien; RIEGLER, Michael; PIRAS, Luca; TRAN, Minh-Triet; LOKOC, Jakub; HUERST, Wolfgang : [Invited papers] Comparing Approaches to Interactive Lifelog Search at the Lifelog Search Challenge (LSC2018). Dans : *ITE Transactions on Media Technology and Applications* 7 (2019), n. 2, pp. 46–59 99
- [Gurrin et al. 2019b] GURRIN, Cathal; SCHOEFFMANN, Klaus; JOHO, Hideo; MUNZER, Bernd; ALBATAL, Rami; HOPFGARTNER, Frank; ZHOU, Liting; DANG-NGUYEN, Duc-Tien : A test collection for interactive lifelog retrieval. Dans : *International Conference on Multimedia Modeling* Springer (Veranst.), 2019, pp. 312–324 viii, 91, 92, 99, 101, 105, 107, 113
- [Gurrin et al. 2014] GURRIN, Cathal; SMEATON, Alan F.; DOHERTY, Aiden R. : Lifelogging: Personal big data. Dans : *Foundations and trends in information retrieval* 8 (2014), n. 1, pp. 1–125 vii, 93, 94, 96
- [GUSTAVO HERNÁNDEZ-PEÑALOZA et ÁLVAREZ 2017] GUSTAVO HERNÁNDEZ-PEÑALOZA, MARCOS Q.; ÁLVAREZ, FEDERICO : A Multi-Sensor Fusion Scheme to Increase Life Autonomy of Elderly People With Cognitive Problems. Dans : *IEEE* (2017) 80
- [Gutiérrez et al. 2021] GUTIÉRREZ, Jesús; RODRÍGUEZ, Víctor; MARTIN, Sergio : Comprehensive review of vision-based fall detection systems. Dans : *Sensors* 21 (2021), n. 3, pp. 947 24

- [Habeeb et al. 2019] HABEEB, Riyaz Ahamed A.; NASARUDDIN, Fariza; GANI, Abdullah; HASHEM, Ibrahim Abaker T.; AHMED, Ejaz; IMRAN, Muhammad : Real-time big data processing for anomaly detection: A survey. Dans : *International Journal of Information Management* 45 (2019), pp. 289–307 [126](#)
- [Haykin et al. 2003] HAYKIN, Simon S.; WIDROW, Bernard; WIDROW, Bernard : *Least-mean-square adaptive filters*. 31. Wiley Online Library, 2003 [84](#)
- [He et al. 2016] HE, Kaiming; ZHANG, Xiangyu; REN, Shaoqing; SUN, Jian : Deep residual learning for image recognition. Dans : *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778 [99](#)
- [Hebb 1949] HEBB, Donald O. : *The organisation of behaviour: a neuropsychological theory*. Science Editions New York, 1949 [35](#)
- [Hegde et al. 2017] HEGDE, Nagaraj; BRIES, Matthew; SWIBAS, Tracy; MELANSON, Edward; SAZONOV, Edward : Automatic recognition of activities of daily living utilizing insole-based and wrist-worn wearable sensors. Dans : *IEEE journal of biomedical and health informatics* 22 (2017), n. 4, pp. 979–988 [44](#)
- [Hegde et Sazonov 2014] HEGDE, Nagaraj; SAZONOV, Edward : SmartStep: A fully integrated, low-power insole monitor. Dans : *Electronics* 3 (2014), n. 2, pp. 381–397 [44](#)
- [Hernández-Peñaloza et al. 2018] HERNÁNDEZ-PEÑALOZA, G.; BELMONTE-HERNÁNDEZ, A.; QUINTANA, M.; ÁLVAREZ, F. : A Multi-Sensor Fusion Scheme to Increase Life Autonomy of Elderly People With Cognitive Problems. Dans : *IEEE Access* 6 (2018), pp. 12775–12789. – ISSN 2169-3536 [4](#), [80](#)
- [Hilbe et al. 2010] HILBE, Johannes; SCHULC, Eva; LINDER, Barbara; THEM, Christa : Development and alarm threshold evaluation of a side rail integrated sensor technology for the prevention of falls. Dans : *International journal of medical informatics* 79 (2010), n. 3, pp. 173–180 [31](#)
- [Hochreiter et Schmidhuber 1997] HOCHREITER, Sepp; SCHMIDHUBER, Jürgen : Long short-term memory. Dans : *Neural computation* 9 (1997), n. 8, pp. 1735–1780 [37](#), [104](#), [126](#), [127](#), [136](#)
- [Hossain et Soh 2007] HOSSAIN, AKM M.; SOH, Wee-Seng : A comprehensive study of bluetooth signal parameters for localization. Dans : *2007 IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications* IEEE (Veranst.), 2007, pp. 1–5 [62](#)
- [Howcroft et al. 2017] HOWCROFT, J.; KOFMAN, J.; LEMAIRE, E. D. : Prospective Fall-Risk Prediction Models for Older Adults Based on Wearable Sensors. Dans : *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25 (2017), Oct, n. 10, pp. 1812–1820. – ISSN 1534-4320 [20](#)
- [Hu et al. 2019] HU, Shisheng; PEI, Yiyang; LIANG, Paul P.; LIANG, Ying-Chang : Deep neural network for robust modulation classification under uncertain noise conditions. Dans : *IEEE Transactions on Vehicular Technology* 69 (2019), n. 1, pp. 564–577 [120](#)

-
- [Huang et al. 2020] HUANG, Hsiang-Yun ; HSIEH, Chia-Yeh ; LIU, Kai-Chun ; HSU, Steen Jun-Ping ; CHAN, Chia-Tai : Fluid Intake Monitoring System Using a Wearable Inertial Sensor for Fluid Intake Management. Dans : *Sensors* 20 (2020), n. 22, pp. 6682 [27](#)
- [Hubara et al. 2017] HUBARA, Itay ; COURBARIAUX, Matthieu ; SOUDRY, Daniel ; EL-YANIV, Ran ; BENGIO, Yoshua : Quantized Neural Networks: Training Neural Networks with Low Precision Weights and Activations. Dans : *J. Mach. Learn. Res.* 18 (2017), pp. 187:1–187:30 [36](#)
- [Iosifidis et al. 2012] IOSIFIDIS, Alexandros ; MARAMI, Ermioni ; TEFAS, Anastasios ; PITAS, Ioannis : Eating and drinking activity recognition based on discriminant analysis of fuzzy distances and activity volumes. Dans : *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* IEEE (Veranst.), 2012, pp. 2201–2204 [28](#)
- [Jayatilaka et Ranasinghe 2017] JAYATILAKA, Asangi ; RANASINGHE, Damith C. : Real-time fluid intake gesture recognition based on batteryless UHF RFID technology. Dans : *Pervasive and Mobile Computing* 34 (2017), pp. 146–156 [v](#), [27](#), [29](#)
- [Jeske 2005] JESKE, D. R. : On maximum-likelihood estimation of clock offset. Dans : *IEEE Transactions on Communications* 53 (2005), Jan, n. 1, pp. 53–54. – ISSN 0090-6778 [83](#)
- [Ji et al. 2012] JI, Shuiwang ; XU, Wei ; YANG, Ming ; YU, Kai : 3D convolutional neural networks for human action recognition. Dans : *IEEE transactions on pattern analysis and machine intelligence* 35 (2012), n. 1, pp. 221–231 [vi](#), [39](#), [40](#)
- [Jia et al. 2014] JIA, Yangqing ; SHELLHAMER, Evan ; DONAHUE, Jeff ; KARAYEV, Sergey ; LONG, Jonathan ; GIRSHICK, Ross ; GUADARRAMA, Sergio ; DARRELL, Trevor : Caffe: Convolutional architecture for fast feature embedding. Dans : *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 675–678 [103](#)
- [J.Klenk et al. 2011] J.KLENK ; BECKER, C. ; F.LIEKEN ; S.NICOLAI ; W.MAETZLERAD ; W.ALT ; W.ZIJLSTRA ; J.M.HAUSDORFF ; LUMMEL, R.C.van ; L.CHIARIH ; U.LINDEMANN : Comparison of acceleration signals of simulated and real-world backward falls. Dans : *IRBM* 33 (2011), n. 3, pp. 368–373 [89](#)
- [Jokanovic et al. 2016a] JOKANOVIC, B. ; AMIN, M. ; AHMAD, F. : Radar fall motion detection using deep learning. Dans : *2016 IEEE Radar Conference (RadarConf)*, May 2016, pp. 1–6. – ISSN 2375-5318 [22](#)
- [Jokanovic et al. 2016b] JOKANOVIC, B. ; AMIN, M. G. ; AHMAD, F. : Effect of data representations on deep learning in fall detection. Dans : *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, July 2016, pp. 1–5. – ISSN 2151-870X [22](#)
- [Karaman et al. 2014] KARAMAN, Svebor ; BENOIS-PINEAU, Jenny ; DOVGALECS, Vladislavs ; MÉGRET, Rémi ; PINQUIER, Julien ; ANDRÉ-OBRECHT, Régine ; GAËSTEL, Yann ; DARTIGUES, Jean-François : Hierarchical Hidden Markov Model in detecting

- activities of daily living in wearable videos for studies of dementia. Dans : *Multi-media Tools and Applications* 69 (2014), Apr, n. 3, pp. 743–771. – URL <https://doi.org/10.1007/s11042-012-1117-x>. – ISSN 1573-7721 57
- [Kheirkhahan et al. 2019] KHEIRKHAHAN, Matin ; NAIR, Sanjay ; DAVOUDI, Anis ; RASHIDI, Parisa ; WANIGATUNGA, Amal A. ; CORBETT, Duane B. ; MENDOZA, Tonatiuh ; MANINI, Todd M. ; RANKA, Sanjay : A smartwatch-based framework for real-time and online assessment and mobility monitoring. Dans : *Journal of biomedical informatics* 89 (2019), pp. 29–40 44
- [Khojasteh et al. 2018] KHOJASTEH, Samad B. ; VILLAR, José R. ; CHIRA, Camelia ; GONZÁLEZ, Víctor M. ; CAL, Enrique De la : Improving fall detection using an on-wrist wearable accelerometer. Dans : *Sensors* 18 (2018), n. 5, pp. 1350 44
- [Kingma et Ba 2014] KINGMA, Diederik P. ; BA, Jimmy : Adam: A method for stochastic optimization. Dans : *arXiv preprint arXiv:1412.6980* (2014) 107, 139
- [Knickerbocker et al. 2018] KNICKERBOCKER, J. ; BUDD, R. ; DANG, B. ; CHEN, Q. ; COLGAN, E. ; HUNG, L. W. ; KUMAR, S. ; LEE, K. W. ; LU, M. ; NAH, J. W. ; NARAYANAN, R. ; SAKUMA, K. ; SIU, V. ; WEN, B. : Heterogeneous Integration Technology Demonstrations for Future Healthcare, IoT, and AI Computing Solutions. Dans : *2018 IEEE 68th Electronic Components and Technology Conference (ECTC)*, May 2018, pp. 1519–1528. – ISSN 2377-5726 7
- [Krizhevsky et al. 2012a] KRIZHEVSKY, Alex ; SUTSKEVER, Ilya ; HINTON, Geoffrey E. : Imagenet classification with deep convolutional neural networks. Dans : *Advances in neural information processing systems* 25 (2012), pp. 1097–1105 vii, 35, 100
- [Krizhevsky et al. 2012b] KRIZHEVSKY, Alex ; SUTSKEVER, Ilya ; HINTON, Geoffrey E. : ImageNet Classification with Deep Convolutional Neural Networks. Dans : PEREIRA, F. (Hrsg.) ; BURGESS, C. J. C. (Hrsg.) ; BOTTOU, L. (Hrsg.) ; WEINBERGER, K. Q. (Hrsg.): *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., 2012, pp. 1097–1105. – URL <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf> 99
- [Krizhevsky et al. 2012c] KRIZHEVSKY, Alex ; SUTSKEVER, Ilya ; HINTON, Geoffrey E. : Imagenet classification with deep convolutional neural networks. Dans : *Advances in neural information processing systems*, 2012, pp. 1097–1105 100
- [Lazarus et al. 2018] LAZARUS, Norman R. ; IZQUIERDO, Mikel ; HIGGINSON, Irene J. ; HARRIDGE, Stephen D. : Exercise deficiency diseases of ageing: the primacy of exercise and muscle strengthening as first-line therapeutic agents to combat frailty. Dans : *Journal of the American Medical Directors Association* 19 (2018), n. 9, pp. 741–743 16
- [Lecun et al. 1998] LECUN, Y. ; BOTTOU, L. ; BENGIO, Y. ; HAFNER, P. : Gradient-based learning applied to document recognition. Dans : *Proceedings of the IEEE* 86 (1998), Nov, n. 11, pp. 2278–2324. – ISSN 0018-9219 vi, 39

-
- [LeCun et al. 2015] LECUN, Yann ; BENGIO, Yoshua ; HINTON, Geoffrey : Deep learning. Dans : *nature* 521 (2015), n. 7553, pp. 436–444 39
- [LeCun et al. 1990] LECUN, Yann ; BOSER, Bernhard E. ; DENKER, John S. ; HENDERSON, Donnie ; HOWARD, R. E. ; HUBBARD, Wayne E. ; JACKEL, Lawrence D. : Handwritten Digit Recognition with a Back-Propagation Network. Dans : TOURETZKY, D. S. (Hrsg.): *Advances in Neural Information Processing Systems 2*. Morgan-Kaufmann, 1990, pp. 396–404. – URL <http://papers.nips.cc/paper/293-handwritten-digit-recognition-with-a-back-propagation-network.pdf> 39
- [LeCun et al. 1998] LECUN, Yann ; BOTTOU, Léon ; BENGIO, Yoshua ; HAFNER, Patrick : Gradient-based learning applied to document recognition. Dans : *Proceedings of the IEEE* 86 (1998), n. 11, pp. 2278–2324 35
- [Leng et Wu 2010] LENG, M. ; WU, Y. : On Clock Synchronization Algorithms for Wireless Sensor Networks Under Unknown Delay. Dans : *IEEE Transactions on Vehicular Technology* 59 (2010), Jan, n. 1, pp. 182–190. – ISSN 0018-9545 83, 84
- [Li et al. 2014] LI, Jiajia ; PEPLINSKI, Shaun J. ; NIA, Sarah M. ; FARAJIDAVAR, Aydin : An interoperable pillbox system for smart medication adherence. Dans : *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE (Veranst.)*, 2014, pp. 1386–1389 33
- [Li et al. 2017] LI, X. ; PANG, T. ; LIU, W. ; WANG, T. : Fall detection for elderly person care using convolutional neural networks. Dans : *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, Oct 2017, pp. 1–6 22, 39
- [Li et al. 2018] LI, Xue ; NIE, Lanshun ; XU, Hanchuan ; WANG, Xianzhi : Collaborative fall detection using smart phone and Kinect. Dans : *Mobile Networks and Applications* 23 (2018), n. 4, pp. 775–788 24
- [Liaqat et al. 2020a] LIAQAT, Sidrah ; DASHTIPOUR, Kia ; ARSHAD, Kamran ; RAMZAN, Naeem : Non invasive skin hydration level detection using machine learning. Dans : *Electronics* 9 (2020), n. 7, pp. 1086 27
- [Liaqat et al. 2020b] LIAQAT, Sidrah ; DASHTIPOUR, Kia ; ARSHAD, Kamran ; RAMZAN, Naeem : Non invasive skin hydration level detection using machine learning. Dans : *Electronics* 9 (2020), n. 7, pp. 1086 50
- [Liciotti et al. 2015] LICOTTI, D. ; MASSI, G. ; FRONTONI, E. ; MANCINI, A. ; ZINGARETTI, P. : Human activity analysis for in-home fall risk assessment. Dans : *2015 IEEE International Conference on Communication Workshop (ICCW)*, June 2015, pp. 284–289. – ISSN 2164-7038 23
- [Liu et al. 2020] LIU, Kai-Chun ; HSIEH, Chia-Yeh ; HUANG, Hsiang-Yun ; CHIU, Li-Tzu ; HSU, Steen Jun-Ping ; CHAN, Chia-Tai : Drinking Event Detection and Episode Identification Using 3D-Printed Smart Cup. Dans : *IEEE Sensors Journal* 20 (2020), n. 22, pp. 13743–13751 v, 28

- [Liu et al. 2011] LIU, L. ; POPESCU, M. ; SKUBIC, M. ; RANTZ, M. ; YARDIBI, T. ; CUDDIHY, P. : Automatic fall detection based on Doppler radar motion signature. Dans : *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, May 2011, pp. 222–225. – ISSN 2153-1633 22
- [Longari et al. 2020] LONGARI, Stefano ; VALCARCEL, Daniel Humberto N. ; ZAGO, Mattia ; CARMINATI, Michele ; ZANERO, Stefano : CANnolo: An anomaly detection system based on LSTM autoencoders for controller area network. Dans : *IEEE Transactions on Network and Service Management* (2020) 131
- [Lucine] LUCINE : <https://www.interaction-healthcare.com/creations/lucine/>. – Accessed: 2018-08-31 6, 7, 66
- [Luo et al. 2017] LUO, Y. ; TENG, K. ; LI, Y. ; MAO, W. ; HENG, C. ; LIAN, Y. : A 93 μ W 11Mbps wireless vital signs monitoring SoC with 3-lead ECG, bio-impedance, and body temperature. Dans : *2017 IEEE Asian Solid-State Circuits Conference (A-SSCC)*, Nov 2017, pp. 29–32 54, 80
- [Ma et al. 2019] MA, Chao ; SHIMADA, Atsushi ; UCHIYAMA, Hideaki ; NAGAHARA, Hajime ; TANIGUCHI, Rin-ichiro : Fall detection using optical level anonymous image sensing system. Dans : *Optics & Laser Technology* 110 (2019), pp. 44–61 24
- [Ma et al. 2018] MA, Jinxin ; OVALLE, Anaelia ; WOODBRIDGE, Diane Myung-kyung : Medhere: A smartwatch-based medication adherence monitoring system using machine learning and distributed computing. Dans : *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* IEEE (Veranst.), 2018, pp. 4945–4948 32
- [Mackensen et al. 2012] MACKENSEN, E. ; LAI, M. ; WENDT, T. M. : Bluetooth Low Energy (BLE) based wireless sensors. Dans : *SENSORS, 2012 IEEE*, Oct 2012, pp. 1–4. – ISSN 1930-0395 63
- [Maggs et al. 2012] MAGGS, M. K. ; O’KEEFE, S. G. ; THIEL, D. V. : Consensus Clock Synchronization for Wireless Sensor Networks. Dans : *IEEE Sensors Journal* 12 (2012), June, n. 6, pp. 2269–2277. – ISSN 1530-437X 81
- [Majumder et al. 2017] MAJUMDER, Sumit ; AGHAYI, Emad ; NOFERESTI, Moein ; MEMARZADEH-TEHRAN, Hamidreza ; MONDAL, Tapas ; PANG, Zhibo ; DEEN, M J. : Smart homes for elderly healthcare—Recent advances and research challenges. Dans : *Sensors* 17 (2017), n. 11, pp. 2496 ix, 64, 65
- [Maróti et al. 2004] MARÓTI, Miklós ; KUSY, Branislav ; SIMON, Gyula ; LÉDECZI, Ákos : The Flooding Time Synchronization Protocol. Dans : *Proceedings of the 2Nd International Conference on Embedded Networked Sensor Systems*. New York, NY, USA : ACM, 2004 (SenSys ’04), pp. 39–49. – URL <http://doi.acm.org/10.1145/1031495.1031501>. – ISBN 1-58113-879-2 82, 83

-
- [Martin et al. 2020a] MARTIN, Pierre-Etienne; BENOIS-PINEAU, Jenny; MANSENCAL, Boris; PÉTERI, Renaud; MASCARILLA, Laurent; CALANDRE, Jordan; MORLIER, Julien : Sports Video Classification: Classification of Strokes in Table Tennis for MediaEval 2020. Dans : HICKS, Steven (Hrsg.); JHA, Debesh (Hrsg.); POGORELOV, Konstantin (Hrsg.); HERRERA, Alba García Seco de (Hrsg.); BOGDANOV, Dmitry (Hrsg.); MARTIN, Pierre-Etienne (Hrsg.); ANDREADIS, Stelios (Hrsg.); DAO, Minh-Son (Hrsg.); LIU, Zhuoran (Hrsg.); QUIROS, José Vargas (Hrsg.); KILLE, Benjamin (Hrsg.); LARSON, Martha A. (Hrsg.): *Working Notes Proceedings of the MediaEval 2020 Workshop, Online, 14-15 December 2020* 2882, CEUR-WS.org, 2020. – URL <http://ceur-ws.org/Vol-2882/paper2.pdf> 119
- [Martin et al. 2020b] MARTIN, Pierre-Etienne; BENOIS-PINEAU, Jenny; PÉTERI, Renaud; MORLIER, Julien : Fine grained sport action recognition with Twin spatio-temporal convolutional neural networks. Dans : *Multimedia Tools and Applications* 79 (2020), n. 27, pp. 20429–20447 39, 40
- [Mattatia 2019] MATTATIA, Fabrice : *RGPD et droits des données personnelles*. Eyrolles, 2019 112
- [McCarthy et al. 2016] MCCARTHY, Cameron; PRADHAN, Nikhilesh; REDPATH, Calum; ADLER, Andy : Validation of the Empatica E4 wristband. Dans : *2016 IEEE EMBS international student conference (ISC) IEEE* (Veranst.), 2016, pp. 1–4 74
- [McCulloch et Pitts 1943] MCCULLOCH, Warren S.; PITTS, Walter : A logical calculus of the ideas immanent in nervous activity. Dans : *The bulletin of mathematical biophysics* 5 (1943), n. 4, pp. 115–133 35
- [McPhail 2016] MCPHAIL, Steven M. : Multimorbidity in chronic disease: impact on health care resources and costs. Dans : *Risk management and healthcare policy* 9 (2016), pp. 143 14
- [Meditkos et al. 2018] MEDITSKOS, Georgios; PLANS, Pierre-Marie; STAVROPOULOS, Thanos G.; BENOIS-PINEAU, Jenny; BUSO, Vincent; KOMPATSIARIS, Ioannis : Multimodal activity recognition from egocentric vision, semantic enrichment and lifelogging applications for the care of dementia. Dans : *J. Vis. Commun. Image Represent.* 51 (2018), pp. 169–190 37
- [Meek] MEEK, Theresa : *How Pharmacy Automation and IoT Devices Are Transforming Medicine*. https://www.insight.com/en_US/learn/content/190053001-automation-and-internet-of-things-transform-pharmacies.html. – Accessed: 2018-08-31 6, 7, 66
- [Merenda et al. 2020] MERENDA, Massimo; PORCARO, Carlo; IERO, Demetrio : Edge machine learning for ai-enabled iot devices: A review. Dans : *Sensors* 20 (2020), n. 9, pp. 2533 36
- [Minvielle et Audiffren 2019] MINVIELLE, Ludovic; AUDIFFREN, Julien : NurseNet: monitoring elderly levels of activity with a piezoelectric floor. Dans : *Sensors* 19 (2019), n. 18, pp. 3851 57

- [MONIZ-COOK et al. 1998] MONIZ-COOK, E. ; AGAR, S. ; GIBSON, G. ; WIN, T. ; WANG, M. : A preliminary study of the effects of early intervention with people with dementia and their families in a memory clinic. Dans : *Aging & Mental Health* 2 (1998), n. 3, pp. 199–211 48
- [Montero-Odasso et al. 2012] MONTERO-ODASSO, Manuel ; VERGHESE, Joe ; BEAUCHET, Olivier ; HAUSDORFF, Jeffrey M. : Gait and cognition: a complementary approach to understanding brain function and the risk of falling. Dans : *Journal of the American Geriatrics Society* 60 (2012), n. 11, pp. 2127–2136 30
- [Moshnyaga et al. 2016] MOSHNYAGA, Vasily ; KOYANAGI, Masaki ; HIRAYAMA, Fumi-yuki ; TAKAHAMA, Akihisa ; HASHIMOTO, Koji : A medication adherence monitoring system for people with dementia. Dans : *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC) IEEE (Veranst.)*, 2016, pp. 000194–000199 33
- [Mousavi et Saba 2010] MOUSAVI ; SABA : Depression, Chronic Diseases, and Decrements in Health: Results from the World Health Survey. Dans : *The Lancet* (2010) 5
- [Mukhopadhyay 2015] MUKHOPADHYAY, S. C. : Wearable Sensors for Human Activity Monitoring: A Review. Dans : *IEEE Sensors Journal* 15 (2015), March, n. 3, pp. 1321–1330. – ISSN 1530-437X 58
- [Nait Aicha et al. 2018a] NAIT AICHA, Ahmed ; ENGLEBIENNE, Gwenn ; SCHOOTEN, Kimberley S. van ; PIJNAPPELS, Mirjam ; KRÖSE, Ben : Deep Learning to Predict Falls in Older Adults Based on Daily-Life Trunk Accelerometry. Dans : *Sensors* 18 (2018), n. 5. – URL <http://www.mdpi.com/1424-8220/18/5/1654>. – ISSN 1424-8220 20, 21, 99
- [Nait Aicha et al. 2018b] NAIT AICHA, Ahmed ; ENGLEBIENNE, Gwenn ; VAN SCHOOTEN, Kimberley S. ; PIJNAPPELS, Mirjam ; KRÖSE, Ben : Deep learning to predict falls in older adults based on daily-life trunk accelerometry. Dans : *Sensors* 18 (2018), n. 5, pp. 1654 39
- [Nguyen et al. 2021] NGUYEN, HD ; TRAN, Kim P. ; THOMASSEY, Sébastien ; HAMAD, Moez : Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management. Dans : *International Journal of Information Management* 57 (2021), pp. 102282 126
- [Nolan et al. 2016] NOLAN, K. E. ; GUIBENE, W. ; KELLY, M. Y. : An evaluation of low power wide area network technologies for the Internet of Things. Dans : *2016 International Wireless Communications and Mobile Computing Conference (IWCMC)*, Sept 2016, pp. 439–444. – ISSN 2376-6506 63
- [Nouredanesh et al. 2021] NOUREDANESH, Mina ; GODFREY, Alan ; HOWCROFT, Jennifer ; LEMAIRE, Edward D. ; TUNG, James : Fall risk assessment in the wild: A critical examination of wearable sensor use in free-living conditions. Dans : *Gait & Posture* 85 (2021), pp. 178–190 31

-
- [Oh et al. 2015] OH, Hyungik ; JALALI, Laleh ; JAIN, Ramesh C. : An intelligent notification system using context from real-time personal activity monitoring. Dans : *ICME*, IEEE Computer Society, 2015, pp. 1–6 92
- [OMS] OMS : *Les chutes*. <http://www.who.int/fr/news-room/fact-sheets/detail/falls>. – Accessed: 2018-09-01 18
- [Ordóñez et Roggen 2016] ORDÓÑEZ, Francisco J. ; ROGGEN, Daniel : Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. Dans : *Sensors* 16 (2016), n. 1, pp. 115 27
- [Organization 2015] ORGANIZATION, World H. : *World report on ageing and health*. World Health Organization, 2015 v, 17
- [Organization et al. 2003] ORGANIZATION, World H. et al. : *Adherence to long-term therapies: evidence for action*. World Health Organization, 2003 32
- [Ozcan et al. 2013a] OZCAN, K. ; MAHABALAGIRI, A. K. ; CASARES, M. ; VELIPASALAR, S. : Automatic Fall Detection and Activity Classification by a Wearable Embedded Smart Camera. Dans : *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 3 (2013), June, n. 2, pp. 125–136. – ISSN 2156-3357 21
- [Ozcan et al. 2013b] OZCAN, K. ; MAHABALAGIRI, A. K. ; CASARES, M. ; VELIPASALAR, S. : Automatic Fall Detection and Activity Classification by a Wearable Embedded Smart Camera. Dans : *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 3 (2013), June, n. 2, pp. 125–136. – ISSN 2156-3357 21
- [PalChaudhuri et al. 2004] PALCHAUDHURI, Santashil ; SAHA, Amit K. ; JOHNSON, David B. : Adaptive Clock Synchronization in Sensor Networks. Dans : *Proceedings of the 3rd International Symposium on Information Processing in Sensor Networks*. New York, NY, USA : ACM, 2004 (IPSN '04), pp. 340–348. – URL <http://doi.acm.org/10.1145/984622.984672>. – ISBN 1-58113-846-6 81
- [Pandey et Janghel 2019] PANDEY, Saroj K. ; JANGHEL, Rekh R. : Recent deep learning techniques, challenges and its applications for medical healthcare system: A review. Dans : *Neural Processing Letters* 50 (2019), n. 2, pp. 1907–1935 37, 99
- [Patel et al. 2012] PATEL, Shyamal ; PARK, Hyung ; BONATO, Paolo ; CHAN, Leighton ; RODGERS, Mary : A review of wearable sensors and systems with application in rehabilitation. Dans : *Journal of NeuroEngineering and Rehabilitation* 9 (2012), Apr, n. 1, pp. 21. – URL <https://doi.org/10.1186/1743-0003-9-21>. – ISSN 1743-0003 58
- [Pérès et al. 2013] PÉRÈS, Karine ; EDJOLO, Arlette ; DARTIGUES, Jean-François ; BARBERGER-GATEAU, Pascale : Recent Trends in Disability-Free Life Expectancy in the French Elderly. Dans : *Annual Review of Gerontology and Geriatrics* 33 (2013), n. 1, pp. 293–311 14

- [Poh et al. 2010] POH, M.; SWENSON, N. C.; PICARD, R. W. : A Wearable Sensor for Unobtrusive, Long-Term Assessment of Electrodermal Activity. Dans : *IEEE Transactions on Biomedical Engineering* 57 (2010), May, n. 5, pp. 1243–1252. – ISSN 0018-9294 58
- [Purwar et al. 2007] PURWAR, Amit; JEONG, Do U.; CHUNG, Wan Y. : Activity monitoring from real-time triaxial accelerometer data using sensor network. Dans : *2007 International Conference on Control, Automation and Systems*, Oct 2007, pp. 2402–2406 58
- [Queralta et al. 2019] QUERALTA, J P.; GIA, Tuan N.; TENHUNEN, Hannu; WESTERLUND, Tomi : Edge-AI in LoRa-based health monitoring: Fall detection system with fog computing and LSTM recurrent neural networks. Dans : *2019 42nd international conference on telecommunications and signal processing (TSP) IEEE* (Veranst.), 2019, pp. 601–604 24
- [Raja et al. 2019] RAJA, Joel M.; ELSAKR, Carol; ROMAN, Sherif; CAVE, Brandon; POUR-GHAZ, Issa; NANDA, Amit; MATURANA, Miguel; KHOUZAM, Rami N. : Apple watch, wearables, and heart rhythm: where do we stand? Dans : *Annals of translational medicine* 7 (2019), n. 17 67
- [Rajagopalan et al. 2017] RAJAGOPALAN, Ramesh; LITVAN, Irene; JUNG, Tzyy-Ping : Fall Prediction and Prevention Systems: Recent Trends, Challenges, and Future Research Directions. Dans : *Sensors* 17 (2017), n. 11. – URL <http://www.mdpi.com/1424-8220/17/11/2509>. – ISSN 1424-8220 4, 54
- [Rajan et van der Veen 2011] RAJAN, R. T.; VEEN, A. van der : Joint ranging and clock synchronization for a wireless network. Dans : *2011 4th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec 2011, pp. 297–300 84
- [Regalia et al. 2019] REGALIA, Giulia; ONORATI, Francesco; LAI, Matteo; CABORNI, Chiara; PICARD, Rosalind W. : Multimodal wrist-worn devices for seizure detection and advancing research: focus on the Empatica wristbands. Dans : *Epilepsy research* 153 (2019), pp. 79–82 44
- [Reisig et Holtfreter 2013] REISIG, Michael D.; HOLTFRETER, Kristy : Shopping fraud victimization among the elderly. Dans : *Journal of Financial Crime* (2013) 49
- [Ren et Peng 2019] REN, Lingmei; PENG, Yanjun : Research of fall detection and fall prevention technologies: A systematic review. Dans : *IEEE Access* 7 (2019), pp. 77702–77722 24
- [René et Alexandre] RENÉ, WALLSTEIN; ALEXANDRE, COTARMANAC'H E. : *L'expansion des réseaux de données sans fil*. <https://www.universalis.fr/encyclopedie/telecommunications-la-communication-sans-fil/3-1-expansion-des-reseaux-de-donnees-sans-fil/>. – Accessed: 2018-08-10 vi, 64

-
- [Ricard et Thélot 2008] RICARD, Cécile; THÉLOT, Bertrand : Plusieurs centaines de milliers de chutes chez les personnes âgées chaque année en France. Dans : *Revue d'Épidémiologie et de Santé Publique* 56 (2008), n. 5, pp. 264 18
- [Rivera et al. 2014] RIVERA, L. R. ; ULMER, E. ; ZHANG, Y. D. ; TAO, W. ; AMIN, M. G. : Radar-based fall detection exploiting time-frequency features. Dans : *2014 IEEE China Summit International Conference on Signal and Information Processing (ChinaSIP)*, July 2014, pp. 713–717 23
- [Rosenblatt 1958] ROSENBLATT, Frank : The perceptron: a probabilistic model for information storage and organization in the brain. Dans : *Psychological review* 65 (1958), n. 6, pp. 386 35
- [Rosero-Montalvo et al. 2017] ROSERO-MONTALVO, P. D. ; PELUFFO-ORDÓÑEZ, D. H. ; GODOY, P. ; PONCE, K. ; ROSERO, E. A. ; VÁSQUEZ, C. A. ; CUZME, F. ; FLORES, S. C. ; MERA, Z. A. : Elderly fall detection using data classification on a portable embedded system. Dans : *2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM)*, Oct 2017, pp. 1–4 20
- [Rougier et al. 2011] ROUGIER, C. ; MEUNIER, J. ; ST-ARNAUD, A. ; ROUSSEAU, J. : Robust Video Surveillance for Fall Detection Based on Human Shape Deformation. Dans : *IEEE Transactions on Circuits and Systems for Video Technology* 21 (2011), May, n. 5, pp. 611–622. – ISSN 1051-8215 21
- [Rowat et al. 2012] ROWAT, Anne ; GRAHAM, Catriona ; DENNIS, Martin : Dehydration in hospital-admitted stroke patients: detection, frequency, and association. Dans : *Stroke* 43 (2012), n. 3, pp. 857–859 25
- [Rubenstein 2006] RUBENSTEIN, Laurence Z. : Falls in older people: epidemiology, risk factors and strategies for prevention. Dans : *Age and Ageing* 35 (2006), 09, n. suppl_2, pp. ii37–ii41. – ISSN 0002-0729 45
- [Rucco et al. 2018] RUCCO, Rosaria ; SORRISO, Antonietta ; LIPAROTI, Marianna ; FER-RAIOLI, Giampaolo ; SORRENTINO, Pierpaolo ; AMBROSANIO, Michele ; BASELICE, Fabio : Type and location of wearable sensors for monitoring falls during static and dynamic tasks in healthy elderly: a review. Dans : *Sensors* 18 (2018), n. 5, pp. 1613 13
- [Rumelhart et al. 1986] RUMELHART, David E. ; HINTON, Geoffrey E. ; WILLIAMS, Ronald J. : Learning representations by back-propagating errors. Dans : *nature* 323 (1986), n. 6088, pp. 533–536 35
- [Sabesan et Sankar 2015] SABESAN, Shivkumar ; SANKAR, Raman : Improving long-term management of epilepsy using a wearable multimodal seizure detection system. Dans : *Epilepsy & Behavior* 46 (2015), pp. 56–57 44
- [Seeman et al. 2010] SEEMAN, Teresa E. ; MERKIN, Sharon S. ; CRIMMINS, Eileen M. ; KARLAMANGLA, Arun S. : Disability trends among older Americans: national health and nutrition examination surveys, 1988–1994 and 1999–2004. Dans : *American journal of public health* 100 (2010), n. 1, pp. 100–107 14

- [Seiderer et al. 2017] SEIDERER, Andreas ; FLUTURA, Simon ; ANDRÉ, Elisabeth : Development of a mobile multi-device nutrition logger. Dans : *Proceedings of the 2nd ACM SIGCHI International Workshop on Multisensory Approaches to Human-Food Interaction*, 2017, pp. 5–12 29
- [Shany et al. 2011] SHANY, Tal ; REDMOND, Stephen J. ; NARAYANAN, Michael R. ; LOVELL, Nigel H. : Sensors-based wearable systems for monitoring of human movement and falls. Dans : *IEEE Sensors journal* 12 (2011), n. 3, pp. 658–670 31
- [Shih et al. 2015] SHIH, C. ; YANG, C. ; CHENG, Y. : Data Alignment for Multiple Temporal Data Streams without Synchronized Clocks on IoT Fusion Gateway. Dans : *2015 IEEE International Conference on Data Science and Data Intensive Systems*, Dec 2015, pp. 667–674 80
- [Simonyan et Zisserman 2014] SIMONYAN, Karen ; ZISSERMAN, Andrew : Two-stream convolutional networks for action recognition in videos. Dans : *arXiv preprint arXiv:1406.2199* (2014) 40
- [Singh et al. 2020] SINGH, Anuradha ; REHMAN, Saeed U. ; YONGCHAREON, Sira ; CHONG, Peter Han J. : Sensor technologies for fall detection systems: A review. Dans : *IEEE Sensors Journal* 20 (2020), n. 13, pp. 6889–6919 13
- [Soh et al. 2015] SOH, P. J. ; VANDENBOSCH, G. A. E. ; MERCURI, M. ; SCHREURS, D. M. M. .. : Wearable Wireless Health Monitoring: Current Developments, Challenges, and Future Trends. Dans : *IEEE Microwave Magazine* 16 (2015), May, n. 4, pp. 55–70. – ISSN 1527-3342 54
- [Sokullu et al. 2020] SOKULLU, Radosveta ; AKKAŞ, Mustafa A. ; DEMIR, Eren : IoT supported smart home for the elderly. Dans : *Internet of Things* 11 (2020), pp. 100239 vi, 48
- [Stevenson et al. 2016] STEVENSON, Jean E. ; ISRAELSSON, Johan ; NILSSON, Gunilla C. ; PETERSSON, Göran I. ; BATH, Peter A. : Recording signs of deterioration in acute patients: The documentation of vital signs within electronic health records in patients who suffered in-hospital cardiac arrest. Dans : *Health Informatics Journal* 22 (2016), n. 1, pp. 21–33. – URL <https://doi.org/10.1177/1460458214530136>. – PMID: 24782478 54
- [Taylor et al. 2014] TAYLOR, Morag E. ; DELBAERE, Kim ; LORD, Stephen R. ; MIKOLAJZAK, A S. ; BRODATY, Henry ; CLOSE, Jacqueline C. : Neuropsychological, physical, and functional mobility measures associated with falls in cognitively impaired older adults. Dans : *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* 69 (2014), n. 8, pp. 987–995 v, 30, 32
- [Taylor et Machado-Moreira 2013] TAYLOR, Nigel A. ; MACHADO-MOREIRA, Christiano A. : Regional variations in transepidermal water loss, eccrine sweat gland density, sweat secretion rates and electrolyte composition in resting and exercising humans. Dans : *Extreme physiology & medicine* 2 (2013), n. 1, pp. 1–30 58

-
- [Tong et al. 2009] TONG, L.; CHEN, W.; SONG, Q.; GE, Y. : A research on automatic human fall detection method based on wearable inertial force information acquisition system. Dans : *2009 IEEE International Conference on Robotics and Biomimetics (RO-BIO)*, Dec 2009, pp. 949–953 58
- [Tong et al. 2013] TONG, L.; SONG, Q.; GE, Y.; LIU, M. : HMM-Based Human Fall Detection and Prediction Method Using Tri-Axial Accelerometer. Dans : *IEEE Sensors Journal* 13 (2013), May, n. 5, pp. 1849–1856. – ISSN 1530-437X 20, 21, 58
- [Uddin 2019] UDDIN, Md Z. : A wearable sensor-based activity prediction system to facilitate edge computing in smart healthcare system. Dans : *Journal of Parallel and Distributed Computing* 123 (2019), pp. 46–53 31
- [Vaswani et al. 2017] VASWANI, Ashish; SHAZEER, Noam; PARMAR, Niki; USZKOREIT, Jakob; JONES, Llion; GOMEZ, Aidan N.; KAISER, Lukasz; POLOSUKHIN, Illia : Attention is All you Need. Dans : *NIPS*, 2017, pp. 5998–6008 132
- [Vemulapalli et al. 2017] VEMULAPALLI, Sai S.; ULAK, Mehmet B.; OZGUVEN, Eren E.; SANDO, Thobias; HORNER, Mark W.; ABDELRAZIG, Yassir; MOSES, Ren : GIS-based Spatial and Temporal Analysis of Aging-Involved Accidents: a Case Study of Three Counties in Florida. Dans : *Applied Spatial Analysis and Policy* 10 (2017), Dec, n. 4, pp. 537–563. – ISSN 1874-4621 16, 45, 91
- [Vhaduri et Poellabauer 2017] VHADURI, S.; POELLABAUER, C. : Wearable device user authentication using physiological and behavioral metrics. Dans : *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Oct 2017, pp. 1–6. – ISSN 2166-9589 67
- [Vigeo] VIGEO : <https://abo.vigeoapp.com>. – Accessed: 2018-08-31 7, 66
- [Vizitiu et al. 2021] VIZITIU, Cristian; BÎRĂ, Călin; DINCULESCU, Adrian; NISTORESCU, Alexandru; MARIN, Mihaela : Exhaustive Description of the System Architecture and Prototype Implementation of an IoT-Based eHealth Biometric Monitoring System for Elders in Independent Living. Dans : *Sensors* 21 (2021), n. 5, pp. 1837 6
- [Wang et al. 2015] WANG, R.; ZHANG, Y.; DONG, L.; LU, J.; ZHANG, Z.; HE, X. : Fall detection algorithm for the elderly based on human characteristic matrix and SVM. Dans : *2015 15th International Conference on Control, Automation and Systems (IC-CAS)*, Oct 2015, pp. 1190–1195. – ISSN 2093-7121 20
- [Wang et al. 2020] WANG, Zhuo; RAMAMOORTHY, Vignesh; GAL, Udi; GUEZ, Allon : Possible life saver: a review on human fall detection technology. Dans : *Robotics* 9 (2020), n. 3, pp. 55 111
- [Williams et al. 1998] WILLIAMS, G.; DOUGHTY, K.; CAMERON, K.; BRADLEY, D. A. : A smart fall and activity monitor for telecare applications. Dans : *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol.20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No.98CH36286)* 3, Nov 1998, pp. 1151–1154 vol.3. – ISSN 1094-687X 13, 18

- [Wu et al. 2007] WU, Jen-Her ; WANG, Shu-Ching ; LIN, Li-Min : Mobile computing acceptance factors in the healthcare industry: A structural equation model. Dans : *International journal of medical informatics* 76 (2007), n. 1, pp. 66–77 [32](#)
- [Xu et al. 2015] XU, Kelvin ; BA, Jimmy ; KIROS, Ryan ; CHO, Kyunghyun ; COURVILLE, Aaron ; SALAKHUDINOV, Ruslan ; ZEMEL, Rich ; BENGIO, Yoshua : Show, attend and tell: Neural image caption generation with visual attention. Dans : *International conference on machine learning* PMLR (Veranst.), 2015, pp. 2048–2057 [134](#), [141](#)
- [Yared et al. 2015] YARED, Rami ; MALLAT, Hady K. ; ABDULRAZAK, Bessam : Ambient technology to support elderly people in outdoor risk situations. Dans : *International Conference on Information and Communication Technologies for Ageing Well and e-Health* Springer (Veranst.), 2015, pp. 35–56 [18](#)
- [Yebda et al. 2020] YEBDA, Thinhinane ; BENOIS-PINEAU, Jenny ; PECH, Marion ; AMIÈVA, Hélène ; GURRIN, Cathal : Detection of Semantic Risk Situations in Lifelog Data for Improving Life of Frail People. Dans : *Proceedings of the 2020 International Conference on Multimedia Retrieval*, 2020, pp. 402–406 [vi](#), [44](#), [49](#), [58](#)
- [Zerveas et al. 2021] ZERVEAS, George ; JAYARAMAN, Srideepika ; PATEL, Dhaval ; BHAMIDIPATY, Anuradha ; EICKHOFF, Carsten : A Transformer-based Framework for Multivariate Time Series Representation Learning. Dans : *KDD*, ACM, 2021, pp. 2114–2124 [132](#)

Liste des acronymes

- AcVC** Accident de vie courante 18
- ADN** Acide DésoxyriboNucléique 7, 37, 66
- ANN** Articial Neural Network 35
- BLE** Bluetooth Low Energy 62, 63, 73, 74, 82
- bpm** Beats per Minute 55, 68
- BVP** Blood Volume Pressure 73, 76
- CNN** Convolutional Neural Networks vi, 20–22, 35–37, 39, 40
- ConvLSTM** Convolutional Long Short-Term Memory 20, 21, 39
- CSV** Comma-separated values 78
- DNN** Deep Neural Network 11
- ECG** Electrocardiogram 31, 59–61
- EDA** Electrodermal Activity 50, 58, 60, 67, 68, 74, 75, 89
- EEG** Électroencéphalogramme 59, 60
- EMG** Électromyographie 59, 60
- GPS** Global Positioning System 61, 66
- GPU** Graphics processing unit 36
- GRU** Gated Recurrent Unit v, 37, 38
- GRU-D** Gated Recurrent Unit with Decay v, 37, 38
- GSM** Global System for Mobile Communications 63
- GSR** Galvanic Skin Response 49, 50
- HMM** Hidden Markov Model 20, 21
- HR** Heart Rate 73
- I3D** Inflated 3D 40
- IBI** InterBeat Interval 73, 76
- IMU** Inertial Measure Units 33
- INSEE** l’Institut National de la Statistique et des Études Économiques 4, 14

- IoT** Internet Of Things 7, 10, 13, 29, 36, 41, 44, 48, 51, 57, 59, 62, 64, 117, 118, 121, 125, 137
- kNN** k-Nearest Neighbor 20, 22, 28
- LDA** Laser Doppler Anemometer 28
- LPV** Lower Peak Value 20
- LSTM** Long Short-Term Memory 10, 21, 24, 27, 37, 39, 127, 136
- MAC** Media Access Control 81
- MAPE** Mean Absolute Percentage Error 27
- MEMS** Micro Electro Mechanical Systems 59
- MLP** Multi Layer Perceptron 35
- NFC** Near Field Communication 59, 62
- OMS** Organisation Mondiale de la Santé 5
- PPG** Photoplethysmogramme 60, 73, 76
- ReLU** Rectified Linear Unit 100
- RFID** Radio frequency identification 27, 33
- RNN** Recurrent Neural Networks 31, 37
- SVM** Support Vector Machine 20–23, 27, 30, 36, 37, 126
- UPV** Upper Peak Value 20
- VPN** Video Pose Network 41
- WS** Wearable Sensors 44
- WSNs** Wireless Sensor Networks 59

Table des matières

1	Introduction générale	4
1.1	Contexte	4
1.1.1	La fragilité : mal des sociétés évoluées	5
1.1.2	Solutions proposées pour les populations fragiles	6
1.1.3	Traitement des données issus des systèmes eHealthcare et la part donnée au Deep Learning	8
1.2	Contributions	9
2	État de l’art	13
2.1	Introduction	13
2.2	Le vieillissement et la fragilité	14
2.2.1	Le vieillissement	14
2.2.2	Du vieillissement à la fragilité	15
2.3	Les objets connectés pour la surveillance des personnes âgées	18
2.3.1	Les systèmes de détection de chute	18
	Détection chute par capteurs portés	19
	Détection de chute par capteurs ambiants	21
	Détection de chute par fusion de données	23
2.3.2	Les systèmes de détection de déshydratation	25
	Détection du risque de déshydratation avec les technologies portables	26
	Détection du risque de déshydratation avec les technologies environnementales	27
	Détection du risque de déshydratation avec la fusion de données issues de différentes technologies	28
2.3.3	Détection du risque de chute chez les personnes âgées	30
2.3.4	Risque d’oubli de prise de médicaments	31
2.3.5	Analyses	34
2.4	Réseaux de neurones profonds pour les technologie eHealthcare	35
2.4.1	Réseaux de neurones profonds	35
2.4.2	L’apprentissage automatique au service des personnes âgées fragiles	36
2.5	Conclusion	41
3	De la taxonomie à la mise en place du dispositif d’acquisition	44
3.1	Introduction	44

3.2	Définition des situations à risques	45
3.2.1	Les situations à risques immédiates	45
	Risque d'accidents domestiques	48
	Risque d'intrusion d'une tierce personne	48
3.2.2	Situations à risques survenant dans le temps	50
	Risque de déshydratation	50
	Oubli de prise de médicament	50
3.2.3	Modèle des risques	52
3.3	Choix des mesures	54
3.3.1	Mesures physiologiques	54
	Paramètres vitaux	54
	Paramètres extra-vitaux	54
3.3.2	Mesures motrices	57
3.3.3	La vidéo	57
3.3.4	Choix de l'emplacement des capteurs	58
3.4	Technologies d'acquisition et de transmission pour la solution connectée	59
3.4.1	Technologies capteurs	59
3.4.2	Les objets connectés et la connectivité	61
	Les technologies de courte portée	62
	Les technologies à moyenne portée	63
	Les technologies à longue portée	63
3.5	Solution retenue	64
3.5.1	Architecture du dispositif	64
3.5.2	Premier prototype	66
	Principe du système	66
	Microsoft Band 2	67
	MetaMotionC Pro	69
	Caméra	69
3.5.3	Limitations du premier prototype	72
3.5.4	Implémentation de la solution finale	73
	Bracelet Empatica E4 Wearable	73
	MetaMotionLR	76
	Caméra USB	77
3.5.5	Acquisition des données	77
3.6	Synchronisations des données	80
3.6.1	Choix de l'approche par fusion de données	80
3.6.2	Synchronisation des données	81
3.6.3	Modélisation mathématique de la désynchronisation	82
3.6.4	Algorithmes de synchronisation dans la littérature	83
	Synchronisation par échange des timestamps	83
	Synchronisation des timestamps par estimation de la fréquence et du déphasage	83
3.6.5	Solution proposée	84

	Résultats de la synchronisation	85
3.6.6	Influence de la taille du filtre	85
	Influence du pas de descente de gradient	86
	Influence de la taille de l'échantillon de données	87
3.7	Conclusion	89
4	Analyses de situations à risques sur des données Lifelog	91
4.1	Introduction	91
4.2	Les données Lifelog	91
4.3	Présentation du dataset NTCIR 13	93
4.4	Mise en correspondance des concepts avec la taxonomie de risques et réannotation des données NTCIR 13	96
4.5	Reconnaissance des risques sémantiques sur les données du corpus NTCIR 13	98
4.5.1	Application des CNN sur la composante visuelle	98
	Choix de l'architecture et réglage du CNN pour la reconnaissance des situations à risque	100
	Préparation des données visuelles	101
	Le dataset	102
4.5.2	Entraînement du modèle	102
	Résultats de la détection de deux situations de risque sémantique	103
4.5.3	Application d'un LSTM sur des données (signal 1D)	104
4.5.4	Architecture LSTM	104
	Experimentations et resultats	107
4.6	Conclusion	108
5	Dataset Birds	109
5.1	Introduction	109
5.2	Dataset BIRDS	109
5.2.1	Protocole d'enregistrement des données	109
5.2.2	Description du corpus BIRDS	112
5.3	Annotation des données	115
5.4	Pré-traitement des données	117
5.4.1	La prise en compte du déphasage initial des capteurs	117
5.4.2	Anonymisation du corpus par image	118
5.4.3	Les données imparfaites	120
5.4.4	Équilibrage des données	122
5.5	Conclusion	123
6	Analyse des données BIRDS	125
6.1	Introduction	125
6.2	Méthodes de traitement des données	127

6.2.1	Les séries temporelles multi-entrées à classifications multiples avec les réseaux de neurones récurrents	127
	Architecture LSTM	128
6.2.2	L'auto-encodeur LSTM avec ajout de l'attention	129
6.3	Les réseaux de neurones auto-encodeurs avec attention	132
6.3.1	Le LSTM encodeur	133
6.3.2	Le décodeur LSTM	133
6.3.3	Les blocs d'attention	134
	L'attention déterministe "molle"	134
	L'attention stochastique "dure"	136
6.4	Expérimentations et résultats	136
6.4.1	Réseaux de neurones utilisés	136
6.4.2	Dataset	137
6.4.3	Paramètres et détails des entraînements	138
6.4.4	Résultats de la classification	140
6.5	Conclusion	143
7	Conclusion générale et perspectives	146
	Annexe A	151
	Bibliographie	153
	Liste des acronymes	173
	Table des matières	176