HAL open science

Modèles statistiques avancés pour la reconnaissance de l'activité physique dans un environnement non contrôlé en utilisant un réseau d'objets connectés

Hamdi Amroun

▶ **To cite this version:**

Hamdi Amroun. Modèles statistiques avancés pour la reconnaissance de l'activité physique dans un environnement non contrôlé en utilisant un réseau d'objets connectés. Intelligence artificielle [cs.AI]. Université Paris-Saclay, 2018. Français. NNT : 2018SACLS406 . tel-02314617v1

HAL Id: tel-02314617

https://theses.hal.science/tel-02314617v1

Submitted on 13 Oct 2019 (v1), last revised 29 Nov 2019 (v2)

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.

université PARIS-SACLAY

NNT : 2018SACLS406

Thèse de doctorat

UNIVERSITÉ PARIS SUD
Comprendre le monde, construire l'avenir

Modèles statistiques avancés pour la reconnaissance de l'activité physique dans un environnement non contrôlé en utilisant un réseau d'objets connectés

Thèse de doctorat de l'Université Paris-Saclay
Préparée à l'université Paris-Sud
Spécialité : Informatique

École doctorale n°580 : sciences et technologies de l'information et de la communication (STIC)

Thèse présentée et soutenue à Orsay, le 26/10/2018, par

Hamdi Amroun

Composition du Jury :

Dirk Heylen Professeur, Université de Twent (Informatique intelligente sociale)	Rapporteur
Mohamed Chetouani Professeur, Sorbonne Université (– ISIR, CNRS)	Rapporteur
Michel-Ange Amorim - Professeur, UPSud (– STAPS)	Président
Jean-Daniel Zucker Directeur de Recherche, Sorbonne Université (–UMMISCO)	Examinateur
Nizar Ouarti Maitre de Conférences IPAL, Singapour	Examinateur
Mehdi Ammi Maitre de Conférences, UPSud (– LIMSI)	Directeur de thèse

Titre : Modèles statistiques avancés pour la reconnaissance de l'activité physique dans un environnement non contrôlé.

Mots clés : Reconnaissance de l'activité physique, activités élémentaires, activités basées sur le contexte, Dempster-Shafer Theory, Optimisation par essaim particulaires, internet des objets.

Résumé : Avec l'arrivée des objets connectés, la reconnaissance de l'activité physique connait une nouvelle ère. De nouvelles considérations sont à prendre en compte afin d'aboutir à un meilleur processus de traitement. Dans cette thèse, nous avons exploré le processus de traitement pour la reconnaissance de l'activité physique dans un environnement non contrôlé. Les activités physiques reconnues, avec seulement une centrale inertielle (accéléromètre, gyroscope et magnétomètre), sont dites élémentaires. Les autres types d'activités dépendantes d'un contexte sont dites « basés sur le contexte ». Nous avons extrait la transformée en cosinus discrète (DCT) comme principal descripteur pour la reconnaissance des activités élémentaires.Afin de reconnaitre les activités physiques basées sur le contexte, nous avons défini trois niveaux de granularité : un premier niveau dépendant des objets connectés embarqués (smartphone, smartwatch et samrt TV). Un deuxième niveau concerne l'étude des comportements des participants en interaction avec l'écran de la smart TV. Le troisième niveau concerne l'étude de l'attention des participants envers la TV.

Nous avons pris en considération l'aspect imperfection des données en fusionnant les données multi capteurs avec le modèle de Dempster-Shafer. A ce titre, nous avons proposé différentes approches pour calculer et approximer les fonctions de masse.

Afin d'éviter de calculer et séléctionner les différents descripteurs, nous avons proposé une approche basée sur l'utilisation d'algorithmes d'apprentissage en profondeur (DNN).

Nous avons proposé deux modèles : un premier modèle consiste à reconnaitre les activités élémentaires en séléctionnant la DCT comme principal descripteur (DNN-DCT). Le deuxième modèle consiste à apprendre les données brutes des activités basées sur le contexe (CNN-brutes). L'inconvénient du modèle DNN-DCT est qu'il est rapide mais moins précis, alors que le modèle CNN-brutes est plus précis mais très lent. Nous avons proposé une étude empirique permettant de comparer les différentes méthodes pouvant accélérer l'apprentissage tout en gardant un niveau élevé de précision. Nous avons ainsi exploré la méthode d'optimisation par essaim particulaires (PSO). Les résultats sont très satisfaisants (97%) par rapport à l'apprentissage d'un réseau de neurones profond avec les méthodes d'optimisation classiques telles que la descente de Gradient Stochastique et l'optimisation par Gradient accéléré de Nesterov. Les résultats de nos travaux suggèrent le recours à de bons descripteurs dans le cas où le contexte n'importe peu, la prise en compte de l'imperfection des données capteurs quand le domaine sous jacent l'exige, l'utilisation de l'apprentissage profond avec un optimiseur permettant d'avoir des modèles très précis et plus rapides.

Title: Advanced statistical models for recognizing physical activity in an uncontrolled environment.

Keywords: Physical Activity Recognition, Basic Activities, Context Based Activities, Dempster-Shafer Theory, Particle Swarm Optimization, Internet of Things.

Abstract: With the arrival of connected objects, the recognition of physical activity is experiencing a new era. New considerations need to be taken into account in order to achieve a better treatment process. In this thesis, we explored the treatment process for recognizing physical activity in an uncontrolled environment. The recognized physical activities, with only one inertial unit (accelerometer, gyroscope and magnetometer), are called elementary. Other types of context-dependent activities are called "context-based". We extracted the DCT as the main descriptor for the recognition of elementary activities. In order to recognize the physical activities based on the context, we defined three levels of granularity: a first level depending on embedded connected objects (smartphone, smartwatch and samrt TV) . A second level concerns the study of participants' behaviors interacting with the smart TV screen. The third level concerns the study of participants' attention to TV. We took into consideration the imperfection aspect of the data by merging the multi sensor data with the Dempster-Shafer model. As such, we have proposed different approaches for calculating and approximating mass functions. In order to avoid calculating and selecting the different descriptors, we proposed an approach based on the use of deep learning algorithms (DNN). We proposed two models: a first model consisting of recognizing the elementary activities by selecting the DCT as the main descriptor (DNN-DCT). The second model is to learn raw data from context-based activities (CNN-raw). The disadvantage of the DNN-DCT model is that it is fast but less accurate, while the CNN-raw model is more accurate but very slow. We have proposed an empirical study to compare different methods that can accelerate learning while maintaining a high level of accuracy. We thus explored the method of optimization by particle swarm (PSO). The results are very satisfactory (97%) compared to deep neural network with stochastic gradients descent and Nesterov accelerated Gradient optimization. The results of our work suggest the use of good descriptors in the case where the context matters little, the taking into account of the imperfection of the sensor data requires that it be used and faster models.

PUBLICATIONS

JOURNAUX PUBLIÉS

H. Amroun and M.Ammi, « Who used my smart object? A flexible appraoch for the recognition of users ». IEEE Access, 2018 (IF=3.4).

H. Amroun, M.Y.Benziani and M.Ammi, « Dempster-Shafer Theory of Evidence for modeling hot water consumption ». IEEE Access 2018 (IF=3.4).

H. Amroun, M.Y.Benziani and M.Ammi, « Dempster-Shafer Theory of Evidence for modeling Nicotine consumption using an e-cigarette». IEEE Access 2018 (IF=3.4).

CONFÉRENCES INTERNATIONALES

H.Amroun, M.Y.Benziani and M.Ammi "Long Short Term Memory Based Model for Modeling Nicotine Consumption Using an Electronic Cigarete and Internet of Things Devices" In the 20th International Conference on Artificial Intelligence and Cognitive Science pp. 334–340, Tokyo, Japan, March 2018.

H.Amroun, M.Y.Benziani and M.Ammi " A Framework Based on Dempster-Shafer Theory of Evidence Algorithm for The Analysis of the TV-Viewers' Behaviors " In the 20th International Conference on Artificial Intelligence and Cognitive Science pp. 424–433, Tokyo, Japan, March 2018.

H.Amroun, N.Ouarti and M.Ammi Recognition of human activity using Internet of Things in a non-controlled environment " In the 14th International Conference on Control, Automation, Robotics and Vision (ICARCV) pp. 1-6, Phuket, Thailande, November 2016.

H.Amroun, M.H.Temkit and M.Ammi recognition of human activity using paired connected objects " In the 10th IEEE International Conference on internet of things (iTHINGS) pp. 17-23, Exeter, UK, June 2017.

H.Amroun, M.H.Temkit and M.Ammi "Best features for CNN classification of human activity using IOT network» In the 10th IEEE International Conference on internet of things (iTHINGS) pp. 17-23, Exeter, UK, June 2017.

H.Amroun, M.H.Temkit and M.Ammi "Impact of the positions transition of a smartphone on human activity recognition» In the 10th IEEE International Conference on internet of things (iTHINGS) pp. 17-23, Exeter, UK, June 2017.

H.Amroun, M.H.Temkit and M.Ammi "DNN-based approach for recognition of human activity raw data in Non-controlled environment» 6th IEEE International Conference on AI & Mobile Services pp.1-4, june, 2017, Honolulu, USA.

H.Amroun, M.H.Temkit and M.Ammi “DNN-based approach for identification of the level of attention of the TV-viewers using IoT network» The 3rd IEEE International Conference on Internet of People pp.1-6, august, 2017, San Francisco, USA

H.Amroun, M.H.Temkit and M.Ammi “Study of the viewers’TV-watching behaviors before, during and after watching a TV program using IoT network» The 3rd IEEE International Conference on Systems, Man, and Cybernetics (SMC 2017), pp.1781-1792, august, 2017, Banff, Canada.

Maxence Bobin, **H.Amroun**, S.Coquillard, F.Bimbard and M.Ammi “DNN based Approach for the Assessment of Elbow Flexion with Smart Textile Sensor » The 3rd IEEE International Conference on Systems, Man, and Cybernetics (SMC 2017), pp.1183-1190, august, 2017, Banff, Canada.

Maxence Bobin, **H.Amroun**, M.Boukallel, F.Bimbard and M.Ammi “Smart Cup for Festival Alcohol Consumption Awareness» The 4th IEEE World Forum on Internet Of Things (WF-IoT) pp.606-612, December, Singapore.

Maxime De Bois, **H.Amroun** and M.Ammi “Energy Expenditure Estimation Through Daily Activity Recognition Using a Smart-phone'» The 4th IEEE World Forum on Internet Of Things (WF-IoT) pp.714-721, December, Singapore.

PRIX ET DISTINCTIONS

Best paper award, IEEE International Conference on Systems, Man, and Cybernetics (SMC 2017),

Best presentation award, 20th the 20th International Conference on Artificial Intelligence and Cognitive Science pp. 334–340, Tokyo, Japan, March 2018.

REMERCIEMENTS

Je tiens à remercier par ces quelques mots toutes les personnes qui m'ont permis de concrétiser cette thèse. J'adresse ainsi en premier lieu et sans ambage mes plus vifs remerciements à Mehdi Ammi, directeur de thèse, qui a su m'initier à la liberté de la recherche et par là-même à ses joies. J'adresse ainsi par la meme occasion, mes remerciements à Nizar Ouarti pour m'avoir accompagné durant ces travaux de thèse. Je remercie très chalheureusement la société IOfMe pour la confiance qu'elle m'a accordée pour la réalisation de ce travail de thèse.

Je tiens à remercier vigoureusement Mohamed Chetouani et D.K.J. Heylen (Dirk), rapporteurs du jury de soutenance, pour avoir accepté d'étudier mon travail de thèse et m'avoir permis de prendre du recul par rapport à ces travaux. Je tiens également à remercier très chaleureusement Marc Schmidt, pour ses indications très lumineuses pour moi à propos du choix des distributions candidates pour les algorithmes de simulation à sauts réversibles de Green et le travail de fond sur les techniques d'apprentissage en profondeur : ses remarques très pertinentes m'ont permis d'améliorer la lisibilité de la thèse de manière certaine. Je remercie enfin du fond du cœur M'Hamed Hamy Temkit et Yacine Benziani pour leurs conseils et indications pour l'exploration des fonctions de croyances pour la reconnaissance de l'activité physique. La rigueur de Yacine et l'ouverture d'esprit de Hamy m'ont permis de prendre connaissance de l'ampleur et l'impact du choix de ce type d'approche.

La thèse se concrétise au final par une soutenance ponctuelle mais elle est le fruit d'un travail de recherche journalier de trois années. Celles-ci se sont écoulées de manière fort agréable pour moi grâce à la rencontre au laboratoire de nombreuses personnes attachantes que je vais maintenant remercier un peu en rafale peut-être mais très chaleureusement : Mohamed Yacine Tsalamlal, Amine Benchiheub, Maxence Bobin, Nassim Hadjadj et Maxime Debois.

Je remercie également les autres thésards ou non qui ont contribué, contribuent et contribueront à la bonne ambiance de l'équipe : Adrien Arnaud et Jean Baptiste corrégé.

De manière générale, la Recherche n'excluant pas complètement d'autres sujets en dehors des méthodes de reconnaissance de l'activité physique, je remercie donc Issam Laradji de l'Université de Colombie Britannique (UBC) à Vancouver pour m'avoir fait découvrir récemment de manière sympa les techniques de représentations diffusives pour les opérateurs pseudo-différentiels.

Enfin je voudrais remercier tous mes amis ainsi que ma famille : mes parents et ma sœur, pour leur soutien et leurs encouragements. J'ai particulièrement une pensée très émue pour mon grand ami Fayçal disparu tragiquement en mai 2003, s'il y a bien une chose qu'il m'a apprise c'est qu'il faut jouir de tous les instants présents et à venir. C'est ce que je désire faire le plus ardamment avec Sara, ma compagne sensationnelle, sans qui je n'aurais jamais pu accomplir cette thèse. Je conclus ainsi ces remerciements en lui adressant ma reconnaissance la plus profonde pour son soutien et son amour.

Sommaire

Introduction et contexte général……10

Chapitre 1 : Etat de l'art

1. Introduction 15
2. L'internet des objets et les objets connectés 16
2.1. L'internet des objets.....16
2.2. Les objets connectés 19
3. Architecture de traitement 22
3.1. Extraction de descripteurs 23
3.2. Fusion de données multi capteurs 27
3.3. Apprentissage automatique pour la reconnaissance de l'activité 40
4. Limites de l'état de l'art.....63
4.1. Reconnaissance de l'activité physique élémentaire et basée sur le contexte.....63
4.2. Prise en compte de l'imperfection des données des objets connectés 65
4.3.Apprentissage en profondeur pour la reconnaissance de l'activité humaine 65

Chapitre 2 : Reconnaissance de l'activité physique élémentaire avec des objets connectés.

1. Introduction 68
2. Vue d'ensemble des traitements.....69
3. Création du corpus 70
3.1. Description de l'expérience.....70
3.2. Pré-analyse du corpus.....72
4. Extraction de descripteurs 74
4.1. Découpage et segmentation des signaux 76
4.2. Taille d'une coupe 77
4.3. Résultats Globaux 77
4.4. La DCT comme principal descripteur.....83
4.5. Influence chevauchement entre activités élémentaires.....90
5. Discussion.....91

Chapitre 3 : Reconnaissance de l'activité physique basée sur le contexte.

1. Introduction …… ..... …95
2. Etat de l'art ……...…97

3. Activités de premier niveau ..... ….100
3.1. Statistiques descriptives sur le corpus d'activités 100
3.2. Annotation des données ……..... …100
3.3. Pré-analyse du corpus …. ..... …101
4. Processus de classification des activités physiques. …..... …..102
4.1. Extraction de descripteurs …..... …103
5. Activités physiques de Deuxième niveau : comportement des visionneurs d'un programme TV …..... …107
5.1. Distinction entre les trois phases de visionnage d'un programme TV …..... ……109
6. Activités physiques de Troisième niveau : détection de l'attention envers la TV …… . …110
6.1. Segmentation des individus…….. …...113
6.2. Le descripteur de l'attention envers la TV……. …113
7. Synthèse des chapitre 2 et 3…… ……..117
8. Limites des études effectuées dans les chapitres 2 et 3… ..... …117

Chapitre 4 : Prise en compte de l'imperfection des données pour la reconnaissance de l'activité physique

1. Introduction……120
2. Etat de l'art……..121
2.1. Approximation des fonctions de masse ……...……..123
3. Vue d'ensemble de la DST pour la reconnaissance de l'activité 127
3.1. Implémentation…….. 131
4. Résultats……135
4.1. Activités physiques élémentaires.....135
4.2. Activités physiques basées sur le contexte.....136
5. Discussion…… 142

Chapitre 5 : Optimisation de l'architecture d'entrainement des données de l'activité physique par apprentissage en profondeur

1. Introduction …… ....…146
2. Etude préliminaire : Apprentissage d'un réseau de neurones profonds avec sélection de descripteurs ……..147
3. Apprentissage en profondeur des données brutes de l'activité physique humaine ……150
3.1. Le modéle proposé……150

3.2. Principe d'apprentissage du modèle ……..... …154
4. Etude comparative pour l'optimisation des réseaux de neurones profonds.…….…….159
4.1 Critères autour de la fonction d'activité ……..…….....……159
4.2. Optimiseurs plus rapides……. ..... …..166
4.3. Discussion.....……174
4.4. Le modèle proposé….. .....……175

Conclusion générale…….…….…….. 179

Références bibliographiques.....……..…….183

Annexes……..……. 201

Abréviations :

3DD : Déviation 3D des signaux d'accélération.

ACC : Accéléromètres.

AE : Activité élémentaire

ABSC : Activité basée sur le contexte

AMB : Activités ambulatoires.

ANN : Artificial Neural Network.

ALR : Classificateur de régression logistique additive.

AR : Autoregressive Model Coefficients.

AV : Vitesse angulaire.

BN : Bayesian Network Classifier.

CART : Classification And Regression Tree.

DA : Daily activities.

DCT : Discrete Cosine Transform.

DNN : Deep Neural Network

DT : Decision Tree-Based Classifier.

DST : Dempster-Shafer Theory of Evidence

DTW : Dynamic Time Warping.

ENV : Capteurs environnementaux.

FBF : Fonction de base floue.

FD : Fonctions de domaine de fréquence.

GYR : Gyroscope.

HMM : Hidden Markov Models.

HRB : La fréquence cardiaque bat au-dessus de la fréquence cardiaque au repos.

HRM : Moniteur de fréquence cardiaque.

HW : Activités ménagères.

KNN : k-Nearest Neighbors classifier.

LAB : Expérience contrôlée en laboratoire.

LDA : Linear Discriminant Analysis.

MCMC : Méthode de Monte-Carlo par chaînes de Markov

MNL : Classificateur monolithique (indépendant du sujet).

MIL : Activités militaires.

NAT : Expérience naturaliste.

LS : Least Squares algorithm.

NB : The Naive Bayes classifier.

Abréviations

N/S : Not Specified.

NDDF : Fonction discriminante de densité normale.

PCA : Analyse en composantes principales.

PHO : Activités liées à l'utilisation du téléphone.

PR : Polynomial Regression.

PSO : Particule Swarm Optimization

RFIS : Système d'inférence floue récurrent.

SD : Sujet Évaluation dépendante.

SFFS : Sequential Forward Feature Selection.

SI : Subject Independent evaluation.

SMA : Signal Magnitude Area.

SMCRF : Semi-Markovian Conditional Random Field.

SPC : User-specific classifier (subject dependent).

SPI : Spiroergometry.

TA : Tilt Angle.

TD : Time-domain features.

TF : Transient Features.

TR : Transitions between activities.

UB : Upper body activities.

VS : Vital sign sensors

INTRODUCTION ET CONTEXTE GENERAL

Contexte de la thèse

La reconnaissance de l'activité physique joue un rôle important dans plusieurs domaines. En effet, la médecine, le bien être, le sport, l'énergie, le bâtiment, l'analyse du comportement et bien d'autres domaines s'intérèssent de plus en plus à l'état de l'activité d'un sujet [1]. Dans le domaine de la santé, par exemple, on voit de plus en plus de dispositifs médicaux, développés par des entreprises privées, se basant sur un système de reconnaissance de l'activité physique pour calculer des indicateurs de santé, telle que la fréquence cardiaque, l'analyse du pouls, la pression artérielle, etc [2]. Dans le sécteur de l'énergie et du bâtiment, la détection de l'activité physique peut renseigner les professionnels de l'énergie sur le comportement des occupants et d'agir en conséquence. En effet, en adaptant la consommation de l'énergie selon les comportements des usagers, ceci peut permettre aux consommateurs de réaliser des économies [3]. Dans le secteur du commerce et du marketing, les professionnels du sécteur s'intéressent de plus en plus à l'étude des comportements de leurs clients afin de leur proposer des services adaptés et personnalisés [4].

L'arrivée des objets connectés a donné une nouvelle ère à ces différents domaines applicatifs. De nouveaux dispositifs portables et connectés intégrant différents types de capteurs ont vu le jour. Par exemple, les smartphones permettant de passer des appels téléphoniques, de se connecter à internet et d'envoyer des textos. Les smartwatches et les bracelets connectés, permettant l'affichage et le calcul du nombre de pas effectués dans la journée, la fréquence cardiaque, etc. D'autres dispositifs commencent à voir le jour comme les semelles connectées, des vêtements connectés, des smart TV et bien d'autres objets [5]. Ces nouveaux dispositifs génèrent ainsi de nouveaux usages (par exemple, allumer son téléviseur avec son smartphone, répondre à un appel téléphonique avec sa smartwatch, etc), c'est ce que nous désignons par le terme « nouveaux usages ou usage » [6].

L'ensemble de ces objets connectés portables sont, en général, portés par les usagers. Ainsi, beaucoup de travaux les ont explorés afin de reconnaitre différents types d'activités. Les techniques ainsi utilisées reposent généralement sur l'utilisation d'approches à base d'algorithmes d'apprentissage automatique [7].

Problématique

Des études récentes ont montré l'importance de la reconnaissance de l'activité physique avec des objets connectés. Cependant, l'étude et le développement des systèmes de reconnaissance de l'activité physique font face à trois problématiques de taille. Tout d'abord les environnements des expériences réalisées ne permettent pas d'explorer efficacement les

méthodes de traitement. La majorité des bases de données disponibles ont été collectées dans des environnements contrôlés et avec des dispositifs intrusifs (capteurs placés à différents endroits du corps, détection de l'attention à la TV avec un traqueur des yeux, etc.). D'autre part, les techniques de reconnaissances de l'activité physiques employées ne sont pas adaptées aux différentes sources de données disponibles (en termes d'incertitude, d'erreurs et plus généralement d'imperfection). Enfin, l'exploration des techniques d'apprentissage en profondeur reste très limitée. La majorité des modèles nécessitent d'être optimisés afin d'avoir des résultats interprétables et surtout des modèles réutilisables pour la reconnaissance d'autres types d'activités.

Objectifs de recherche

Afin de répondre à ces problématiques, il est nécessaire de les étudier sur deux axes : un axe technique et un axe expérimental. L'axe technique se focalise sur trois principales contributions.

Premièrement, afin de reconnaitre les activités physiques, nous avons proposé de leurs attribuer deux catégories : i) les activités de types « debout », « assis », « allongé » et « marcher » détectées avec seulement des accéléromètres, des gyroscopes et des mégnétomètres, sont appelées activités élémentaires. La transformée en cosinus discrète (DCT) a été extraite comme principal descripteur à donner en entrée des différents algorithmes d'apprentissage pour classification.

Ii) Les activités détectées avec l'ensemble des capteurs disponibles sont dites activités basées sur le contexte. Une architecture à trois niveaux de granularité a été proposée afin de reconnaitre ce type d'activités.

Deuxièmement, la prise en compte de l'imperfection des données capteurs a été examinée en utilisant la Théorie de Dempster Shafer (DST). Deux principales contributions ont été approtées aux travaux existants : i) une nouvelle approche de calcul des fonctions de masse a été proposée. Ii) la fusion des fonctions de masse nécéssite un temps et complexité de calcul élévés. Nous avons proposé une approche permettant de prendre en compte cet aspect, en approximant les fonctions de masse ainsi calculées avant l'étape de la fusion.

Troisièmement, nous avons proposé une étude approfondie sur l'optimisation des architectures d'entrainements des réseaux de neurones profonds. Deux modèles d'apprentissage profond ont été proposé : i) le premier modèle prenant en entrée les descripteurs extraits auparavant et ii) un deuxième modèle prenant en entrée les données brutes des différents capteurs. La lenteur du procéssus d'entrainement ainsi que le problème d'explosion et/ou disparition de gradients ont été exploré. Une heuristique permettant d'avoir un meilleur compromis entre la vitesse d'entrainement et la précision de classification, a été implémentée.

Pour l'axe expérimental, nous avons mené une experience impliquant 33 participants dans leurs domiciles. L'ensemble des participants portait trois objets connectés : un Smartphone, une Smartwatch et une Smart TV. Ce type de dispositifs est beaucoup moins intrusif par rapport aux objets utilisés par la plupart des travaux. L'environnement de l'expérience était non controlé, les participants ne recevaient aucune instruction par rapport au déroulement de leurs activités.

Les objectifs et motivations de cette thèse se déclinent en quatre axes principaux :

1) Proposer un cadre expérimental non contrôlé pour la collecte de données des activités de 33 participants. Définir l'activité physique élémentaire et extraire la DCT comme principal descripteur de ce type d'activité.
2) Définir l'activité physique basée sur le contexte et proposer trois niveaux de granularité de reconnaissance de ce type d'activités.
3) Explorer la théorie de Dempster-Shafer afin de prendre en compte l'incertitude et l'imperfection des données capteurs.
4) Explorer l'apprentissage en profondeur afin de reconnaitre les données brutes de l'activité humaine. Discuter les techniques d'optimisation de l'entrainement et proposer une nouvelle approche permettant de conserver la rapidité d'entrainement du modèle tout en bénéficiant d'une grande précision de reconnaissance.

Organisation du manuscrit

Dans cette thèse, nous avons adopté une démarche incrémentale (comme résumé dans le Tableau 1). Le manuscrit est divisé en 5 chapitres. L'objectif du chapitre 1 est de faire un état de l'art sur deux aspects : i) aspect technologique (matériel, objets connectés, etc.) et ii) aspect traitement de données (calcul et séléction de descripteurs, apprentissage automatique, fusion de données multicapteurs, etc.). Le chapitre 2 aborde le corpus de données collectées, la notion d'activité physique élémentaire et enfin l'extraction de la DCT comme principal descripteur. Le chapitre 3, présente la notion d'activité physique basée sur le contexte ainsi que la définition des niveaux de granularité de la reconnaissance de l'activité physique basée sur le contexte. Le chapitre 4 présente une approche basée sur la théorie de Dempster-Shafer pour la prise en compte de l'imperfection des données capteurs. Le chapitre 5 présente une étude quantitative sur l'utilisation des réseaux de neurones profonds pour la reconnaissance des activités physiques. Une stratégie permettant de traiter les problèmes d'optimisation des réseaux de neurones profonds a été proposée. Enfin, nous terminons par une conclusion, les limites de la thèse, une bibliographie et une série d'annexes.

Les travaux de recherche présentés dans ce projet ont été réalisés dans le cadre du projet EGG financé par la société IofMe, dans le cadre d'une collaboration avec le CNRS. Notre objectif est de reconnaitre des activités physiques avec des objets connectés du quotidien pouvant se connecter à une plateforme nommée EGG. Cette plateforme est un thermostat connecté développé en interne dont le but est de l'utiliser afin de réguler automatiquement la température d'une maison en fonction de l'activité physique des occupants. J'ai effectué cette thèse au sein du laboratoire LIMSI-CNRS rattaché à l'école doctorale STIC de l'université de Paris-Sud/Saclay. Mes travaux de thèse m'ont permis d'entreprendre des collaborations internes au laboratoire avec Maxence Bodin, Maxime Debois et Franck Bimbard. Ainsi que des collaborations externes : Pr. Mark Schimdth (British Columbia University, Canada), Mohamed Yacine Benziani (Massachusetts Institute of Technolog, USA) et M'Hamed Hamy Temkit (Arizona State University, USA).

Tableau 1 : Organisation du manuscrit.

Chapitre	Objectifs
Chapitre 1 : Etat de l'art.	• Etudier les objets connectés et leurs architectures • Etudier les différentes techniques d'extraction de descripteurs • Etudier les différents systèmes de fusion de données multicapteurs • Etudier les différents systèmes d'apprentissage automatique (en ligne, hors ligne) • Etudier les différents algorithmes d'apprentissage automatique utilisés pour la reconnaissance de l'activité physique. • Une étude particulière sur les techniques d'optimisation des réseaux de neurones profonds pour la reconnaissance de l'activité physique.
Chapitre 2 : Reconnaissance des activités physiques élémentaires.	• Définition des activités physiques élémentaires • Etude comparative sur les principaux descripteurs utilisés. • Proposer une étude quantitative montrant l'impact du choix de la DCT comme principal descripteur.
Chapitre 3 : Reconnaissance des activités physiques basées sur le contexte.	• Définir les activités physiques basées sur le contexte • Définir une architecture de reconnaissance des activités physiques basée sur la notion de granularité. • Proposer un cadre expérimental pour reconnaitre ces différents niveaux de granularité (comportements devant la TV, attention à la TV, etc.).
Chapitre 4 : prise en compte de l'imperfection des données des activités physiques.	• Explorer la théorie de Dempster-Shafer • Proposer deux approches pour calculer les fonctions de masse • Proposer une approche afin d'approximer les fonctions de masse
Chapitre 5 : Optimisation de l'architecture	• Développer un modèle d'apprentissage profond avec sélection de descripteurs

d'entrainement des données de l'activité physique par apprentissage en profondeur.	• Développer un modèle d'apprentissage profond sans sélection de descripteurs • Etude comparative sur les techniques d'optimisation des réseaux de neurones profonds pour la reconnaissance de l'activité physique. • Proposer une nouvelle approche couplant l'algorithme du Gradient accéléré de Nestervo avec l'optimisation de l'essaim particulaire (PSO) pour l'optimisation d'un algorithme d'apprentissage des données brutes de l'activité humaine.

Chapitre 1

Etat de l'art

1. Introduction 16

2. L'internet des objets 17

2.1. Usages de l'internet des objets 18

2.2. Architecture de traitement dans l'internet des objets 20

2.3. Dispositifs commerciaux et plateformes de recherche 21

3. Procéssus de traitement 24

3.1. Extraction de descripteurs 28

3.2. Fusion de données multi capteurs 40

3.3. Apprentissage automatique pour la reconnaissance de l'activité 42

4. Limites de l'état de l'art 61

Conclusion.. … 64

1. Introduction

La définition de l'activité physique la plus répandue est celle qui a été proposée par Caspersen, en 1985. Il s'agit de « l'ensemble des mouvements corporels produits par la mise en action des muscles squelettiques et entrainant une augmentation substantielle de la dépense énergétique au-dessus du métabolisme de repos » [356]. Quant à la notion de la reconnaissance de l'activité physique, la plupart des travaux la définissent comme étant « l'ensemble des techniques et modèles statistiques ou quantitatifs permettant d'identifier avec une certaine précision l'état de l'ensemble des mouvements corporels », grace à l'utilisation de différents dispositifs et capteurs [22-63].

Depuis quelques années, nous assistons à un développement et diffusion exceptionnel de la microélectronique produisant des capteurs et des appareils mobiles et portables avec une variété de caractéristiques (différents types de capteurs, de systèmes embarqués hautes technologies, etc.). Leur puissance de calcul élevée, leur petite taille et leur faible coût ont permis aux différents utilisateurs d'interagir avec ces appareils en permanence. Ceci a fait émergé le domaine de la détection omniprésente (ou Ubiquitous Sensing, en anglais), un domaine de recherche actif dont l'objectif principal est d'extraire des connaissances à partir des données acquises par des capteurs embarqués [1]. En particulier, la reconnaissance des activités humaines est devenue une tâche d'un grand intérêt dans plusieurs domaines, principalement pour les applications médicales, militaires et de sécurité [2]. Dans le domaine de la santé, les patients atteints de diabète, d'obésité ou de maladie cardiaque doivent souvent suivre une série d'exercices bien définies dans le cadre de leur traitement [3]. Par conséquent, reconnaître leurs activités quotidiennes telles que la marche, la course ou le repos devient très utile pour fournir des informations au personnel de santé sur le comportement du patient [4]. De même, les patients atteints de pathologies mentales pourraient être surveillés pour détecter des activités anormales et ainsi prévenir des conséquences indésirables (une chute, par exemple) [5][6]. Enfin, dans les scénarios tactiques et militaires, des informations précises sur les activités des soldats ainsi que leurs emplacements et leurs conditions de santé sont très bénéfiques pour leur performance et leur sécurité [7]. De telles informations sont également utiles pour soutenir la prise de décision dans les environnements de combat et d'entraînement [8].

L'arrivée des objets connectés sur le marché a permis de faire émerger ces applications de suivi en les rendant plus flexibles. En effet, la détection des activités, qui se faisant traditionnellement, avec des capteurs fixés à différents endroits du corps, est aujourd'hui

réalisée grâce à un ou plusieurs objets connectés (smartphone, smartwatch, etc.) sans avoir à attacher les objets connectés dans différents endroits du corps [9]. Ce qui rend la tâche de détection plus flexible et pratique. Ainsi, le caractère intrusif des capteurs est largement diminué [12].

La mise en place d'un système de reconnaissance de l'activité physique exploitants des objets connectés du quotidien fait émergé plusieurs contraintes matérielles, logicielle (quel type d'objet connecté, quelle application, etc.) et de traitement (environement de l'expérience, stockage, méthode de reconnaissance à utiliser, robustesse des systèmes, etc.) [10].

Nous proposons dans ce chapitre de faire une revue de littérature sur la reconnaissance des activités physiques en utilisant des objets connectés. Nous aborderons principalement deux points essentiels :

Le premier point, concerne les notions d'internet des objets (IoT, Internet of Things en anglais) et objets connectés. Nous analysons les différentes architectures et composantes existantes (architecture, capteurs, systèmes embarqué, télécom, etc.), ainsi que les dispositifs commerciaux et les plateformes de recherche.

Le deuxième point, consiste à aborder les différentes architectures de traitement pour la reconnaissance de l'activité physique. Cette partie sera subdivisée en trois sous parties : i) les méthodes d'extractions de descripteurs utilisées. Ii) les méthodes de fusion de données multicapteurs et enfin iii) les méthodes d'apprentissage automatiques utilisées.

2. L'internet des objets

L'Internet des objets repose sur l'idée que tous les objets seront connectés à Internet et seront donc capables d'émettre de l'information et éventuellement de recevoir des commandes [11]. On parle aussi du paradigme de"ubiquitous computing", c'est à dire informatique omniprésente, ambiante, ou pervasive [12]. Ce nouveau paradigme informatique est basé non plus sur le PC, mais sur des objets quotidiens intégrant des capteurs et des capacités de communication [13]. L'internet des objets propose de créer une continuité entre le monde réel et le monde numérique : il donne une existence aux objets physiques dans le monde numérique [14].

Sur le plan fonctionnel, l'Internet des objets désigne une informatique qui se fond dans notre quotidien pour nous simplifier la vie, nous faire gagner du temps, décharger notre cerveau de la mémorisation de données logistiques (itinéraires, agenda, etc.) [15]. Il permet de créer de nouveaux usages. C'est-à-dire de nouvelles manières ou façons de se servir de l'objet connecté. Par exemple, des informations en temps réel sur la localisation de ses amis à l'aide d'un smartphone car le smartphone n'est, originellement, pas conçu pour suivre le déplacement de ces amis [16]. Il permet aussi de faire des mesures exhaustives, là où on se contentait dans le passé d'un simple panel, comme par exemple avec la mesure du trafic automobile dans les rues d'une ville [17].

L'internet des objets est largement connecté au sujet "Web Squared" [18] (le successeur du Web 2.0 selon Tim O'reilly). Il s'agit d'un Web basé moins sur les pages HTML que sur la publication de données structurées émises par les médias sociaux, les objets, les bases de données publiques (mouvement OpenData), etc. L'Internet des objets va générer une somme colossale de données qu'il faudra analyser et interpréter. Il requiert donc des architectures d'un nouveau genre appelé "BigData" capables de prendre en charge de grandes masses de données à la manière des infrastructures des grands du Web (**Figure 1**).

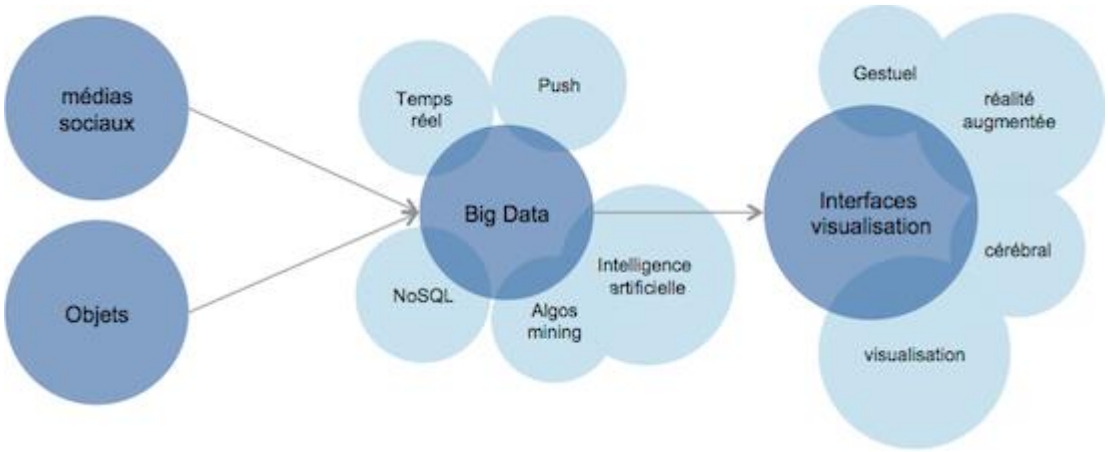


Figure 1 : Les différentes composantes de l'internet des objets (logicielles, type de données, domaine d'application, etc.) [19].

Dans la partie suivante, nous proposons d'aborder les principales catégories d'objets connectés les plus répandues. Nous aborderons ensuite les principaux types d'usage de ces objets. Ainsi, nous présenterons une vue d'ensemble de l'architctecture de traitement dans les objets connectés. Nous présenterons également une vue globale du protocole de communication, stockage et transfert de données dans les objets connectés. Enfin, nous présenterons quelques exemples d'objets connectés les plus répandus sur le marché ainsi que quelques plateformes fabriquées en laboratoires de recherche.

2.1. Les usages de l'internet des objets

L'état de l'art distingue deux grandes familles d'usage de l'Internet des objets [13] : (i) un usage sans réatroaction et ii) un usage avec réatroaction. Nous proposons de les détailler dans les parties suivantes.

2.1.1 Usages sans rétroaction

Dans ce cas, l'Internet des Objets est utilisé pour faire du suivi continu au quotidien, comme par exemple [12]:

- **consommation de l'objet** (énergie pour une voiture)
- **état de l'objet** (niveau d'huile pour une voiture)
- **mesure du contexte environnemental** (météo, trafic routier, foule dans les transports,…)
- **mesure des paramètres personnels d'un individu** ou Quantified Self (activité sportive,…)
- **activité d'une infrastructure** (réseau de télécommunication, réseau d'énergie, …)

Les utilisateurs des données peuvent être : l'usager final, un fournisseur, une administration publique, etc. Les données remontent des indicateurs permettant de réaliser [23] :

- des **tableaux de bords analytiques**, pour des fonctions décisionnelles (ex : augmenter le nombre de train sur un réseau)
- la **détection d'évènements anormaux** (ex : saturation des transports lors d'une manifestation)
- de **disposer d'informations sur un individu**, ses comportements et consommations, … (ex : les performances sportives de la semaine)
- de **ramener les informations d'un individu à une population** à des fins de comparaison pour lui-même ou pour le fournisseur de service (ex : un individu consomme plus de gaz que les usagers vivant dans un appartement de même surface)
- de **proposer** une compétition entre l'individu et ses pairs afin d'optimiser l'usage du service (ex : faire baisser sa consommation électrique pour gagner des prix)

2.1.2. Usages avec rétroaction

Dans ce cas, l'Internet des Objets permet de [24]:

- **piloter** les objets à distance, par exemple dans des applications domotiques (ex : allumer la lumière lorsque on rentre à la maison)
- **d'envoyer des notifications** au fournisseur (ex : approvisionnement d'aliments car les rayons du supermarché sont vides)
- **d'envoyer des notifications** aux usagers (ex : ralentir car l'autoroute est saturée)

Nous proposons dans la séction suivante d'exposer les différentes architectures de traitement dans les objets connectés, les protocoles de stockage, transfert et communications de données. Enfin, nous aborderons quelques exemples de dispositifs commerciaux et plateformes proposées dans le milieu accadémique.

2.2. Architecture de traitement dans les objets connectés

Les objets connectés sont des objets électroniques connectés, partageant des informations avec un ordinateur, une tablette ou un smartphone... et capables de percevoir, d'analyser et d'agir selon les contextes et notre environnement. Ce sont des dispositifs permettant de collecter, stocker, transmettre et traiter des données issues du monde physique (**Figure 2**) [25].

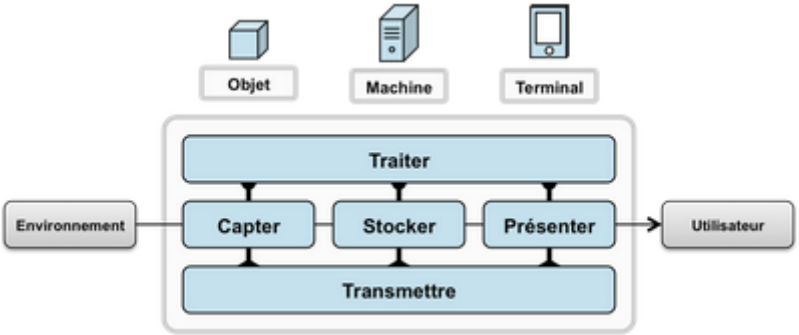


Figure 2 : Architecture de traitements dans un objet connecté [26].

Les objets dont il est question ici sont donc des sources de données, identifiés et identifiables de façon unique et ayant un lien direct ou indirect avec Internet. Selon le procéssus de collecte, stockage, transmission et traitement des données issues du monde physique, L'état de l'art a distingué deux types d'architectures [27-28]:

- **Les architectures passives** : utilisent généralement un tag (puce RFID, code barre 2D). Elles embarquent une faible capacité de stockage (de l'ordre du kilo-octet) leur permettant d'assurer un rôle d'identification. Elles peuvent parfois, dans le cas d'une puce RFID, embarquer un capteur (température, humidité) et être réinscriptibles [27].
- **Les architectures actives** : elles peuvent être équipés de plusieurs capteurs, d'une plus grande capacité de stockage, être doté d'une capacité de traitement ou encore être en mesure de communiquer sur un réseau [28].

L'architecture globale de communication entre objets connectés est représentée dans la **Figure 3**. Précisons le rôle des différents processus présentés sur ce schéma :

- Capter désigne l'action de transformer une grandeur physique analogique en un signal numérique.
- Concentrer permet d'interfacer un réseau spécialisé d'objet à un réseau IP standard (e.g. Wifi) ou des dispositifs grand public.

- Stocker qualifie le fait d'agréger des données brutes, produites en temps réel, méta taguées, arrivant de façon non prédictible.
- Enfin, présenter indique la capacité de restituer les informations de façon compréhensible par l'Homme, tout en lui offrant un moyen d'agir et/ou d'interagir.

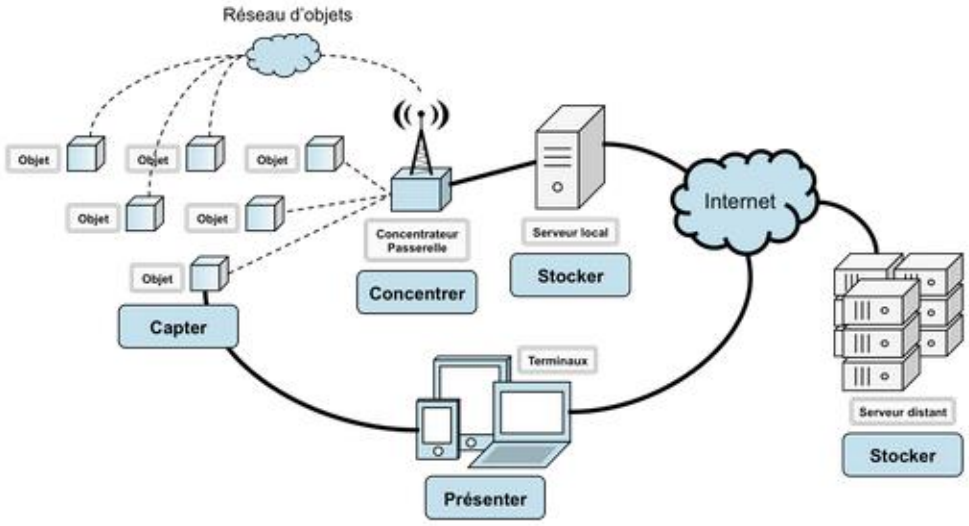


Figure 3 : Architecture de communication stockaqge et transfert de données d'objets connectés [26].

Le traitement des données est un processus qui peut intervenir à tous les niveaux de la chaîne, depuis la capture de l'information jusqu'à sa restitution. Une stratégie pertinente, et commune quand on parle d'Internet des objets, consiste à stocker l'information dans sa forme intégrale. On collecte de manière exhaustive, sans préjuger des traitements qu'on fera subir aux données. Cette stratégie est possible aujourd'hui grâce à des architectures distribuées type NoSQL, capables d'emmagasiner de grandes quantités d'information tout en offrant la possibilité de réaliser des traitements complexes en leur sein (Map/Reduce par exemple) [29].

2.3. Dispositifs commerciaux et plateformes de recherche

Il existe de nombreuses applications grand public de l'Internet des objets. En voici quelques-unes, mettant en évidence les concepts présentés dans les séctions précédentes.

- Fitbit propose de mesurer son activité physique tout au long de la journée à l'aide d'un capteur. Ce dispositif, transfert l'ensemble des informations capturées à un site web qui peut alors estimer la qualité du sommeil, la distance parcourue à pied ou encore les calories brûlées [33].

- Karotz (anciennement Nabaztag) est un objet communiquant permettant, entre autre, de reconnaître des tags RFID et d'interagir avec son propriétaire de façon sonore, gestuelle ou visuelle. Connecté à Internet, il peut diffuser des informations et être piloté depuis un smartphone [34]. **Le tableau 1** présente quelques exemples de dispositifs commerciaux ayant été utilisés pour la reconnaissance de l'activité physique.

Tableau 1 : Synthèse de quelques dispositifs commerciaux utilisés pour la reconnaissance de l'activité physique

Nom du dispositif	Type de capteurs	Type d'activités	Où est fait le calcul ?
Fitbit [33]	Accéléromètre, gyroscope, magnétomètre et cardiofréquencemètre (selon le type et la version du dispositif)	Distance parcourue, le nombre de calories brulées et estime la qualité du sommeil	Site web
Karotz [34]	accéléromètres	Gestuelle	Site web
Xiaomi [35]	Accéléromètre, gyroscope, magnétomètre.	Activités quotidiennes, nombre de pas parcourus.	En locale ou web

Dans le domaine de la reconnaissance de l'activité physique, différents objets connectés ont été utilisés. Des smartphones [35], des smartwatches [9], des bracelets connectés, des semelles connectées [36], du textile connecté [37], des lunettes connectées [38], etc. D'autres objets sont plutôt fabriqués en laboratoire, à titre d'exemple les travaux résumés dans [39] où les auteurs ont développé un textile intelligent permettant le suivi de l'activité des patients atteint d'accident vasculaire cérébral. De plus, ils ont développé un ensemble d'objets permettant le suivi de la rééducation post-AVC (par exemple, un verre connecté pour le suivi du niveau d'hydratation des patients) (**Figure 4**). L'avantage de ce type de dispositif est le faible cout des capteurs intégrés et la flexibilité et facilité d'usage. **Le tableau 2**, présente une synthèse de quelques travaux de recherche ayant développé leurs propres dispositifs.

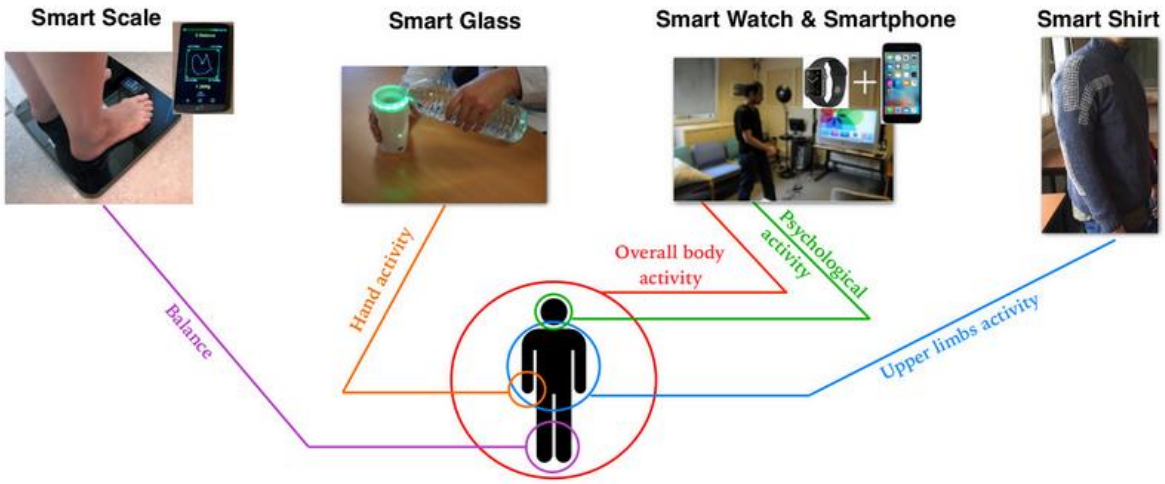


Figure 4 : Plateforme d'objets connectés pour le suivi et la rééducation post-AVC [39].

Tableau 2 : Synthèse de quelques travaux de recherche ayant développés des dispositifs connectés.

Nom du dispositif	Type de capteurs	Type d'activités	Où est fait le calcul ?
Ermes[37]	Acceléomètres, gyroscopes et magnétomètres	Debout, assis, allongé, marcher	local
eWatch[59]	accéléromètres	Debout, assis, allongé, marcher, ouvrir/fermer la porte	Local
Tapia[78]	Accéormètres, gyroscopes et magnétomètre	Debout, assis, allongé, marcher, courir, monter et déscendre les escaliers	Local vs PC
Vigilante[77]	Accéormètresn, gyroscopes et magnétomètre	Debout, assis, allongé, marcher, courir.	Local
Kao[57]	Accéormètresn, gyroscopes et magnétomètre	Marcher, courir, sauter.	Local
Brezmes[22,23]	Accéormètresn,	Debout, assis,	Pc

	gyroscopes et magnétomètre	allongé, marcher	
COSCAR[71]	Accéormètresn, gyroscopes et magnétomètre	Debout, assis, allongé, marcher, monter/déscendre les escaliers.	PC
ActiServ[26,28]	Accéormètresn, gyroscopes et magnétomètre	Debout, assis, allongé, marcher, courir, sauter, denser.	PC

Contrairement aux dispositifs du commerce, les plateformes de recherche disposent de leurs propres espaces de stockage. D'autres plateformes se connectent à distance et envoient leurs données vers des PC.

La partie suivante présente une revue de littérature sur les procéssus de traitement pour la reconnaissance de l'activité. Nous présenterons les principaux travaux sur l'extraction de descripteurs, les modèles de fusion de données multicapteurs, les méthodes d'apprentissage automatiques utilisées ainsi que le mode de fonctionnement de quelques systèmes de reconnaissance de l'activité.

3. Procéssus de traitement

Les travaux de recherche dans le domaine de la reconnaissance de l'activité physique mettent en évidence trois principales étapes: i) extraction de descripteurs, ii) fusion des données multicapteurs et iii) application des modèles d'apprentissage automatique [56-84].

Afin de permettre la reconnaissance des activités humaines, les données brutes doivent d'abord passer par le processus d'extraction des descripteurs et de fusion (que ce soit une fusion par concaténation ou bien une fusion en développant un modèle dédié). Ensuite, le modèle de reconnaissance est construit à partir de l'ensemble des instances de caractéristiques fusionnées au moyen de techniques d'apprentissage automatique. Une fois le modèle entrainé, des instances non vues (c'est-à-dire, des fenêtres temporelles) peuvent être évaluées dans le modèle de reconnaissance, ce qui donne une prédiction sur l'activité réalisée [40].
Nous proposons de commencer par étudier les différents traitements et descripteurs extraits. Nous aborderons, dans un premier temps, les descripteurs issues des données des accéléromètres, gyroscopes et magnétomètres (appelés centrales inertielles). Puis, dans un

second lieu, nous aborderons l'extraction de descripteurs dans d'autres types de de données capteurs appélés environementaux et physiologiques.

3.1. Extraction de descripteurs

Les activités humaines sont effectuées pendant des périodes relativement longues (de l'ordre de quelques secondes ou minutes, voire quelques heures) par rapport à la fréquence d'échantillonnage des capteurs (qui peut atteindre 250 Hz) [41]. En outre, un seul échantillon sur un instant de temps spécifique (par exemple, l'accélération de l'axe Y est de 2,5 g, ou la fréquence cardiaque de 130 bpm) ne fournit pas suffisamment d'informations pour décrire l'activité réalisée [42].

En pratique, les signaux ne sont généralement pas identiques, même s'ils proviennent du même sujet effectuant la même activité [43]. C'est la motivation principale pour appliquer des méthodologies d'extraction de caractéristiques (FE : Feature Extraction, en anglais) à chaque fenêtre temporelle, un filtrage des informations pertinentes permet d'obtenir des mesures quantitatives permettant de comparer les signaux. Deux approches ont été déjà proposés dans la littérature pour extraire des caractéristiques à partir de séries chronologiques: i) les approches statistiques et ii) les approches structurelles. Les approches statistiques, telles que la transformée de Fourier et la transformée en ondelettes, utilisent les caractéristiques quantitatives des données pour extraire des caractéristiques, alors que les approches structurelles tiennent compte de la relation morphologique entre les données ou les signaux [44]. Le critère de choix de l'une ou l'autre de ces méthodes dépend de la nature du signal.

La section suivante portera sur les techniques d'extraction de descripteurs les plus utilisés : i) les signaux des centrales inertielles, comme l'accélération ainsi que ii) les signaux environnementaux et ii) les signaux vitaux. Les données GPS ne sont pas prises en compte car elles sont principalement utilisées pour calculer la vitesse [45] ou inclure des connaissances sur le lieu où l'activité a été réalisée [46].

a) Descripteurs pour les signaux des centrales inertielles

Les données des centrales inertielles sont typiquement des données des accéléromètres, gyroscopes et magnétomètres. Les signaux d'accélération sont très fluctuants et oscillatoires, ce qui rend difficile la reconnaissance des modèles sous-jacents en utilisant leurs valeurs brutes. Les systèmes de reconnaissance de l'activité existants basés sur des données d'accéléromètres utilisent l'extraction de caractéristiques statistiques du domaine temporel ou fréquentiel (la DCT, FFT, etc.) [47-48]. Les caracteristiques employées sont conçues pour gérer la forte variabilité inhérente aux signaux d'accélération [49].

Le Tableau 3 résume les méthodes d'extraction de caractéristiques pour les signaux d'accélération. La définition des fonctionnalités est indiquée ci-dessous pour un signal donné Y = $\{y_1, \dots, y_n\}$ [8].

- Mesures de tendance centrale telles que la moyenne arithmétique $\overline{y}$ et le carré moyen (RMS) (Equations 2.1 et 2.2).
- Mesures de dispersion telles que l'écart type σ_y, la variance σ_y^2, et l'écart absolu moyen (MAD) (équations 2.3, 2.4 et 2.5).
- Les mesures de transformation de domaine telles que l'énergie, où F_i est le i-ème composant de la transformée de Fourrier de Y (équation 1.6).

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (1.1)$$

$$RMS\ (Y) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2} \quad (1.2)$$

$$\sigma_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2} \quad (1.3)$$

$$\sigma_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2 \quad (1.4)$$

$$MAD(y) = \sqrt{\frac{1}{n-1}}\sum_{i=1}^{n} |y_i - \overline{y}| \quad (1.5)$$

$$Energy\ (y) = \frac{\sum_{i=1}^{n} F_i^2}{n} \quad (1.6)$$

Tableau 3 : Synthèse des méthodes d'extraction de descripteurs pour les données accéléromètriques.

Type de domaine d'extraction de descripteurs	La méthode d'extraction
Domaine temporel	La moyenne, écart type, la variance, gamme interquartile (IQR), écart absolu moyen du domaine (MAD), corrélation entre les axes, L'entropie, kurtosis [18-23].
Domaine fréquentiel	La transformée de Fourrier (FT) [50], La transformée en cosinus discrète (DCT) [51].
Autres	L'analyse en composante principales (PCA) [22-26], Analyse discriminante linéaire (LDA) [52], Le modèle autorégressif (AR) et le filtre HAAR [53].

b) Descripteurs des signaux environnementaux :

Les signaux environnementaux sont principalement des signaux issus de capteurs de pression de l'air, de tempréature, d'humidité, etc. Les attributs environnementaux ainsi que les signaux d'accélération, ont été utilisés pour enrichir la connaissance du contexte. Par exemple, les valeurs de la pression de l'air et de l'intensité de la lumière sont utiles pour déterminer si l'individu est à l'extérieur ou à l'intérieur de la maison [54]. De plus, les signaux audio sont utiles pour détecter que l'utilisateur a une conversation plutôt que d'écouter de la musique [55]. **Le tableau 4** résume les méthodes d'extraction de descripteurs des attributs environnementaux les plus répandues dans l'état de l'art.

Plusieurs attributs ont été étudiés. Par exemple, l'altitude, l'humidité, la lumière et la tempréture ont été extrait à l'aide de descripteurs du domaine temporel, alors que les caracteristiques des signaux audio ont été extraits à l'aide de descripteurs du domaine fréquentiel.

Tableau 4 : Synthèse des methodes d'extraction de descripteurs en utilisant des données de capteurs environementaux.

Attribut	Features
Altitude	Domaine temporel [56]
Audio	Domaine fréquentiel [54]
Pression barométrique	Domaine temporel [57]
humidité	Domaine temporel [58]
Lumière	Domaine temporel et domaine fréquentiel [59]
température	Domaine temporel [60]

c) Descripteurs des signaux physiologiques :

Les signaux physiologiques comme L'électroencéphalographie (EEG), L'électromyogramme (EMG) et l'Électro-oculographie (EOG), ont été utilisé dans plusieurs travaux de recherche afin de reconnaitres des activités physiques basées sur le contexte (Cf. chapitre 3) [58-60]. Les signaux sont généralement prétraités afin d'extraire une série de descripteurs après avoir divisé les signaux en fenetres de tailles égales [52-60].

Diviser la série temporelle mesurée en fenêtres temporelles est une solution pratique pour entamer un problème de reconnaissance de l'activité. Par conséquent, un facteur clé est la sélection de la longueur de la fenêtre car la complexité de calcul de toute méthode de sélection de descripteurs dépend du nombre d'échantillons à sélectionner. Ainsi, avoir des fenêtres courtes peut améliorer les performances de la méthode de sélection de descripteurs. En revanche, cela entraînerait des couts de calcul plus élevés du fait que l'algorithme de

reconnaissance est appliqué d'une manière itérative à chaque sélection afin de faire des tests. En outre, les fenêtres de temps courtes peuvent ne pas fournir suffisamment d'informations pour décrire complètement l'activité réalisée. Inversement, si les fenêtres sont trop longues, il pourrait y avoir plus d'une activité dans une seule fenêtre de temps [61]. Différentes longueurs de fenêtre ont été utilisées dans la littérature: 0,08s [228], 1s [295], 1,5s [236], 3s [284], 5s [200-221], 7s [283], 12s [279], ou jusqu'à 30s [62]. Ce choix est conditionné par les activités à reconnaître et aux attributs mesurés. Le signal de fréquence cardiaque, par exemple, nécessitait des fenêtres temporelles de 30 secondes [63], à contrario, pour des activités telles que la déglutition, des fenêtres de temps de 1,5s ont été employées [64].

Les fenêtres temporelles peuvent également se chevaucher [65] [54]. L'état de l'art a montré que les fenêtres qui se chevauchent ont pour but de gérer les transitions avec plus de précision [66].

Certains descripteurs de l'ensemble de données traitées peuvent contenir des informations redondantes ou non pertinentes pouvant affecter négativement la précision de la reconnaissance de l'activité [66]. La mise en œuvre de techniques de sélection des caractéristiques les plus appropriées est une pratique suggérée pour réduire les calculs et simplifier les modèles d'apprentissage [66-67]. Le critère d'information bayésien (BIC) et la longueur minimale du descripteur (MDL) ont été largement utilisés pour les problèmes généraux d'apprentissage automatique [67]. Dans le domaine de la reconnaissance de l'activité, il existe des méthodes communes déjà utilisées comme la redondance minimale et la pertinence maximale (MRMR)[36].
Des approches itératives ont également été évaluées pour sélectionner des caractéristiques. L'état de l'art a montré que le nombre de sous-ensembles de caractéristiques est en O(2^n) [301-354], l'évaluation de tous les sous-ensembles possibles n'est généralement pas réalisable par calcul. Ainsi, des méthodes métaheuristiques telles que des algorithmes évolutifs multiobjectifs ont été utilisées pour explorer l'espace des sous-ensembles de descripteurs possibles [71].

3.2. Fusion de données multi capteurs

L'extraction de descripteurs est souvent une étape permettant d'entrainer un modèle d'apprentissage automatique pour la classification des activités physiques [66]. Certains travaux ont exploré la fusion des caracterisques extraites afin de faire une prédiction de l'activité physique [102-106].

L'une des avancées majeures dans l'exploitation des réseaux d'objets connectés est la technique de fusion des capteurs afin d'améliorer la précision de reconnaissance de l'activité

physique et réduire la large bande passante des données provenant de multiples capteurs [72]. Dans le domaine de la santé, les techniques de fusion de capteurs combinent les données capturées par les capteurs à deux niveaux : i) un niveau de fonctionnalité et ii) un niveau de décision [73]. Le niveau de fonctionnalités consiste à calculer les descripteurs extraits de chaque source de données. Le niveau de décision consiste à appliquer des critères probabilistes (prédiction, par exemple) afin de déduire l'activité physique.

Dans le domaine de la télésurveillance, par exemple, les signaux d'un capteur ECG et d'un accéléromètre triaxial ont été fusionnés au niveau des caractéristiques et des décisions afin d'implémenter une méthode d'analyse pour la surveillance à domicile des personnes âgées. Ce système était capable d'enregistrer et d'analyser en continu les données de l'ECG et de l'accéléromètre [74].

Une autre motivation pour la fusion des données capteurs est la réduction de la complexité de la logique d'application. Dans les systèmes conventionnels, les mesures de capteurs brutes ou prétraitées de manière traditionnelle atteignent directement l'application, qui doit traiter des flux de données imprécis, ambigus et incomplets. En revanche, le prétraitement par fusion de capteurs permet de normaliser l'entrée de l'application, simplifiant ainsi le développement, la maintenance et l'extensibilité de l'application [75]. Ainsi, déplacer une partie de la logique de prétraitement à un niveau inférieur à celui de l'application (typiquement, dans ou sur la surface du niveau du middleware) est à la fois rentable et souvent plus efficace. Cette « séparation des préoccupations » est, en outre, une approche qui favorise la modularité logicielle et, par conséquent, la maintenabilité [76].

La fusion de données des objets connectés doit répondre à plusieurs critères (temps de calcul, stockage des données, protocole de communication, etc.) [75]. Il est donc nécessaire d'établir une strétegie optimale pouvant répondre à ses critères et permettre un déploiement efficace de ces méthodes.

Nous allons commencer par aborder les différentes stratégies de fusion de données capteurs dans le cas des objets connectés. Puis, nous aborderons les différents niveaux de fusion les plus répandues dans l'état de lart. A ce titre, nous aborderons deux aspects : i) niveau de fusion dans les objets connectés, sans faire abstraction de la reconnaissance de l'activité et ii) nous aborderons les différents niveaux de fusion appliqués à la reconnaissance de l'activité physique.

3.2.1. Stratégies de fusion de capteurs dans le cas des objets connectés

Indépendamment du niveau d'abstraction, la fusion des capteurs peut être regroupée de manière compétitive, complémentaire et coopérative [75]. La fusion compétitive implique

l'utilisation de multiples sources d'informations équivalentes et est utilisée pour obtenir la redondance et l'auto-étalonnage. Cependant, il est très rares dans la pratique car même lorsque plusieurs capteurs équivalents sont utilisés, ils sont généralement placés à des endroits différents sur le corps et peuvent donc fournir des informations complémentaires.

Dans la fusion complémentaire, chaque capteur capture différents aspects des phénomènes surveillés et il est utilisé pour améliorer la précision et la fiabilité du système. L'information de haut niveau est obtenue par l'analyse conjointe des signaux complémentaires. Enfin, la fusion coopérative entre en jeu lorsque plusieurs signaux de capteurs sont nécessaires pour obtenir des informations qui ne pourraient pas être obtenues en regardant indépendamment l'un de ces signaux. D'après la littérature, dans le domaine des objets connectés, la fusion coopérative est la forme la plus courante de fusion de capteurs [76]. En effet, ce type de modèles a l'avantage de faire participer tous les capteurs embarqués dans les objets connectés afin d'augmenter la précision de détection. Ainsi, ce type d'approches a l'avantage d'apporter la bonne information sans contraindre la qualité du système (charge de communication, complexité de calcul, etc) [77].

En termes de traitement de données, l'état de l'art subdivise les méthodes de fusion multicapteurs en trois catégories principales: i) la fusion au niveau des données, ii) la fusion au niveau des caractéristiques et iii) la fusion au niveau décisionnel [74].

Dans cette partie, nous nous concentrerons sur l'approche du triplet (niveau de données, de caractéristiques et de décisions) basée sur le niveau de traitement.

Trois types de méthodes de fusion ont été proposés dans l'état de l'art :i) des approches de fusion de données centralisées, ii) des approches distribuées et ii) des approches hybrides. L'approche centralisée repose sur un centre de fusion dans lequel le traitement est effectué, tandis que dans l'approche distribuée, chaque capteur exécute un traitement indépendant sur ses propres donnés et transmet les résultats à un nœud de fusion où l'analyse globale est effectuée. Enfin, dans la fusion de données hybride, la collecte de données et le prétraitement sont généralement effectués avec une approche distribuée. Un nœud central est responsable de la fusion des données collectées à partir des sources distribuées et de l'exécution du calcul au niveau de la décision.

Les tableaux 5 et 6 décrivent certaines caractéristiques et techniques pertinentes de ces niveaux de fusion.

Tableau 5 : Caractéristiques de la fusion à différents niveaux.

Niveau de fusion	Le modèle	Charge de communication	Complexité de traitement	Perte d'information	Perte de performance
Niveau de données [75-81]	compétitif	Haute	Haute	non	Non
Niveau de caractéristique [82-86]	Compétitif, complémentaire, coopératif	Moyenne	Moyenne	Oui	Oui
Niveau de -- décision [87-91]	Compétitif, complémentaire, coopératif	basse	De bas en haut	Oui	Oui

La fusion au niveau des données s'avère plus couteuse en termes de charge de communication. Toutefois, ce type d'approches a la particularité de conservation de l'information. En revanche, des techniques de fusion au niveau des caracteristiques et de décision sont plus exposées à la perte d'information et ne sont pas gourmandes en termes de charge de communication.

Tableau 6 : Techniques et applications utilisées à différents niveaux de fusion.

Niveau de fusion	Utilisation	Technique
Niveau des données [76-83]	Analyse spectrale, adaptation de données, estimation des paramètres, calibration et robustesse.	Traitement de signaux, transformer les coordonnées, filtre de Kalman, moyenne pondérée, ACP.
Niveau de caractéristique [77-91]	Classification.	Machine learning (réseaux de neurones, classification, …)
Niveau de décision [91-96]	Action décisionnelle.	Systèmes experts, intelligence artificielle.

Les techniques de fusion au niveau des données sont utilisées pour diverses applications (analyse spectrale, faire de la calibration de systèmes de capteurs, etc). Les systèmes de fusion au niveau des caracteristiques sont plutôt utilisés pour faire de la classification en reconnaissance de formes. Enfin, les techniques de fusion au niveau de la décision sont

utilisées dans des systèmes décisionels informatisés. Les techniques utilsiées dans ces différents systèmes reposent essentiellement sur de la modélisation statistique [88-96].

Nous proposons dans la partie suivante de détailler les différents niveaux de fusion de données provenant des objets connectés, sans faire abstraction de l'aspect applicatif (reconnaissance de l'activité, e-santé, etc). Puis, nous aborderons ces niveaux de fusion dans le cadre de la reconnaissance de l'activité physique.

3.2.1.1. Niveaux de fusion

Si le système implique plusieurs capteurs homogènes mesurant les mêmes phénomènes physiques, les données du capteur peuvent être directement fusionnés. Les données générées à partir de sources hétérogènes ne peuvent pas être fusionnées directement et des techniques de fusion de caractéristiques ou de décision doivent être appliquées (voir Chapitre 4).

Nous allons, ci-dessous, présenter les principaux niveaux de fusion et les travaux connexes qui s'y rapportent.

a) Fusion au niveau des données :

Au niveau d'abstraction le plus bas, l'état de l'art suppose généralement que les systèmes de communication, de stockage et de traitement sont fiables. Ces systèmes sont axés sur les algorithmes de fusion qui combinent plusieurs sources homogènes de données sensorielles brutes. Le but est d'obtenir des informations plus précises et donc informatives [78].

Les technologies clés dans les études sur la fusion au niveau des données sont principalement la conception et la mise en œuvre du débruitage, de l'extraction de caractéristiques, de la classification et de la compression de données [79].

La fusion au niveau des données nécessite des approches de traitement centralisées et est utilisée en combinaison avec des réseaux de capteurs, typiquement pour la redondance et donc pour améliorer la robustesse du système [80].

Dans le domaine des objets connectés et capteurs portables, il est plus fréquemment adopté pour la séparation à la source de signaux mixtes [81]. Dans le cas d'applications de surveillance et suivi de l'activité cérébrale et cardiaque, par exemple, où les mesures ne peuvent pas être prises directement à la source et donc les données recueillies sont en réalité une combinaison de signaux physiologiques homogènes (par exemple avec les signaux EEG et ECG). Un certain nombre de paramètres sont affectés par des exigences d'application spécifiques et des choix de conception de méthodes et de techniques de fusion de capteurs

au niveau des données. Nous présentons ci-dessous un apperçu des paramètres de fusion de niveau de données (Data-Level-Fusion, [80]) les plus significatifs qui ont été identifiés.

- **Nombre de sources :** De multiples sources de données homogènes sont collectées et fusionnées au niveau des données.
- **Taux d'échantillonnage :** il peut affecter la possibilité d'effectuer une fusion de données en ligne. De plus, si les capteurs sont échantillonnés à des débits différents, les techniques de synchronisation deviennent plus complexes.
- **Synchronisation des sources :** Ces techniques nécessitent une synchronisation fine entre les sources de capteurs.
- **Périodicité des capteurs :** Les données provenant de sources multiples peuvent être collectées à un moment fixe ou régulier (détection périodique).
- **Mémoire tampon de données :** La fusion au niveau des données peut être effectuée échantillon par échantillon ou sur des tampons de données.
- **Stratégie d'agrégation :** elle détermine la manière dont les données atteignent le nœud de fusion.
- **La plateforme du nœud du capteur :** La même plate-forme de nœud ainsi que des plates-formes hétérogènes pourraient être utilisées.

b) Fusion au niveau des caractéristiques :

Les ensembles de descripteurs extraits de plusieurs sources de données (générées à partir de différents nœuds de capteurs ou d'un seul nœud équipé de plusieurs capteurs physiques) peuvent être fusionnés pour créer un nouveau vecteur de caractéristiques haute dimension qui représente l'entrée pour la classification / le modèle pour l'étape de reconnaissance [81].

A ce niveau de fusion, l'apprentissage automatique et la reconnaissance de motifs, selon le type d'application, seront appliqués aux vecteurs de caractéristiques multidimensionnels qui peuvent ensuite être fusionnés pour former des vecteurs de caractéristiques conjoints à partir desquels la classification est faite [82].

Un autre système a été proposé récemment [80]. Ce système a la pricnipale caractéristique d'etre évolutif pour la détection de mouvements et de données physiologiques d'utilisateurs possédant un capteur portable et la fusion de données au niveau décisionnel. Ce système intègre et analyse les données de la température corporelle, de l'emplacement géographique actuel, de l'électrocardiographie, de la posture et de la détection des chutes en temps réel afin de déterminer l'état de santé de l'utilisateur. D'autres travaux [83], ont présenté un système de surveillance et de réadaptation non obstructif à faible coût pour détecter la posture de

l'utilisateur et en identifiant les chutes normales. Leur système utilise la fusion au niveau des caractéristiques des données sensorielles fournies par un réseau de capteurs sans fil.

L'état de l'art a proposé une extension de la fusion à décision multi-capteurs [84]. L'objectif était d'avoir des applications de surveillance biométrique et médicale. Ce système comprend un capteur ECG, un capteur de température, un accéléromètre et fournit des modèles de rétroaction haptique distinctifs à l'état de santé de l'utilisateur. Les auteurs ont utilisé les informations biométriques collectées par les capteurs pour surveiller l'état de santé de la personne en temps réel.

Comme il n'est pas pratique de simplement concaténer les ensembles de caractéristiques, il est généralement utile d'appliquer des algorithmes de sélection de caractéristiques pour obtenir le vecteur de caractéristiques le plus significatif. Par conséquent, l'un des principaux avantages de la fusion au niveau des caractéristiques est la détection des caractéristiques corrélées générées par différents signaux de capteurs afin d'identifier un sous-ensemble d'éléments qui améliore la précision de la reconnaissance. Cependant, le principal inconvénient est que, pour trouver le sous-ensemble d'entités le plus significatif, de grands ensembles d'apprentissage sont généralement requis. Nous proposons de résumer ci-dessous les paramètres de fusion les plus importants au niveau des caractéristiques qui ont été identifiés en analysant l'état de l'art.

- **Domaine des caractéristiques (domaine des descripteurs) :** les descripteurs peuvent être extraits dans la fonction de domaine temporel, dans le domaine fréquentiel ou dans une combinaison des deux.
- **Méthode d'extraction des descripteurs :** Les descripteurs sont normalement extraits périodiquement sur des fenêtres de données fixes ou adaptables.
- **Normalisation des descripteurs :** Lorsque des valeurs de caractéristiques individuelles peuvent varier à la fois dans la portée et la distribution, la tâche de normalisation des caractéristiques doit être effectuée.
- **Sélection de descripteurs :** Il existe un large éventail de méthodes de sélection de caractéristiques dont l'objectif est d'identifier un sous-ensemble d'entités (descripteurs) optimal (ou sous-optimal).
- **La diversité de la source (capteur ou objet connecté) :** Les signaux de données peuvent provenir de capteurs homogènes ou de sources hétérogènes.
- **Taille de la fenêtre de découpage de données :** Il est souvent défini a priori, mais dans certains cas, il est adapté dynamiquement en fonction de la variance des valeurs de caractéristiques précédentes.
- **Chevauchement entre fenêtres :** choisir d'introduire un chevauchement entre les fenêtres adjacentes.

- **Modèle de traitement :** Peut être réalisée de manière centralisée sur un seul nœud de fusion ou répartie entre les nœuds de capteurs.

c) Fusion au niveau de la décision :

La fusion au niveau de la décision est le processus de sélection (ou de génération) d'une hypothèse à partir de l'ensemble des hypothèses générées par les décisions individuelles (locales et souvent plus faibles) de plusieurs capteurs [80]. La sortie de fusion au niveau de la décision est une décision unique obtenue à partir des décisions locales de capteurs multiples (homogènes ou hétérogènes). Par conséquent, il utilise les informations qui ont déjà été extraites à un certain niveau grâce à un traitement préliminaire au niveau des données ou des caractéristiques du capteur, de sorte qu'une décision de haut niveau puisse être prise.
Les méthodes de fusion de données centralisées impliquent un centre de fusion dans lequel les mesures ou les vecteurs de caractéristiques de chacun des capteurs sont traités pour former une décision globale. Dans la fusion distribuée, chaque capteur prend une décision indépendante basée sur ses propres observations et transmet ces décisions au nœud de fusion où une décision globale est prise. Comme la technique de fusion distribuée transmet moins d'informations au nœud de fusion, ses performances peuvent être dégradées par rapport à l'approche centralisée. Cette approche fournit une solution plus pratique aux systèmes en temps quasi-réel, et elle offre un avantage lorsqu'il y a une détection simultanée par les différents capteurs [80-108].
L'état de l'art a identifié une approche alternative qui utilise les avantages des techniques centralisées et distribuées est connue sous le nom de fusion hybride [99,105]. En effet, chaque capteur opère indépendamment du réseau de capteurs qui l'entoure. Ainsi, une liste de cibles candidates est générée indépendamment de chaque capteur. Cette hypothèse de détection préliminaire est une forme de décision. L'espace des hypothèses combinées se concentre uniquement sur les cibles candidates qui apparaissent dans chaque domaine individuel. Cela entraîne une réduction significative du nombre d'hypothèses pour les processus ultérieurs. Cette approche n'est applicable que si la probabilité de détection est élevée dans chaque domaine de capteur, sinon la probabilité de détection sera déterminée par le capteur le moins performant.
Nous proposons de résumer ci-dessous les principales propriétés de la fusion au niveau de la décision.

- **La méthode de fusion au niveau de la décision (Decision-Fusion):** comprend l'inférence bayésienne, la logique floue, l'analyse heuristique (le vote) et l'inférence classique.
- **Diversité de la source (capteur, objet connecté) :** Les décisions peuvent provenir de sources de capteurs homogènes ou hétérogènes.

- **Périodicité de classification :** Les résultats de classification à source unique sont générés périodiquement.
- **Modèle de traitement :** La fusion de capteurs de niveau de la décision est généralement une tâche répartie entre les nœuds de capteurs.

Le suivi de l'activité physique sensibilise les individus aux habitudes personnelles pour la santé. Par conséquent, il est crucial de pouvoir suivre avec précision les activités physiques que les gens effectuent au cours de leur vie quotidienne [85]. Les capteurs inertiels tels que les accéléromètres, les gyroscopes et les magnétomètres sont les dispositifs portables (ou wearables, en anglais) les plus couramment utilisés dans la reconnaissance de l'activité humaine [86].

Les modèles de fusion de capteurs pour la reconnaissance de l'activité physique peuvent être classés en termes de (i) niveaux d'information; (ii) les objectifs du processus de fusion; (iii) le domaine d'application; (iv) les types de capteurs utilisés; (v) la configuration de la suite de capteurs; (vi) le processus de fusion; (vii) les caractéristiques d'E / S, et ainsi de suite [87]. C'est ce que nous proposons d'aborder dans la partie suivante.

3.2.1.2. Niveaux de fusion pour la reconnaissance de l'activité

Dans cette section, nous proposons de définir les principaux travaux et caracteristiques des niveaux de fusion pour la reconnaissance de l'activité. La distinction entre les niveaux de fusion présentés dans la section 3.2.1.1 est celle-ci réside principalement dans la nature des taches à détecter. En effet, dans la section précédente, les différents modèles abordés présente une architecture globale de traitement. Ici, la sortie des modèles de fusion est la reconnaissance des activités physique. Nous proposons donc de présenter les différents niveaux de fusion pour la reconnaissance de l'activité et de discuter les principaux travaux ayant exploité ce type de modèles.

a) Fusion au niveau des données :

Dans le domaine de la fusion multi-capteurs au niveau des données, les données brutes provenant des capteurs portables sont combinées directement [88]. Cependant, ce niveau de fusion n'a pas été complètement exploré par la communauté scientifique en reconnaissance d'activités. La raison en est que les systèmes actuels de reconnaissance d'activités ne considèrent qu'un ensemble limité d'activités différentes (de 4 à 16) à distinguer. Par conséquent, la manière la plus pratique de traiter les données consiste à extraire d'abord les caractéristiques des nœuds de capteurs, puis à transmettre les caractéristiques pour la reconnaissance d'activité. Cela réduira la charge de communication et par conséquent la

consommation d'énergie du système de reconnaissance d'activité [79]. Cependant, une des limitations de cette approche est que la transmission des caractéristiques abstraites de haut niveau ne satisferait pas aux exigences d'une reconnaissance d'activité.

Une des rares études ayant explorées la fusion directe de données dans la reconnaissance d'activité est décrite dans [89]. Cette étude propose une méthode de reconnaissance d'activité quotidienne en fusionnant les données de deux capteurs inertiels attachés à un pied et à la taille du sujet, respectivement basé sur l'utilisation des modèles de type SVMs. Leur méthode a correctement reconnu les activités avec 88,1% de précision, ce qui était 12,3% plus élevé que l'utilisation d'un seul accéléromètre au niveau de la hanche.

b) Fusion au niveau des caractéristiques :

Dans le domaine de la reconnaissance d'activité, la caractéristique extraite de plusieurs capteurs avec différentes modalités telles que les accéléromètres, les gyroscopes [88], les magnétomètres [226,227], les capteurs de pression, les microphones [228], les capteur de température, les capteur de lumière [129] , pourrait être des caractéristiques du domaine temporel telles que la moyenne, l'écart-type [222-354], la variance [227-259], l'énergie, l'entropie, la corrélation entre axes, l'amplitude du signal [124] et la racine moyenne [225], les percentiles [226] ou domaine fréquentiel des caractéristiques telles que la transformée de Fourrier (FFT) et de cosinus discrète (DCT) [228], l'énergie spectrale et l'entropie [230].

Les fonctions du domaine temporel sont généralement extraites des fenêtres de tailles fixes. Ces fenetres peuvent se chevaucher ou non [221,322]. L'étape suivante consiste à sélectionner l'ensemble optimal de caractéristiques en utilisant des méthodes de sélection de caractéristiques (telles que la technique de fenêtrage, l'analyse discriminante du noyau [322], l'heuristique minimale de redondance maximale [326], la sélection de caractéristiques par corrélation [321,325]). Ensuite, les caractéristiques sélectionnées sont introduites dans des classificateurs supervisés (Naïf de Bayes [322], SVM [124,126], arbres de décision [321], réseaux de neurones [115, 354], K plus proches voisins [325]). Il existe aussi des méthodes de classification non supervisées telles que LDA, AMF [327] en fonction de la disponibilité des étiquettes (labels en anglais) pour chaque classe d'activité afin de détecter les activités comprenant des activités physiques statiques telles que s'asseoir, se tenir debout, s'allonger, regarder TV [315,322], et dynamique comme marcher et monter les escaliers [322], descendre les escaliers, courir [325], monter et descendre [321].

Le tableau 7 résume notre analyse de la littérature sur la fusion de données multicapteurs au niveau des caractéristiques dans les systèmes de reconnaissance d'activité. Par souci de clarté, nous n'avons rapporté que les études les plus représentatives (en termes de citations, de nouveauté et pertinence).

Le tableau résume les différents domaines de descripteurs utilisés (temps seul, fréquence seul ou temps fréquence), le type de descripteur extrait (moyenne, ecart type, DCT, etc), la méthode de sélection de descripteurs utilisée (fenetrage, etc), le type de sources de données (accéléromètre, audio, etc), la taille de la fenetre de données, s'il y a cheuvauchement ou pas entre les fenetres, le type de modèle utilisé (Bayésien, etc) et enfin le type d'activité physique détecté.

Tableau 7: Synthèse de la littérature sur la fusion de capteurs au niveau des caractéristiques dans les systèmes de reconnaissance d'activité.

Attribue	[222]	[224,350]	[226,326]	[225,312]	[215]
Domaine des descripteurs	Temps	Temps, fréquence	Temps et fréquence	Temps	Temps
Descripteurs	Moyenne, écart type	Moyenne, variance ou écart-type, énergie, entropie, corrélation entre axes, aire d'amplitude du signal, angle d'inclinaison, coefficients autorégressifs (AR), FFT, coefficients de transformée en cosinus discrète (DCT), différences d'altitude	Moyenne, écart type, percentiles corrélation, énergie spectrale et entropie	Valeur moyenne, écart-type, (médiane), 75e, 90e percentiles et corrélation entre les grandeurs vectorielles, fréquence dominante du signal respiratoire qu'est la fréquence respiratoire, énergie spectrale, entropie	Moyenne empirique, moyenne quadratique, écart-type, variance, écart absolu moyen, histogramme cumulatif, intervalle interquartile, taux de passage par zéro, taux de croisement moyen, nième percentile
Méthode de	Technique	Technique de	Minimal-	Sélection de	N.A

sélection de descripteurs	de fenêtrage, analyse discriminante du noyau	fenêtrage	redondance maximale-pertinence	caractéristiques basée sur la corrélation (CFS)	
Diversité des sources	accéléromètres	capteur accéléromètre, la pression et le microphone	Accéléromètres, Capteur respiratoire	un accéléromètre à deux axes, une lumière, un capteur de température et des microsecondes	N.A
Taille de la fenêtre de données	0.5 secondes	3.5 secondes		secondes	10 secondes
Fenêtre se chevauchant	Non	Oui	Non	Non	Non
Modèle de traitement	Naive Bayes	Support vector machine	Support vector machine	Decision Trees, k-Nearest Neighbor, Naive-Bayes et le Bayes Net classifier	Décision Trees, Neural Networks
Activités	assis, debout, couché, marcher, monter les escaliers	marcher, marcher sur un tapis roulant, courir, rouler sur un tapis roulant, monter et descendre les	Non spécifié	marcher, se tenir debout, s'asseoir, courir, monter et descendre les escaliers	marcher, faire du jogging, monter et descendre les escaliers, assis,

		escaliers, monter un ascenseur, monter un ascenseur, sauter, faire du vélo, faire du vétusté, regarder la télévision, passer l'aspirateur, conduire une voiture, monter un bus			debout, allongé

c) Fusion au niveau de la décision pour la reconnaissance d'activité

L'objectif principal de la fusion de décisions est d'utiliser un classifieur méta-niveau où les données du capteur sont d'abord prétraitées en extrayant les caractéristiques de celles-ci [90]. Ces caractéristiques peuvent également être des caractéristiques du domaine temporel [91], y compris la valeur moyenne, l'écart-type [229], la médiane, les percentiles [330,331], le nombre de pics, l'amplitude moyenne des pics [193] ou comme corrélation entre les axes [129] et les capteurs, l'énergie et l'entropie [130] ou la combinaison des deux [230,231]. Ensuite, les caractéristiques extraites de chaque capteur sont données en entrées des classificateurs (classificateur de base) comme k-NN [132], HMM [193, 212,333], SVM, arbre de décision [234], Naive Bayes [135], réseaux de neurones. [112,134], pour détecter leurs étiquettes de classe individuellement. Ces étiquettes de classe sont le type d'activités telles que la marche et la course, le saut [123] et les activités non liées à l'exercice, comme monter des escaliers, passer l'aspirateur, se brosser les dents [129], assis et debout [136,137], applaudir, lancer un objet, se pencher [123], travailler sur l'ordinateur, bouger [130,131], fermer et ouvrir la porte [112], se coucher, tourner à gauche et à droite [132], s'allonger [138]. Enfin, ces classes sont combinées en utilisant différentes techniques de fusion incluant l'inférence classique (vote majoritaire [112,134], régression logistique [93]), vote et ensemble [130,139], boosting [140], L'inférence bayésienne [112] et la méthode de Dempster-Shafer [123].

Les deux approches les plus communes à ce niveau de fusion sont le vote majoritaire et Bayes naïf. Dans le cas du vote majoritaire, tous les capteurs sont pondérés de manière égale sans utiliser aucune statistique antérieure [112,134]. L'étiquette finale est simplement l'étiquette de classe qui est la plus répandue parmi tous les classificateurs de base. L'approche bayésienne naïve combine le modèle probabiliste de Bayes avec une règle de décision. Une règle courante consiste à classer une instance d'entrée comme appartenant à la classe qui maximise la probabilité a posteriori [112,141].

Le tableau 9 regroupe les méthodes de fusion de décision qui ont été appliquées (théorie de Dempster Shafer, classificateur, etc), les méthodes de classification utilisées, les activités détectées et enfin le mode de traitement utilisé (centralisé ou distribué).

Tableau 8 : Paramètres de la fusion par décision dans la reconnaissance de l'activité physique.

Attribue	[223]	[229]	[330]	[293]	[234]
Méthode de fusion de décision	La théorie de Dempster-Shafer	Arbres de décision	classificateur d'ensembles	Régression logistique	Vote majoritaire
Méthode de classification de base	Régression logistique	SVM	Arbres de decision	LDA et HMM	Réseaux de neurons réccurents
Les activités détectées	sauter, plier, poinçonner, agitant, applaudir, lancer, assis, debout	debout, marcher, courir, monter et descendre les escaliers, s'asseoir, passer l'aspirateur, se brosser les dents	marcher, courir, monter et descendre les escaliers, s'asseoir, passer l'aspirateur, se brosser les dents	tâches d'assemblage d'atelier	marcher, s'asseoir, debout
Méthode de traitement	Centralisé	Centralisé	distribué	Centralisé	distribué

L'état de l'art a montré que les méthodes de fusion présentent plusieurs limitations. En effet, la plupart des méthodes emplyées souffrent de manque de précision de reconnaissance (de l'ordre de 60%-69% de précision de reconnaissance) [102-109]. Ceci étant problématique, principalement pour les applications de santé où la précision de reconnaissance reste un facteur clé de la robustesse du système [112-118].

L'état de l'art a montré l'efficacité d'avoir recours aux modèles d'apprentissage automatique [110-119]. En effet, ces techniques proposent une variété de modèles (SVM, réseaux de neurones, etc) qui ont montré leurs efficacités et leurs précisions de reconnaissance [120-129].

Nous proposons, dans la partie suivante, d'aborder quelques travaux de reconnaisance de l'activité physique ayant utilisés les méthodes d'apprentissage automatique. Nous aborderons les modèles d'apprentissage supervisés et semi supervises. Puis, nous proposons une évaluation des systèmes de reconnaissance de l'activité physique « En ligne » (au fur et à mesure de l'arrivée des données ou apprentissage à la volée) et les systèmes de reconnaissance de l'activité « Hors ligne ». Enfin, nous aborderons les modèles d'apprentissage en profondeur.

3.3. Apprentissage automatique pour la reconnaissance de l'activité

Dans un contexte d'apprentissage automatique, les modèles doivent être construits à partir d'un ensemble d'exemples donnés ou d'instances dénommées observations. Un tel ensemble d'entrées est appelé un ensemble d'apprentissage. Chaque instance est un vecteur de caractéristiques extrait de signaux dans une fenêtre temporelle. Les exemples dans l'ensemble d'apprentissage peuvent être étiquetés ou non, c'est-à-dire associés à une classe connue (par exemple, marche, course, etc.). Dans certains cas, les données d'étiquetage ne sont pas réalisables car cela peut nécessiter un expert pour examiner manuellement les exemples de données et attribuer une étiquette en fonction de leur expérience. Ce processus est généralement fastidieux, coûteux et prend beaucoup de temps dans de nombreuses applications d'exploration de données. Il existe deux approches d'apprentissage, à savoir l'apprentissage supervisé et non supervisé, qui traitent respectivement des données étiquetées et non étiquetées. Comme un système de reconnaissance d'activité humaine doit renvoyer une étiquette comme marcher, s'asseoir, courir, etc., la plupart des systèmes de reconnaissance de l'activité humaine fonctionnent de manière supervisée. En effet, il peut être très difficile de discriminer les activités dans un contexte complètement non supervisé. Certains autres systèmes fonctionnent de manière semi-supervisée, ce qui permet de ne pas étiqueter une partie des données.

Nous proposons de présenter les principaux travaux de recherche qui ont traité ces aspects de l'apprentissage.

3.3.1. Approches supervisées pour la reconnaissance de l'activité

L'étiquetage des données détectées par des individus effectuant différentes activités est une tâche relativement facile. Certains travaux proposent des systèmes permettant de stocker des données de capteurs des objets connectés dans un support non volatile pendant qu'une personne de l'équipe de recherche supervise le processus de collecte et enregistre manuellement les étiquettes d'activité [54]. D'autres disposent d'une application mobile embarquée qui permet à l'utilisateur de sélectionner l'activité à effectuer à partir d'une liste d'activités déjà établie [179]. En procédant ainsi, chaque échantillon est apparié à une étiquette d'activité, puis stocké dans un serveur.

L'apprentissage supervisé a été un domaine très productif, ce qui a amené un grand nombre d'algorithmes.

Nous proposons ci après de lister quelques classificateurs les plus répandus dans la Reconnaissance de l'activité humaine.

- L'algorithme du C4.5 est peut-être le classificateur d'arbre de décision le plus largement utilisé dans la littérature [191, 196]. Il est basé sur le concept de gain d'information pour sélectionner les attributs qui devraient être placés dans les nœuds supérieurs [193]. Les arbres de décision ont été évalués à complexité égale à O (log n) pour n attributs, et génèrent des modèles faciles à comprendre pour les humains.
- Les méthodes bayésiennes calculent les probabilités *à posteriori* pour chaque classe en utilisant des probabilités conditionnelles estimées de l'ensemble d'apprentissage. Le classificateur Bayesian Network (BN) [123] et le classificateur Naive Bayes (NB) [215], qui est un cas particulier de BN, sont les principaux modèles de cette famille de classificateurs qui ont été largement utilisés dans l'état de l'art. Un problème clé dans les réseaux bayésiens est la construction de la topologie, car il est nécessaire de faire des hypothèses sur l'indépendance entre les descripteurs. Par exemple, le classificateur NB suppose que tous les descripteurs sont conditionnellement indépendants en fonction d'une valeur de classe, mais une telle hypothèse ne tient pas dans de nombreux cas. En effet, dans la reconnaissance de l'activité physique, les signaux d'accélération sont fortement corrélés, ainsi que les signaux physiologiques tels que la fréquence cardiaque, la fréquence respiratoire et l'amplitude de l'ECG.
- Les méthodes IBL (Instance-Based Learning) [111] ont également été proposées. Ces modèles classent une instance en fonction des instances les plus similaires de l'ensemble d'apprentissage. À cette fin, ils définissent une fonction de distance pour mesurer la similarité entre chaque paire d'instances. Cela rend les classificateurs IBL très coûteux dans leur phase d'évaluation car chaque nouvelle instance à classer doit être comparée à l'ensemble d'apprentissage. Un tel coût élevé en termes de calcul et de

stockage rend habituellement les modèles IBL peu pratiques à mettre en œuvre dans un appareil mobile, comme les smartphones.

- Les (SVM) [242] et les réseaux de neurones artificiels (RNA) [352] ont également été utilisés dans les systèmes de reconnaissance de l'activité physique bien qu'ils ne fournissent pas un ensemble de règles compréhensibles pour les humains. Les SVMs s'appuient sur des fonctions de noyaux qui projettent toutes les instances vers un espace de dimension supérieur dans le but de trouver une surface de décision linéaire (c'est-à-dire, un hyperplan séparateur). Les réseaux de neurones reproduisent le comportement des neurones biologiques dans le cerveau humain, propagent les signaux d'activation et codent les connaissances dans les connexions du réseau. En outre, les réseaux de neurones artificiels sont des approximateurs de fonctions universels. Le coût de calcul élevé et la nécessité d'une grande quantité de données d'apprentissage sont deux inconvénients communs des réseaux de neurones [142,152].
- Les Classificateurs Ensembles (ensemblistes) (CE) combinent la sortie de plusieurs classeurs pour améliorer la précision de la classification. Quelques exemples peuvent être donnés comme le renforcement et l'empilement [111]. Les ensembles de classifieurs sont plus couteux en termes de calcul, car ils sont plusieurs modèles à entrainer.

Tableau 9: les principaux algorithmes utilisés dans l'état de l'art pour la reconnaissance de l'activité physique.

Type	Classifieurs
Decision tree	C4.5, ID3 [92] [93]
Bayesian	Naıve Bayes et Bayesian Networks [94]
Instance Based	k-nearest neighbors [95]
Neural Networks	Multilayer Perceptron [96]
Domain transform	Support Vector Machines ([158, 159, 160])
Fuzzy Logic	Fuzzy Basis Function, Fuzzy Inference System ([169, 135, 127])
Regression methods	MLR, ALR ([196, 179])
Markov models	Hidden Markov Models, Conditional Random Fields ([116, 110])
Classifier ensembles	Boosting and Bagging ([184, 179])

3.3.2. Approches semi-supervisées

Très peu de travaux dans le domaine de la reconnaissance de l'activité physique ayant utilisé des techniques basées sur l'apprentissage semi supervisé, c'est-à-dire en ayant une partie des données sans étiquettes [103, 102, 121, 155, 164]. En pratique, l'annotation de données peut être difficile dans certains scénarios, en particulier lorsque la granularité des activités est très élevée ou que l'utilisateur n'ait pas disposé à coopérer avec le processus de collecte de données. Comme l'apprentissage semi-supervisé est très peu utilisé dans la mise en place d'outils de reconnaissance de l'activité physique, il n'existe pas d'algorithmes ou de méthodes standard d'implmentation de ce type d'approches.

Les systèmes étudiés jusqu'à présent reposent sur de grandes quantités de données d'entraînement étiquetées. Néanmoins, dans certains cas, l'étiquetage de tous les cas peut ne pas être réalisable (vu la taille de l'ensemble de données et l'absence d'un expert pouvant faire l'étiquetage, etc). Par exemple, pour assurer une procédure de collecte de données dans un environnement naturel (où les participants auront un comportement natuel et non controlé), il est recommandé aux utilisateurs d'effectuer des activités dans un environnement non contrôlé. Si les activités changent très souvent, certaines étiquettes peuvent être manquées (car ce changement dans l'activité entraine certainement un changement d'étiquette). Ces données non marquées peuvent toujours être utiles pour former un modèle de reconnaissance au moyen d'un apprentissage semi-supervisé (meme si une partie des données ou étiquettes est perdue).

Nous proposons de résumer les principales contributions et approches de reconnaissance de l'activité physique par apprentissage semi-supervisé.

Stikic et al. [90, 89] ont développé une technique d'apprentissage semi-supervisée basée sur des multi-graphes qui propage les étiquettes à travers un graphique qui contient à la fois des données étiquetées et non-étiquetées. Chaque nœud du graphe correspond à une instance alors que chaque arête encode les similitudes entre une paire de nœuds en tant que valeur de probabilité. La topologie du graphe est donnée par l'algorithme des k-plus proches voisins dans l'espace des caractéristiques. Une matrice de probabilité Z est estimée en utilisant à la fois la distance euclidienne dans l'espace des caractéristiques et la similarité temporelle [190]. Une fois les étiquettes propagées dans tout le graphique (c'est-à-dire que toutes les instances soient étiquetées), la classification est effectuée avec un classificateur de type Support Vector Machine (SVM) qui repose sur un noyau de fonction de base radiale Gaussien. Le classificateur utilise également la matrice de probabilité Z pour introduire des connaissances sur le niveau de confiance de chaque étiquette. La précision globale atteignait 89,1% et 96,5% après l'évaluation de deux jeux de données publics et l'attribution d'étiquettes pour seulement 2,5% des données d'apprentissage.

De nombreuses recherches se sont intérréssées à une stratégie bien connue dans l'apprentissage semi-supervisé qu'est la co-formation, le co-entrainement ou le co-apprentissage (ou co-training en anglais), proposée par Blum et Mitchel [121]. Cette approche nécessite que l'ensemble d'apprentissage ait deux sous-ensembles d'attributs suffisants et redondants, une condition qui ne tient pas toujours dans un contexte de reconnaissance de l'activité physique. Guan et al. [147] ont proposé une extension de la co-formation afin de répondre à cette problématique. Le système a été testé avec dix activités et comparé à trois autres classificateurs entièrement supervisés (les k-plus proches voisins, Naive Bayes et un arbre de décision). L'amélioration maximale du taux d'erreur atteinte par le co-training était de 17% à 14% lorsque 90% des données d'apprentissage n'étaient pas étiquetées. Si 20% ou plus des données d'entraînement sont étiquetées, la différence de taux d'erreur entre le co-training et le meilleur classificateur supervisé ne dépasse pas 1,3%.

Ali et al. [17] ont mis en œuvre une technique appelée Multiple Eigenspaces (MES) basée sur l'analyse en composantes principales combinée avec des modèles de Chaines de Markov cachées. Le système est conçu pour reconnaître les gestes des doigts. Les individus portaient un gant de capteurs avec deux accéléromètres bi-axiaux échantillonnant à 50Hz. Cinq mouvements de rotation et de translation différents de la main de l'individu ont été reconnus avec jusqu'à 80% de précision. Ce système devient difficile à analyser car aucun détail n'est fourni sur la quantité de données étiquetées ni sur la procédure d'évaluation.

Huynh et al. [53] ont combiné plusieurs Eigenspaces avec des machines à vecteurs de support (SVM) pour reconnaître huit activités quotidiennes. Onze accéléromètres ont été placés sur les chevilles, les genoux, les coudes, les épaules, les poignets et la hanche des individus. La quantité de données d'entraînement étiquetées variait de 5% à 80% et la précision globale se situait entre 64% et 88%, respectivement. Leur approche a également présentée de meilleurs résultats que l'algorithme Naive Bayes supervisé, qui a été utilisé comme référence. Toutefois, les activités telles que serrer la main, monter les escaliers et descendre les escaliers étaient souvent confuses.

La plupart des approches de reconnaissance de l'activité physique basées sur l'utilisation d'objets connectés, semi-supervisées estiment d'abord les étiquettes de toutes les instances de l'ensemble d'apprentissage, puis appliquent un algorithme d'apprentissage supervisé conventionnel. Le processus d'estimation d'étiquette est souvent coûteux en termes de calcul [89-90].

Dans l'ensemble, le domaine de la reconnaissance des activités physiques dans le cas semi-supervisée n'a pas atteint sa maturité et a besoin de contributions supplémentaires pour surmonter les problèmes mentionnés.

Après avoir abordé les travaux de reconnaissance de l'activité physique du point de vue méthode d'apprentissage (supervisé, semi-supervisé), nous proposons, dans la partie

suivante, de discuter une autre dimension de la reconnaissance de l'activité physique. En effet, l'apprentissage peut etre supervisé ou semi-supervisé, il peut aussi etre « en ligne » (au fur et à mesure de l'arrivée des données) ou « hors ligne » (données stockées et figées). Ainsi, nous proposons d'aborder ces deux aspects dans la section suivante.

3.3.3. Systèmes de reconnaissance de l'activité physique

Cette partie présente l'état de l'art des systèmes de reconnaissance d'activité humaine utilisant des objets connectés portables. En outre, elle comprend également une évaluation qualitative de ces systèmes en fonction de plusieurs aspects, tels que le type d'activités reconnues, les capteurs utilisés, leur consommation d'énergie et la complexité de calcul, la précision, et bien d'autres aspects.

Trois types de capteurs ont été exploré pour déployer des systèmes de reconnaissance de l'activité physique [74] : i) des capteurs hybrides (thermique, photovoltaique, etc). ii) capteurs externes et environementaux (humidité, température, lumière, caméras, etc) et enfin iii) des capteurs portables (smartphones, smartwatch, etc).

Compte tenu des objectifs de ces travaux de thèse, nous ne couvrons dans ce manuscrit que les dispositifs portables.

Les modèles de reconnaissance de l'activité utilisant des capteurs portables possèdent, comme nous l'avons vu dans la section précédente, deux types d'entrainement : supervisé et semi-supervisé.

L'état de l'art a montré que les approches utilisant des modèles supervisés (SVM, Réseaux de neurones, etc) possèdent deux modes de réponse : un mode de réponse « En ligne » et un mode de réponse « Hors ligne » [75-89] (**Figure 5**).

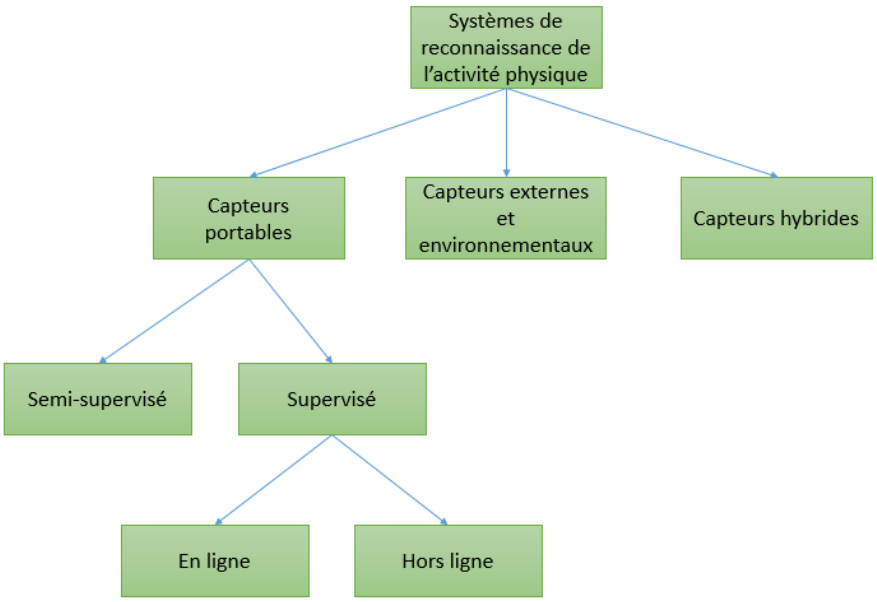


Figure 5 : architecture des systèmes de reconnaissance de l'activité physique.

Le mode de réponse « Hors ligne » fourni un retour immédiat sur les activités réalisées. Le mode de réponse « Hors ligne » a besoin de plus de temps pour reconnaître les activités en raison de demandes de calcul élevées ou il est destiné à des applications qui ne requièrent pas de retour d'informations en temps réel. Le mode de réponse « En ligne » fourni une réponse immédiate de l'activité à prédire.

À notre connaissance, les systèmes semi-supervisés actuels ont été mis en œuvre et ont tous été évalués en mode « Hors ligne ». Cette architecture de la **Figure 5** a été adoptée car les systèmes de chaque classe ont des objectifs très différents et des défis associés et devraient être évalués séparément.

A titre d'exemple, une approche totalement supervisée très précise peut ne pas fonctionner correctement dans un scénario semi-supervisé, alors qu'un système hors ligne efficace peut ne pas être capable de fonctionner en ligne en raison de contraintes de traitement. De plus, il existe un nombre important de systèmes appartenant à chaque groupe, ce qui favorise également la comparaison et l'analyse au moyen de la taxonomie proposée dans la Figure 5.

Enfin, nous présentons une évaluation qualitative des différents systèmes de reconnaissance de l'activité humaine. Elle englobe donc les aspects suivants:

- Type de capteurs et attributs mesurés.
- Dispositif d'intégration.
- Niveau d'intrusion, qui peut être faible, moyen ou élevé.
- Type de protocole de collecte de données, qui pourrait être une expérience contrôlée ou une expérience naturaliste ou non contrôlée.
- Niveau de consommation d'énergie, qui peut être faible, moyen ou élevé.
- Niveau de flexibilité de l'utilisateur, qui peut être spécifique à l'utilisateur ou monolithique.
- Méthode (s) d'extraction de caractéristiques
- Algorithme (s) d'apprentissage.
- Précision globale pour toutes les activités.

Dans la section qui suit, nous allons aborder les systèmes de reconnaissance de l'activité humaine « En ligne ».

3.3.3.1. Systèmes « En ligne »

Les applications de systèmes de reconnaissance d'activité en ligne peuvent être facilement visualisées. Dans le domaine de la santé, la surveillance et le suivi continu des patients atteints

de pathologies physiques ou mentales devient cruciale pour leur protection, leur sécurité et leur rétablissement.
Maurer et al. [59] ont présenté eWatch comme un système de reconnaissance d'activités en ligne qui intègre des capteurs et un microcontrôleur dans un appareil pouvant être porté comme montre de sport. Quatre capteurs sont inclus dans ce système, à savoir un accéléromètre, un capteur de lumière ambiante, un thermomètre et un microphone. Ce sont des capteurs passifs et, comme ils sont intégrés dans l'appareil, aucune communication sans fil n'est nécessaire; Ainsi, eWatch est très économe en énergie.
En terme de traitement algorithmique pour la reconnaissance des activités, un C4.5 a été utilisé après avoir extrait des caractéristiques dans le domaine temporel, la précision globale atteignait 92,5% pour la détection de six activités de déplacement (courir, marcher, …), bien qu'elles atteignent moins de 70% pour des activités telles que descendre et monter les escaliers. Le temps d'exécution pour l'extraction et la classification des caractéristiques est inférieur à 0,3 ms, ce qui rend le système très réactif. Cependant, dans le système eWatch, les données ont été collectées dans des conditions contrôlées, c'est-à-dire qu'un expérimentateur principal a été supervisé et a donné des directives spécifiques aux sujets sur la façon d'exécuter les activités, le nombre de sujets est aussi à discuter, alors que les profils sont diversifiés entre hommes et femmes, le nombre de sujets est 15 personnes [59].

Vigilante [43] est également présenté en tant qu'application mobile pour la reconnaissance de l'activité humaine en temps réel à l'aide d'une plateforme android. Le bracelet thoracique BioHarness BT de Zephyr a été utilisé pour mesurer l'accélération et les signaux physiologiques tels que la fréquence cardiaque, la fréquence respiratoire, l'amplitude de la forme d'onde respiratoire et la température cutanée. Des caractéristiques statistiques et temporelles ont été extraites des signaux d'accélération tandis que les caractéristiques transitoires et la régression linéaire ont été appliquées aux signaux physiologiques.
En ce qui concerne les algorithmes d'apprentissage appliqués ils ont utilisé la régression logistique et le C4.5 pour reconnaître cinq activités avec une précision globale allant jusqu'à 96,8%. L'application peut fonctionner jusqu'à 12,5 heures continues avec un temps de réponse ne dépassant pas 8% de la longueur de la fenêtre. Contrairement à d'autres approches, Vigilante a été évalué complètement en ligne pour fournir des résultats plus réalistes.
Vigilante est moyennement économe en énergie car il nécessite une communication Bluetooth permanente entre la sangle du capteur et le smartphone.

Tapia et al. [78] ont développé un système qui reconnaît 17 activités de déambulation et de gymnase telles que lever des poids, ramer, faire des pompes, etc., avec des intensités différentes (un total de 30 activités). Une étude approfondie a été réalisée, comprenant 21

participants et des études à la fois dépendantes et indépendantes du sujet. La précision moyenne de la classification était de 94,6% pour l'analyse dépendante du sujet alors que 56% de la précision était atteinte dans l'évaluation indépendante du sujet. Si les intensités ne sont pas prises en compte, la précision globale indépendante du sujet est de 80,6%. Ce système fonctionne avec un matériel très intrusif, c'est-à-dire que cinq accéléromètres ont été placés sur le bras et le poignet, la hanche, la cuisse et la cheville de l'utilisateur, ainsi qu'un moniteur de fréquence cardiaque sur la poitrine. En outre, tous ces capteurs nécessitent une communication sans fil, impliquant une consommation d'énergie élevée. Enfin, le dispositif d'intégration est un ordinateur portable, ce qui permet de meilleures capacités de traitement, mais empêche la portabilité.

Berchtold et al [17, 18,19] introduisaient ActiServ comme un système portable de reconnaissance d'activité. Le système a été implémenté sur le téléphone NeoFreeRunner. Ils utilisent un système d'inférence floue pour reconnaitre la marche et les activités en fonction des signaux fournis par l'accéléromètre du téléphone seulement. Cela fait d'ActiServ un système très efficace et portable. La précision globale varie entre 71% et 97%. Cependant, pour atteindre le niveau de précision le plus élevé, le système nécessite une durée d'exécution de l'ordre de quelques jours! Lorsque les algorithmes sont exécutés pour répondre à un temps de réponse en temps réel, la précision tombe à 71%. ActiServ peut également atteindre jusqu'à 90% après la personnalisation, en d'autres termes, une analyse dépendante du sujet. A partir des matrices de confusion calculées, l'activité étiquetée comme la marche était souvent confondue avec le cyclisme, tandis que la position debout et assise ne pouvait pas être différenciée lorsque l'orientation du téléphone portable était changée.

Riboni et al. [71] ont présenté COSAR, un système de reconnaissance d'activités basé sur le contexte utilisant un raisonnement statistique et ontologique sous la plateforme Android. Le système reconnaît les activités de déplacement, se brosser les dents, se promener et écrire sur un tableau. COSAR rassemble les données de deux accéléromètres, l'un dans le téléphone et l'autre sur le poignet de l'individu, ainsi que sur le GPS du téléphone portable. COSAR utilise le capteur GPS, il a été catalogué comme un système modérément économe en énergie. COSAR utilise un concept intéressant de matrice d'activité potentielle pour filtrer les activités en fonction de l'emplacement de l'utilisateur. Par exemple, si la personne est dans la cuisine, elle ne fait probablement pas de vélo. Une autre contribution est la classification statistique des activités avec une variante historique. Par exemple, si les prédictions pour les cinq dernières fenêtres étaient {courir, courir, marche, courir, courir}, la troisième fenêtre était probablement une erreur de classification (par exemple, l'utilisateur effectuant un mouvement atypique) et l'algorithme devrait le corriger automatiquement. Cependant, ceci introduit un délai supplémentaire à la réponse du système, en fonction du nombre de fenêtres analysées pour la variante historique. La précision globale était d'environ 93%, bien que, dans certains

cas, le fait de rester immobile ait été confondu avec l'écriture sur un tableau noir, ainsi qu'avec la marche en descendant.

Kao et al. [57] ont présenté un dispositif portable pour la détection d'activité en ligne. Un accéléromètre triaxial est placé sur le poignet de l'utilisateur, échantillonnant à 100 Hz. Ils appliquent des traitements pour extraire des caractéristiques du domaine temporel et l'analyse discriminante linéaire (LDA) pour réduire la dimension de l'espace des entités. Ensuite, un algorithme de la fonction de base floue - qui utilise des règles If-Then floues - classe alors les activités. Une précision globale de 94,71% a été atteinte pour sept activités: se brosser les dents, frapper à la porte, manger, travailler sur un PC, courir, marcher et se balancer. Le système signale un temps de réponse moyen inférieur à 10 ms, ce qui confirme sa faisabilité. Tous les calculs sont effectués dans un système embarqué qui devrait être porté par l'utilisateur en tant que périphérique supplémentaire. Ceci présente certains inconvénients en termes de portabilité, de confort et de coût. De plus, la taille de la fenêtre temporelle a été fixée à 160 ms. Compte tenu de la nature des activités reconnues, cette granularité excessive provoque des mouvements accidentels lorsque deux activités peuvent être confondus, par exemple.

Brezmes et al. [22,23] propose une application mobile pour la reconnaissance de l'activité physique sous la plate-forme Nokia. Ils ont utilisé le classificateur k-plus proches voisins (KNN), coûteux en termes de calcul et non évolutif pour les téléphones mobiles, car il nécessite la disponibilité permanente de l'ensemble d'apprentissage pour faire une nouvelle prédiction de l'activité physique- qui peut être assez volumineux - à stocker dans l'appareil. En outre, leur système exige que chaque nouvel utilisateur recueille des données d'entraînement supplémentaires afin d'obtenir des résultats plus précis.

Ermes et al. [37] ont développé un système en ligne qui a atteint une précision moyenne globale de 94%, mais ils n'ont appliqué qu'une évaluation dépendant du sujet. En outre, leurs données ont été recueillies à partir de seulement trois sujets, ce qui pose un problème de flexibilité pour soutenir les nouveaux utilisateurs.

Chaque système de reconnaissance de l'activité physique en ligne a ses propres avantages et inconvénients. Ainsi, la sélection d'une approche particulière pour une étude de cas réelle dépend des exigences de l'application et du cadre de l'expérience. Si la portabilité et l'obstruction sont les principaux problèmes, eWatch serait une option appropriée pour les activités de déambulation. Mais si un ensemble plus large d'activités doit être reconnu, le système COSAR devrait être pris en compte, bien qu'il soit couteux en autonomie d'énergie en raison de la communication Bluetooth et du capteur GPS. Le système proposé par Tapia et al. serait un meilleur choix pour surveiller les habitudes et comportements de ses utilisateurs, mais il peut être trop intrusif. Dans l'ensemble, la plupart des systèmes présentent des niveaux de précision similaires (plus de 92%), mais comme chacun fonctionne avec un

jeu de données et un ensemble d'activités spécifiques, il n'y a pas de preuves significatives pour affirmer qu'un système est plus précis que les autres. Vigilante est la seule approche qui recueille des informations sur les signaux vitaux, ce qui ouvre un spectre plus large d'applications à des fins de santé et bien-être, par exemple. En outre, COSAR et Vigilante travaillent sous la plate-forme Android - présentée comme la plate-forme de smartphones la plus vendue en 2010 par Canalys [79] - facilitant le déploiement dans les smartphones actuels. Le coût du système est également un aspect important, en particulier lorsque l'objectif de l'application est étendu à des centaines ou des milliers d'utilisateurs. Vigilante, COSAR, eWatch, et le travail de Kao et al. nécessitent du matériel spécialisé tel que des capteurs et des ordinateurs embarqués alors qu'ActiServ et le système proposé par Brezmes et al. ont seulement besoin d'un téléphone conventionnel.

Les systèmes de reconnaissance de l'activité physique « En ligne » sont en général précis (71%-98%) de précision. Les algorithmes utiliséss sont de diverses natures (arbres de décision, k plus proches voisins, etc). L'inconvénient des approches «En ligne» est la temps de traitement parfois élevé et une forte demande en énergie [78-80].

Enfin, nous présentons un aperçu des principaux travaux dans ce contexte, dans le **tableau 10.**

Tableau 10 : Synthèse des systèmes « En ligne » de reconnaissance de l'activité physique.

Ref.	Activités (#)	capteurs	Type d'expérience	énergie	flexibilité	traitement	descripteurs	apprentissage	précision
Ermes[37]	AMB(5)	Acc (poignet, genou, taille)	N/S	élevée	SPC	élevé	TD,FD	DT	94%
eWatch[59]	AMB(6)	Acc, ENV (poignet)	LAB	basse	MNL	bas	TD,FD	C45,NB	94%
Tapia[78]	EXR(30)	HRM, Acc (5 places)	LAB	élevée	Les deux	élevé	TD,FD,HB	C45,NB	86%(SD) 56%(SI) 56%(SI)
Vigilante [77]	AMB(5)	Acc, VS (poitrine)	NAT	moyenne	MNL	bas	TD,FD,TF	C45	96.8%(SD) 92.6%(SI)
Kao[57]	AMBDA(7)	Acc (poignet)	N/S	moyenne	MNL	bas	TD,LDA	FBF	94.71%
Brezmes[22,23]	AMB(5)	Acc (téléphone)	N/S	basse	SPC	élevé	TD,FD	KNN	80%
COSCAR[71]	AMB,DA(10),DA(10)	GPS, acc (watch, phone)	NAT	moyenne	MNL	moyen	TD	COSAR	93%
ActiServ[26,28]	AMB, PHO(11)	Acc (phone)	N/S	basse	SPC	élevé	$\overline{y}, \sigma_y^2$	RFIS	71%-98%

Dans la partie suivante, nous allons parcourir les différents systèmes de reconnaissance de l'activité physique de type « hors ligne ou offline » supervisés.

3.3.3.2. Systèmes « hors ligne »

Il existe des systèmes dans lesquels l'utilisateur n'a pas besoin de recevoir une rétroaction immédiate. Par exemple, les applications qui analysent les habitudes de patients en terme d'activités physiques et nutritionnelles chez les patients atteints de maladie cardiaque, de diabète ou d'obésité, ainsi que les applications qui estiment le nombre de calories brûlées après une série d'exercices physiques [15, 80]. Ils peuvent fonctionner hors ligne. Dans tous ces cas, les données recueillies peuvent être analysées quotidiennement - voire hebdomadairement - pour tirer des conclusions sur les comportements de la personne.

Le travail de Parkka et al. [64] considère sept activités: coucher, ramer, faire du vélo, rester immobile, courir, marcher et skier. Vingt-deux signaux ont été mesurés, y compris l'accélération, les signaux vitaux et les variables environnementales. Cela nécessite un certain nombre de capteurs sur la poitrine, le poignet, le doigt, le front, l'épaule, le haut du dos et l'aisselle de l'individu. Par conséquent, ce système a été catalogué comme très intrusif. Les caractéristiques du domaine temporel et du domaine fréquentiel ont été extraites de la plupart des signaux tandis qu'un dispositif de reconnaissance vocale était appliqué au signal audio [66]. Cela implique non seulement des exigences de traitement élevées, mais aussi des problèmes de confidentialité dus à l'enregistrement continu du discours de l'utilisateur. Trois méthodes de classification ont été évaluées, à savoir un arbre de décision généré automatiquement, un arbre de décision personnalisé qui introduit la connaissance du domaine et l'inspection visuelle des signaux, ainsi qu'un réseau de neurones. Les résultats indiquent que la plus grande précision était de 86%, donnée par la première méthode, bien que des activités telles que faire de l'aviron, la marche et skier n'aient pas été discriminées avec précision. Parkka et al ont mentionné que l'une des causes d'une telle erreur de classification est le manque de synchronisation entre les performances de l'activité et les annotations ou les labels.

Le travail de Bao et Intelle [16] a apporté des contributions significatives dans le domaine de la reconnaissance des activités physiques. Leur système reconnaît 20 activités, y compris la marche et les activités quotidiennes telles que le brossage de dents, l'aspiration, regarder la télévision et le travailler sur PC. Toutes les données ont été étiquetées par l'utilisateur. Cinq accéléromètres bi-axiaux ont été initialement placés sur le genou, la cheville, le bras et la hanche de l'utilisateur, mais ils ont conclu qu'avec seulement deux accéléromètres - la hanche et le poignet - la précision de reconnaissance n'est pas significativement diminuée (environ 5%). En extrayant des caractéristiques du domaine temporel et fréquentiel avec le classificateur d'arbre de décision et le C4.5, la précision globale était de 84%. Les activités ambulatoires ont été reconnues très précisément (jusqu'à 95% de précision), mais les activités telles que les étirements, le frottement, monter et descendre les escaliers étaient souvent

confuses. L'inclusion d'informations de localisation est suggérée pour surmonter ces problèmes. Une telle idée a ensuite été adoptée et mise en œuvre par d'autres systèmes [71].
Le système proposé par Khan et al. [81] non seulement reconnaît les activités de déambulation, mais aussi les transitions entre elles, par exemple, s'asseoir pour marcher, s'asseoir pour s'allonger, et ainsi de suite. Un accéléromètre a été placé sur la poitrine de l'individu, échantillonnant à 20 Hz, et envoyant des données à un ordinateur via Bluetooth pour le stockage. Trois groupes de caractéristiques ont été extraits des signaux d'accélération: (1) les coefficients du modèle autorégressif; (2) l'angle d'inclinaison (TA), défini comme l'angle entre l'axe Z positif et le vecteur gravitationnel g, ainsi que la (3) Zone d'amplitude du signal (SMA), qui est la somme des valeurs absolues de tous trois signaux.
L'analyse discriminante linéaire a été utilisée pour réduire la dimensionnalité du vecteur de caractéristiques et des activités et transitions classifiées d'un réseau de neurones avec une précision de 97,76%. Les résultats indiquent que le TA joue un rôle clé dans l'amélioration de la précision de la reconnaissance, ce qui était prévisible car les valeurs d'inclinaison du capteur sont clairement différentes pour les activités allongé et debout, étant donné que le capteur est placé sur la poitrine.
L'un des inconvénients de ce système est sa grande complexité de calcul, car il nécessite une réduction du bruit, une reconnaissance d'état, des caractéristiques temporelles et fréquentielles, un LDA et un réseau de neurones pour reconnaître les activités.
Le système proposé par Zhu et Sheng [80] utilise des modèles de Markov cachés (HMM) pour reconnaître les activités de déambulation. Deux accéléromètres, placés sur le poignet et la taille du sujet, sont connectés à un PDA via un port série. Le PDA envoie les données brutes via Bluetooth à un ordinateur pour traiter les données. Cette configuration est gênante et inconfortable parce que l'utilisateur doit porter tout le dispositif et composants de ce système ce qui peuvent interférer avec le cours normal des activités. Les caractéristiques extraites sont la vitesse angulaire et la déviation 3D des signaux d'accélération.
La classification des activités opère en deux étapes. En premier lieu, un réseau de neurones établit une discrimination entre des activités stationnaires (par exemple, assis et debout) et non stationnaires (par exemple, la marche et la course). Ensuite, un HMM reçoit la sortie de l'ANN et génère une prédiction d'activités. Un problème important lié à ce système est que toutes les données ont été collectées auprès d'un seul individu, ce qui ne permet pas de tirer des conclusions solides sur la flexibilité du système.
Nous avons également Centinela [83], un système qui combine les données d'accélération avec les signaux vitaux pour obtenir une reconnaissance précise de l'activité. Centinela reconnaît cinq activités de déplacement et inclut une plate-forme de collecte de données en temps réel, qui ne nécessite qu'un seul appareil de détection et un téléphone portable. Les caractéristiques temporelles et fréquentielles ont été extraites des signaux d'accélération

tandis que la régression polynomiale et les caractéristiques transitoires [83] sont appliquées aux signaux physiologiques. Après avoir évalué huit classificateurs différents et trois différentes tailles de fenêtre temporelle, Centinela atteint une précision globale de plus de 95%. Les résultats indiquent également que l'incorporation de signaux physiologiques permet une amélioration significative de la précision de la classification. À titre de compromis, Centinela s'appuie sur des ensembles de classificateurs représentant un coût de calcul plus élevé, et il nécessite une communication sans fil avec un capteur externe, ce qui augmente les dépenses d'énergie.

Randel et al. [69] ont introduit un système de reconnaissance des activités de la marche qui calcule le Root Mean Square (RMS) à partir des signaux d'accélération et utilise un réseau de neurones pour la classification et la reconnaissance de ces activités. La précision globale était de 95% en utilisant des données d'entrainement spécifiques à l'utilisateur, mais aucun détail n'a été fourni concernant les caractéristiques des sujets, le protocole de collecte de données et la matrice de confusion.

Le système proposé dans [49] a utilisé des filtres de type HAAR pour extraire des caractéristiques et l'algorithme C4.5 à des fins de classification. Les filtres HAAR sont destinés à réduire les calculs d'extraction de caractéristiques, par rapport aux fonctions traditionnelles TD (descripteurs temporels et FD (descripteurs fréquentiels). Cependant, l'étude n'a collecté que des données auprès de quatre individus ayant des caractéristiques physiques inconnues, ce qui pourrait être insuffisant pour permettre une reconnaissance flexible des activités sur les nouveaux utilisateurs.

He et al. [87, 84, 85] atteignaient jusqu'à 97% de précision mais ne considéraient que quatre activités: la course, l'immobilité, le saut et la marche. Ces activités sont de nature très différente, ce qui réduit considérablement le niveau d'incertitude, permettant ainsi une plus grande précision.

Chen et al. [26] ont introduit une approche Dynamic LDA intéressante pour ajouter ou supprimer des classes d'activités et des données d'entraînement en ligne, c'est-à-dire que le classificateur n'a pas besoin d'être ré-entrainé à partir de zéro. Avec un classificateur à fonction Fuzzy Basis, ils ont atteint 93% de précision pour huit déplacements et activités quotidiennes. Néanmoins, toutes les données ont été collectées à l'intérieur du laboratoire, dans des conditions totalement contrôlées.

Enfin, Vinh et al. [86] utilisaient des champs aléatoires conditionnels semi-markoviens pour reconnaître non seulement les activités mais aussi des comportements suivant des contextes tels que le dîner, le transport, le déjeuner et le bureau. Ces comportements sont composés de séquences de sous-ensembles d'activités provenant d'un ensemble de 20 activités. Leurs résultats indiquent 88,38% de précision de reconnaissance.

Contrairement aux systèmes en ligne, les systèmes de reconnaissance de l'activité physique hors-ligne ne sont pas affectés de manière significative par les problèmes de traitement et de stockage car les calculs requis sont effectués sur un serveur disposant de grandes capacités de calcul et de stockage. De plus, les dépenses d'énergie ne sont pas analysées en détail car un certain nombre de systèmes ne nécessitent ni dispositifs d'intégration ni communication sans fil, de sorte que la durée de vie de l'application ne dépend que des spécifications du capteur. Les activités ambulatoires sont reconnues très précisément dans les travaux référencés ci après [58, 81, 83]. Ces systèmes placent un accéléromètre sur la poitrine du sujet, ce qui permet d'éviter les ambiguïtés dues aux mouvements corporels brusques qui surviennent lorsque le capteur est au poignet ou à la hanche [60].

D'autres activités quotidiennes telles que s'habiller, préparer la nourriture, se doucher, travailler sur PC et utiliser un téléphone sont considérées dans [61]. Cela introduit des défis supplémentaires étant donné qu'en réalité, un individu peut utiliser son téléphone en marchant, en étant assis ou couché, présentant ainsi différents modèles d'accélération. De même, dans [16], des activités telles que manger, lire, marcher et monter des escaliers peuvent se produire simultanément, mais aucune analyse n'est présentée pour résoudre ce problème. La discrétion et le non intrusivité est une caractéristique souhaitable de tout système de reconnaissance de l'activité, mais le fait d'avoir plus de capteurs permet la reconnaissance d'un ensemble plus large d'activités.

Le système proposé par Cheng et al. [27] reconnaît les mouvements de la tête et les activités telles que la déglutition, la mastication et la parole, mais nécessite des capteurs intrusifs sur la gorge, la poitrine et le poignet. Dans les applications de santé impliquant des personnes âgées ou des patients atteints de maladies cardiaques, les capteurs intrusifs ne sont pas pratiques. Les études présentées dans [26, 27, 53] sont basées sur des données collectées dans des conditions totalement contrôlées alors que d'autres travaux [86, 67, 61, 87, 84, 85] ne spécifient pas la procédure de collecte de données, on ne sait donc pas si les conditions sont contrôlées ou pas. C'est un problème critique, car un environnement de laboratoire où l'environnement est totalement contrôlé affecte le développement normal des activités humaines [16, 75]. Le nombre de sujets joue également un rôle important dans la validité de toute étude sur les systèmes de reconnaissance des activités physiques.

Une seule personne a recueilli des données dans [88-86], tandis que dans [49], les données ont été recueillies auprès de quatre individus. La collecte de données auprès d'un petit nombre de personnes pourrait s'avérer insuffisante pour permettre une reconnaissance flexible des activités des nouveaux utilisateurs. **Le tableau 11** résume les travaux de pointe sur la reconnaissance de l'activité humaine hors ligne supervisée à l'aide de capteurs et objets connectés portables.

Tableau 11 : Synthèse des systèmes offlines de reconnaissance de l'activité physique.

ref	Activités (#)	capteurs	ID	Type d'expérience	flexibilité	descripteurs	apprentissage	précision
Bao [25]	AMB DA(20)	Acc(Poignet, Cheville Cuisse hanche	aucun	NAT	MNL	TD,FD	KNN,C45,NB	84%
Hanai [57]	AMB(5)	Acc (hanche)	pc	N/S	MNL	HAAR filters	C45	93.91%
Parkka [88]	AMB DA(9)	Acc, env, vs (22 signaux)	pc	NAT	MNL	TD,FD	DR,KNN	86%
He [60,58]	AMB(4)	Acc (poches)	pc	N/S	MNL	AR,DCT,PCA	SVM	92%
Zhu [116]	AMB TR(12)	Acc Poignet Taille	pc	N/S	MNL	AV,3DD	HMM	90%
Altun [22]	AMB(19)	Acc, Gyr Poitrine Pied poignet	pc	N/S	SPC	PCA,SFFS	BN,LS,KNN,DTW	87%-99%
Cheng [36]	UB(11)	Acc Hanche poignet	aucun	NAT	MNL	TD	LDA	77%
McGlynn [83]	DA(5)	Acc jacket	N/S	LAB	MNL	DTW	NB,HMM	84.3%
Pham [92]	AMB DA(4)	Acc Poitrine hanche	N/S	N/S	SPC	Energie relative	ANN	97%
Vinh [110]	AMB DA(21)	Acc, VS poitrine	smartphone	N/S	Les deux	TD	Boosting	88.38%
Centinela [79]	AMB(5)	Acc poitrine	pc	N/S	N/S	TD,FD,PR,TF	FBF	95.79%
Khan [71]	AMB TR(15)	Acc,spi	tablette	N/S	spc	AR,SMA,TA,LDA	CART	97.97%
Jatoba [66]	AMB(6)	Acc 2 poignets	N/S	NAT	MNL	TD,FD	SVM	86%
Chen [35]	AMB DA HW(8)	Acc 6 places	pc	NAT	SPC	TF,FD	SMCRF	93%

Minnen [84]	AMB, MIL(14)	Acc poignet	pc	LAB	MNL	TD,FD	NB,HMM	90%

Parmi les techniques d'apprentissage automatique qui ont connu un succès ces dernières années, les approches à base d'apprentissage en profondeur. Nous proposons, dans la partie suivante, de faire une revue synthétique de littérature sur les principaux travaux dans la reconnaissance de l'activité physique.

3.3.4. Apprentissage en profondeur de l'activité physique

L'exploration des réseaux de neurones ordinaires a montré ses limites. En effet, l'apprentissage de ce type de modèles nécéssite une étape d'extraction, séléction et réduction de la dimensionnalité du vecteur des descripteurs, très fastidieuse à mettre en place [160-163]. Le développement de la recherche dans le domaine de l'apprentissage automatique a fait émerger de nouveaux modèles : les réseaux de neurones profonds. La principale caractéristique de ce type d'approches est leur capacité à apprendre une grande quantité de données avec une grande précision. De plus, dans certains cas, ce type de modèle ne nécéssite pas d'extraction ni de séléction de descripteurs [161-170].

L'un des modèles qui ont été implémenté récemment est le modèle des réseaux de neurones convolutionnels. En effet, dans [155] un réseau de neurones convolutionnel profond (CONVNET) a été proposé pour réaliser un système de reconnaissance de l'activité humaine efficace en utilisant des capteurs de smartphones en exploitant les caractéristiques inhérentes aux activités et aux signaux de séries temporelles 1D, tout en fournissant un moyen d'adaptation automatique et adaptable aux données et extraire des fonctionnalités robustes à partir de données brutes. Les expériences ont montré que les CONVNETS dérivent effectivement des caractéristiques pertinentes et plus complexes avec chaque couche supplémentaire, bien que la différence de niveau de complexité des caractéristiques diminue avec chaque couche supplémentaire. Une période plus longue de corrélation locale temporelle peut être exploitée (1 × 9-1 × 14) et une taille de regroupement faible (1 × 2-1 × 3) s'avère bénéfique. Ces modèles ont également obtenu une classification presque parfaite sur les activités non stationnaires comme courir, en particulier celles très similaires qui étaient auparavant perçues comme très difficiles à classer. Enfin, les CONVNETS s'avèrent plus précis que les autres techniques d'exploration de données de pointe dans les systèmes de reconnaissance de l'activité pour l'ensemble de données de référence recueillies auprès de 30 sujets volontaires, atteignant une performance globale de 94,79% sur l'ensemble de test avec des données brutes et 95,75% avec des informations supplémentaires (activités basées sur le contexte) [155].

La recherche actuelle suggère que les réseaux de neurones convolutionnels profonds sont adaptés pour automatiser l'extraction de caractéristiques à partir d'entrées de capteurs brutes. Cependant, les activités humaines sont constituées de séquences complexes de mouvements moteurs, et la capture de cette dynamique temporelle est fondamentale pour un système réussi de reconnaissance de l'activité. Basé sur le succès récent des réseaux neuronaux récurrents pour les domaines de séries temporelles, les auteurs d'un récent travail [156] ont proposé un cadre générique d'apprentissage profond pour la reconnaissance d'activité basée sur des unités convolutives et récurrentes LSTM, qui: (i) est adapté aux capteurs portables multimodaux; (ii) peut effectuer la fusion de capteurs naturellement; (iii) n'exige pas de connaissances spécialisées dans la conception de caractéristiques; et (iv) modélise explicitement la dynamique temporelle des activations de caractéristiques. Ils ont évalué leur approche sur deux ensembles de données, dont l'un a été utilisé dans un défi de reconnaissance des activités publiques. Leurs résultats ont montré que leur approche dépasse de près de 4% à 9% en moyenne les réseaux concurrents non récurrents.

D'autres études se sont contentés d'utiliser qu'un seul accéléromètre et appliqué un réseau de neurones convolutifs (CNN) aux données brutes. Par exemple, Yang et al [157], ont développé une approche pour modifier le noyau de convolution sans modifier la taille des données d'apprentissage pour adapter les caractéristiques des signaux d'accélération triaxiaux. Le grand ensemble de données qu'ils ont construit se compose de 31688 échantillons provenant de huit activités typiques. Les résultats de l'expérience montrent que le CNN fonctionne bien, ce qui peut atteindre une précision moyenne de 93,8% sans aucune extraction préalable de caractéristiques [157].

D'autres approches [160] ont proposé de combiner la puissance du traitement d'images (de la vision par ordinateur) et les signaux portables afin d'analyser des images prises à partir d'une caméra portable égocentrique passive avec les informations contextuelles, telles que l'heure et le jour de la semaine, pour apprendre et prédire les activités quotidiennes d'un individu. Ils ont recueilli un ensemble de données de 40103 images égocentriques sur une période de six mois avec 19 classes d'activités et démontré l'avantage de techniques d'apprentissage en profondeur pour apprendre et prédire les activités quotidiennes. La classification est effectuée à l'aide d'un réseau de neurones convolutionnels (CNN) avec une méthode de classification qu'ils ont introduit appelée un ensemble de fusion tardive. Ceci incorpore des informations contextuelles pertinentes et augmente la précision de classification pour atteindre une précision globale de 83,07% en prédisant l'activité d'une personne dans les 19 classes d'activité.

D’autres travaux ont été développés dans le domaine de la gestuelle pour reconnaitre des mouvements de la main. Comme par exemple dans [161] où les auteurs ont proposé un algorithme pour la reconnaissance du geste de la main des conducteurs en utilisant des réseaux de neurones convolutifs 3D. Ils ont combiné des informations provenant de multiples échelles spatiales pour la prédiction finale. Ils ont aussi utilisé l'augmentation de données spatiotemporelles pour un entrainement plus efficace et pour réduire les surapprentissages potentiels. Leur méthode atteint un taux de classification correct de 77,5% sur le jeu de données de défi VIVA.

Dans le domaine de la vision par ordinateur, dans [162-164] les auteurs ont proposé un modèle basé sur un algorithme d'apprentissage en profondeur, qui permet de reconnaître l'activité physique humaine basée sur les données du squelette du corps humain à partir du capteur de Microsoft Kinect. Ce modèle utilise les données de squelettes humains de l'ensemble de données CAD-60 pour reconnaître l'activité physique humaine sans utiliser aucune connaissance préalable. Un réseau neuronal convolutif est utilisé pour l'extraction de caractéristiques, et un perceptron multicouche est utilisé comme classificateur. Le modèle peut reconnaître douze types d'activités avec une précision de 81,8%.

Toutes ces approches ont permi une avancée majeure dans le domaine de la reconnaissance de l’activité physique. La précision de reconnaissance atteinte est, dans la plupart des travaux, élevée, dépassant ainsi les 80%.

L’état de l’art a montré que, malgré la bonne précision de classification que fournissent ce type de modèles, l’entrainement et le déploiement des modèles de réseaux de neurones profonds présentent plusieurs limitations. En effet, plusieurs problèmes sont prosées, notamment l’optimisation du procéssus d’entrainement afin de rendre les modèles plus réactifs tout en gardant une bonne précision de reconnaissance. Nous propsons d’aborder ce point dans la partie suivante.

Optimisation du processus d’apprentissage de l’activité physique

Dans la partie précédente, nous avons présenté quelques travaux remarquables de l’utilisation des réseaux de neurones profonds pour la reconnaissance de l’activité physique. Ces modèles ont certainement donné satisfaction, en termes de précision de classification. Cependant, les récents travaux ont signalé certaines limitations, notamment la lenteur du processus d’apprentissage qui se complexifie de plus en plus quand la base de données des activités physiques grandie.

Dans cette partie, nous présentons un aperçu des travaux ayant pris en compte ces aspects et lister leurs solutions pour y remédier.

L'ensemble des méthodes d'optimisation d'un processus de reconnaissance de l'activité physique est résumé dans le **tableau 12** suivant :

Tableau 12: Synthèse de quelques travaux sur l'optimisation du processus de reconnaissance de l'activité.

Ref.	#Activités	type d'optimisation	Réduction	Technique d'apprentissage
Ermes[37]	4	Sélection de descripteurs	Temps de calcul	Fusion prédictive
eWatch[59]	12	ACP	Temps de calcul	Arbres de décision
Tapia[78]	15	Réduction de la dimension,	Temps	SVM
Vigilante[77]	24	Non	Non	Réseau de neurones profond
Kao[57]	9	ACP	Temps de calcul	Fusion
Brezmes[22,23]	15	ACP, sélection de descripteurs	Temps de calcul	SVM, Naif bayes
COSCAR[71]	10	Sélection de descripteurs	Temps/espace	Réseau de neurones
ActiServ[26,28]	7	Réduction de la dimension	Temps/espace	Réseau de neurones

En résumé :

- Les méthodes d'apprentissage en profondeur, malgré leur précision, restent très lentes en termes de temps d'apprentissage.
- Les approches proposonées dans la littéarture ne prennent pas suffisamment en compte l'optimisation des systèmes de reconnaissance de l'activité. Le centre d'interet pour la plupart des travaux reste la précision de détection.

4. Limites de l'état de l'art

Reconnaissance de l'activité physique élémentaire et basée sur le contexte

La comparaison quantitative des approches de reconnaissance de l'activité a été entravée du fait que chaque système fonctionne avec un ensemble de données différent. Dans les

domaines de recherche tels que l'exploration de données, il existe des ensembles de données standards pour valider l'efficacité d'une nouvelle méthode. Les approches de reconnaissance de l'activité physique ne reposent pas sur ce principe. Chaque groupe de recherche recueille des données auprès de différentes personnes, utilise un ensemble d'activités différent et une méthodologie d'évaluation différente. Dans ce sens, les ensembles de données devraient être publiquement ouverts à la communauté scientifique afin d'être utilisés comme points de référence pour évaluer de nouvelles approches. Plusieurs universités et institutions ont publié leurs ensembles de données dans [12, 10, 11, 9]. Un autre ensemble de données est fourni par le Défi de la reconnaissance des activités de 2011 [1], auquel les chercheurs du monde entier ont été invités à participer.

Les participants effectuent certainement des activités d'une manière différente en raison de caractéristiques physiques particulières. Ainsi, les signaux d'accélération mesurés d'un enfant par rapport à une personne âgée devraient être très différents. Un modèle de reconnaissance de l'activité humaine peut être monolithique ou spécifique à l'utilisateur, chacun ayant ses propres avantages et inconvénients. Un terrain d'entente pour tirer le meilleur parti des deux mondes pourrait créer des classificateurs spécifiques au groupe, regroupant des individus ayant des caractéristiques similaires telles que l'âge, le poids, le sexe et les conditions de santé, entre autres. Ensuite, notre hypothèse est qu'un système de reconnaissance de l'activité serait plus efficace si l'on avait un modèle de reconnaissance pour les jeunes hommes en surpoids, un pour les enfants normaux, un autre pour les femmes âgées, et ainsi de suite.

L'étude de la littérature a montré qu'il y a principalement deux types d'activités : (i) activités ne dépendant pas du contexte dans lequel se déroule, et (ii) des activités basées sur le contexte. La plupart des travaux ne feront pas cette distinction au niveau conceptuel. C'est-à-dire qu'ils ne prennent pas en compte cette catégorisation pour mettre en œuvre une stratégie de traitement. En effet, les activités basées sur le contexte apportent, en plus de ce qu'apportent les activités élémentaires, de l'information sur le contexte sous forme de données capteurs, interaction avec les objets, etc. Ceci peut influencer considérablement les méthodes de traitement telles que l'extraction de descripteurs, et peut, dans d'autres cas, influencer la compréhesion de l'activité.

Les travaux antérieurs ont non seulement estimé les activités mais aussi les comportements [86]. Sur la base de cette information, le système pourrait prédire ce que l'utilisateur est sur le point de faire. Cela devient particulièrement utile pour certaines applications telles que celles basées sur des publicités. Par exemple, si l'utilisateur va déjeuner, il peut recevoir de la publicité sur les restaurants à proximité.

La reconnaissance des activités humaines a été en quelque sorte individualisée, c'est-à-dire que la majorité des systèmes prédisent des activités chez un seul utilisateur. Bien que

l'information provenant des réseaux sociaux se soit avérée efficace pour reconnaître les comportements humains [153], la reconnaissance des modèles d'activité collective peut être poussée plus loin. Si nous pouvions rassembler les modèles d'activité d'un échantillon significatif de personnes dans certaines régions (par exemple, une ville, un état ou un pays), ces informations pourraient être utilisées pour estimer les niveaux de sédentarité, les habitudes d'exercice et même les conditions de santé d'une cible. En outre, ce type d'application participative centrée sur l'être humain n'exigerait pas de méthode d'incitation économique. Les utilisateurs seraient prêts à participer au système tant qu'ils recevraient des informations sur leur état de santé et leurs performances physiques, par exemple. De telles données provenant de milliers ou de millions d'utilisateurs peuvent également être utilisées pour alimenter des algorithmes de classification, améliorant ainsi leur précision globale.

L'hypothèse qu'un individu n'effectue qu'une seule activité à la fois est vraie pour les activités de base (qu'on appelle aussi activités élémentaires, par exemple, marcher, courir, s'allonger, etc.). En général, les activités humaines se chevauchent et sont simultanées. Une personne pourrait marcher en se brossant les dents ou en regardant la télé pendant le déjeuner. Étant donné que seulement quelques travaux ont été rapportés dans ce domaine, nous prévoyons de grandes opportunités de recherche dans ce domaine (Helaoui et al. [50]).

Les activités explorées dans ce chapitre sont plutôt simples. En fait, beaucoup d'entre elles pourraient faire partie d'habitudes ou de comportements plus complexes. Imaginant, par exemple, le problème de la reconnaissance automatique lorsqu'un utilisateur joue au tennis. Une telle activité est composée de plusieurs instances de marche, de course, etc., entre autres, avec une certaine séquence logique et une certaine durée. La reconnaissance de ces activités composites à partir d'un ensemble d'activités atomiques enrichirait sûrement la connaissance du contexte mais, en même temps, apporterait une incertitude supplémentaire. Blanke et al. [20] donnent un aperçu de ce sujet et proposent une solution à travers plusieurs couches d'inférence.

En termes de sélection de description, très peu de travaux ont exploré la DCT comme principal descripteur. L'exploration de ce type de descripteur a donné de très bons résultats sur des cas mono-capteur, alors que dans la vie courante, les objets connectés embarquent une multitude de capteurs. Il serait donc intéressant d'explorer cette nouvelle voie de recherche, combinant tous les aspects cités auparavant (contexte, etc.).

Prise en compte de l'imperfection des données des objets connectés

Les données issuent des capteurs embarqués dans les objets connectés sont par définition imparfaites. Les modèles d'apprentissage automatique utilisés dans la plupart des travaux ne prennent pas en compte cette contrainte dans le procéssus de reconnaissance de l'activité physique.

Des techniques telles que la théorie des croyances de Dempster-Shafer ont été utilisées. Les résultats de reconnaissance étaient sensibles au calcul des fonctions de masse. En effet, il n'y a pas de méthode générique permettant de calculer ces fonctions de masse. Celles-ci peuvent etre effectués de diverses manières : statistiques, vote majoritaire, etc. Les fonctions de masse calculées seront fusionnées par la règle de combinaison de Dempster-Shafer. La fusion des fonctions de masse nécéssite un temps et complexité de calcul proportionnel à la taille du cadre de discernement. Ceci peut limiter grandement l'application de la théorie de Dempster-Shafer. Ainsi, la plupart des travaux n'ont pas suffisamment exploré ces différentes contraintes.

Apprentissage en profondeur pour la reconnaissance de l'activité humaine

Les techniques d'apprentissage en profondeur ont permis de résoudre les limites des modèles d'apprentissage automatiques classiques. En effet, la plupart des modèles utilisés nécéssitent une étape d'extraction, séléction et réduction du vecteur de descripteurs. Des travaux récents ont montré qu'il était possible d'utiliser des modèles d'apprentissage profonds afin de répondre à ces limitations en entrainant ces modèles avec des données brutes. En revanche, plusieurs limites ont été soulevées, comme par exemple l'absence d'une stratégie claire permettant de lutter contre le problème de disparition-explosion de Gradients et la lenteur d'entrainement de ce type de modèles.

La plupart des travaux sur la reconnaissance de l'activité physique n'ont pas exploré suffisamment ce type de contraintes. Il serait intéressant de développer des techniques permettant d'optimiser le processus d'apprentissage en termes de temps de calcul, précision vs gain en temps d'apprentissage vs apprentissage des données brutes, etc.

Conclusion

Dans ce chapitre, nous avons exposé quelques travaux relatifs à la reconnaissance de l'activité en utilisant des objets connectés. Nous avons commencé par discuter les différents objets connectés, leurs architectures, leurs natures et leurs procéssus de stockage et traitement de données.

L'état de l'art était axé sur trois éléments essentiels : 1) l'extraction de descripteurs, dans laquelle nous avons abordé les différents descripteurs extraits et discuté leurs avantages et inconvénients. Nous avons montré que la plupart des travaux préfèrent varier les types de descripteurs à extraire.

Les systèmes de reconnaissance de l'activité physiques (système supervisé, semi-supervisé, en ligne, hors ligne) ont été abordé. Nous avons soulevé les avantages de chacun des systèmes (En ligne, Hors ligne). Nous avons conclus que les systèmes «En ligne » souffrent

de problème d'économie d'énergie et la charge de calcul élevé. Alors que les systèmes « Hors ligne » sont plus précis et souffrent aussi de lenteurs (selon le type d'apprentissage à effectuer).

2) La fusion de données multi capteurs a aussi été abordée. Nous avons étudié les différentes méthodes proposées. Nous avons discuté les avantages et inconvénients de chaque approche. Nous avons vu que les méthodes de fusion peuvent etre adaptées à la reconnaissance de l'activité principalement dans les problèmes de suivi des patients en explorant des capteurs fixés à différents endroits d'une pièce ou maison.

Enfin 3) nous avons discuté les différentes approches d'apprentissage profond pour la reconnaissance de l'activité. Nous avons vu que la plupart des travaux utilisant ce type d'approches n'exploraient pas suffisamment certains aspect (notamment l'explosion/Disparition de Gradients et les problèmes liés à l'optimisation du processus d'apprentissage et de convergence).

Les chapitres suivants vont apporter des éléments de réponses aux différentes interrogations soulevés, notamment :

Dans le chapitre 2, nous présenterons le nouveau corpus de données ainsi que le concept d'activité physique élémentaire. Nous proposons une étude comparative afin d'extraire le meilleur descripteur pour la reconnaissance des activités physiques élémentaires. Dans le chapitre 3, nous présenterons une nouvelle architecture de reconnaissance des activités basées sur le contexte. Cette architecture sera basée sur une notion de granularité. Le chapitre 4, aborde le problème de l'imperfection des données capteurs. Le chapitre 5 aborde le problème d'optimisation de l'entrainement d'un réseau de neurones profond.

Chapitre 2

Reconnaissance de l'activité physique élémentaire

1. Introduction 67

2. Vue d'ensemble des traitements.....68

3. Création du corpus 69

3.1. Description de l'expérience.....69

3.2. Pré-analyse du corpus.....71

4. Extraction des descripteurs 73

4.1. Découpage et segmentation des signaux 75

4.2. Taille d'une coupe 76

4.3. Résultats Globaux 76

4.4. DCT comme principal descripteur.....82

4.5. Influence chevauchement entre activités élémentaires.....89

Conclusion…….. 96

Chapitre 2

1. Introduction

L'état de l'art a montré que les activités physiques sont subdivisibles en plusieurs petites activités ou micro activités [166-178]. Par exemple, l'activité « jouer au tennis » se compose de plusieurs activités, dites basiques ou élémentaires, de type « courir », de plusieurs séquences d'activités de type « debout », etc. Les activités physiques sont alors ségmentées en plusieurs activités afin de faciliter les traitements par les différents algorithmes. Les types d'activités résultantes de cette opréation de ségmentation sont en général les suivantes : « debout », « assis », « courir », « monter/descendre les escaliers », « marcher » et « allongé » [201-203]. L'exploration de ce type d'approches pour la reconnaissaance d'autres activités plus complexes reste très limitée.

La multiplication d'objets connectés portables et les données capteurs qu'ils génèrent ne nous permet pas de reconnaitre ce type d'activités, comme « jouer au tennis », avec précision. En effet, chaque sujet effectue cette activité d'une manière très différente. Un cadre comparatif entre les activités de deux sujets différents devient difficile dans ce cas. L'une des solutions que nous proposons est d'avoir recours à la ségmentation par activités basiques ou élémentaires composants cette activité. Ainsi, chaque activité « complexe » sera ségmentée en plusieurs activités basiques ou élémentaires. La reconnaissance de l'activité « complexe » sera faite par le biais des activités élémentaires ainsi définies.

Les premières réflexions menées, dans le cadre de mes travaux de thèse, concernent l'utilisation de modèles d'apprentissage automatique les plus répandus dans l'état de l'art afin de reconnaitre des activités physiques élémentaires.

Nous définissons l'activité physique élémentaire comme étant l'ensemble A={'debout', 'assis', 'allongé', 'marcher'}. Cet ensemble décrit les activités qui ne dépendent d'aucune donnée évènementielle comme le lieu, la météo, l'état psychologique, etc. Nous considérons donc que, quelle que soit l'activité physique effectuée par un individu, elle est nécéssairement incluse dans l'ensemble A. De plus, seules les données des accéléromètres, gyroscopes et magnétomètres peuvent les caractériser.

La première étape du processus de reconnaissance de l'activité physique concerne la collecte de données. Bien que l'état de l'art ait montré l'existence d'une multitude de bases de données, celles-ci restent difficilement généralisables à d'autres cadres d'études. En effet, les bases de données existantes comportent plusieurs limitations rendant leur réutilisation compliquée (temps de collecte de données limité ou insuffisant, environnements contrôlés, type d'objets connectés non adaptés, nombre de participants insuffisant, etc.).

La deuxième étape, consiste à extraire les descripteurs les plus informatifs afin de les donner en entrées aux différents algorithmes d'apprentissage pour la classification.

Enfin, la dernière étape consiste à appliquer les techniques d'apprentissage automatique afin de classifier les descripteurs ainsi extraits. Dans cette étude, nous avons exploité trois objets connectés : un smartphone (ou téléphone intelligent), une smartwatch (ou montre intelligente) et une smart TV (ou une télécommande connectée à un boitier TV) dans un environnement non contrôlé, où les participants effectuent librement leurs activités sans instructions préalables.

La qualité de la donnée ainsi que les descripteurs à extraire ont une influence majeure sur la classification des activités. Des données entachées de bruits ou contenants des valeurs aberrantes ne permettent pas d'aboutir à une bonne précision de reconnaissance. De même, les descripteurs non adaptés ou peu représentatifs de l'information à extraire des signaux des capteurs embarqués dans les objets connectés, influencent négativement la bonne classification des activités. Nous allons donc étudier les descripteurs utilisés dans la littérature permettant d'avoir les meilleures performances de reconnaissance des activités élémentaires.

Les contributions de ce chapitre concernent : premièrement, créer un corpus de données des activités de 33 participants. Deuxièmement, les données collectées seront prétraitées afin d'extraire les principaux descripteurs. Plusieurs descripteurs seront explorés : des descripteurs du domaine temporel et ceux du domaine fréquentiel. Une étude comparative incluant plusieurs algorithmes d'apprentissage sera réalisée afin de choisir le meilleur modèle à entrainer avec le meilleur descripteur. Troisièmement, nous proposons une étude approfondie sur l'utilisation de la Transformée en Cosinus Discrète (DCT) comme principal descripteur. En particulier, nous montrerons l'intérêt de ségmenter et chevaucher des fenêtres des signaux des accéléromètres, des gyroscopes et des magnétomètres, des trois objets connectés. Ainsi, un paramétrage efficace de la DCT a été proposé. Nous verrons comment calculer la meilleure taille de découpe des signaux des données ainsi que la meilleure taille du descripteur (DCT) à retenir pour la classification finale.
Enfin, nous montrerons l'intérêt de chevaucher entre deux activités physiques élémentaires et son impact sur la reconnaissance de l'activité physique.
Nous avons effectué une étude préliminaire afin de tester la pertinence du type de l'environement de l'experience, le type et le nombtre de capteurs à choisir ainsi que le type d'activité à détécter. Les résultats sont présentés dans l'annexe F.

2. Vue d'ensemble des traitements

La figure 1 met en évidence les différentes étapes permettant la reconnaissance des activités élémentaires.

Les données issues des capteurs des trois objets connectés ont été collectées. Puis, quatre types d'annotations ont été appliqués à l'ensemble de données afin de les classifier en quatre activités élémentaires. Les données vont ensuite subir une série de traitements afin d'extraire des descripteurs représentatifs de ces activités. Les descripteurs ainsi extraits vont être présentés en entrées des différents algorithmes d'apprentissage automatique. Nous proposons ici d'utiliser cinq types d'algorithmes de classification les plus utilisés dans l'état de l'art: SVM à noyau Gaussien (SVM) [97], un réseau de neurones de type perceptron multicouches (MLP) [98], un arbre de décision (DT) [65], un modèle naïf de Bayes (NB) [99] et enfin l'algorithme des forets aléatoires (RF) [100].

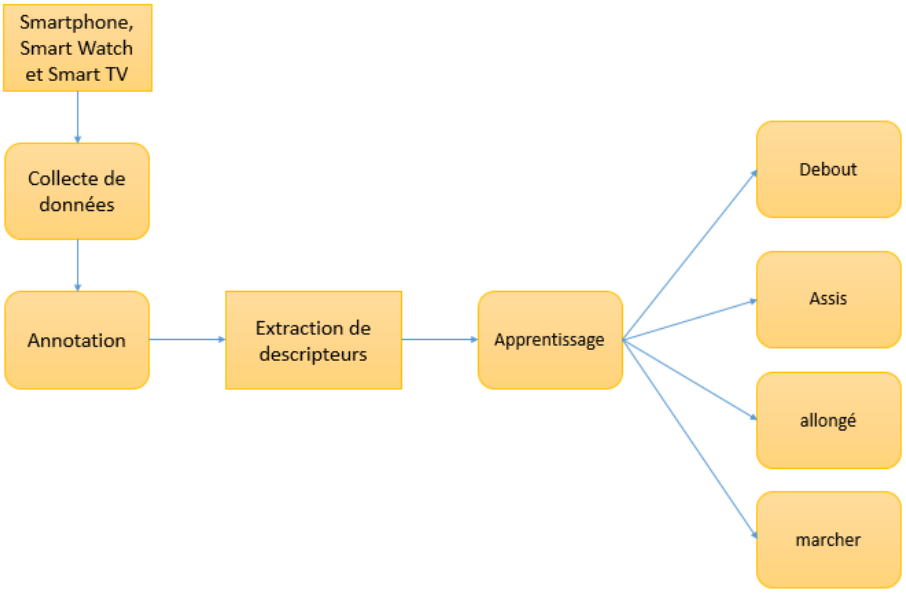


Figure 1: Vue d'ensemble des traitements pour la reconnaissance de l'activité physique élémentaire.

Dans la partie suivante, nous proposons de décrire le corpus de données, le processus de collecte des données et l'environnement de l'expérience.

3. Création du corpus

Nous commençons par décrire l'experience réalisée. Puis, nous proposons d'effectuer une pré-analyse du corpus de données. Nous présenterons ensuite une série de traitements afin d'extraire les descripteurs les plus informatifs. Ainsi, nous menerons une étude approfondie afin de montrer la pertinence du choix de la DCT comme principal descripteur. Enfin, nous présenterons une étude permettant la prise en compte des données de transition entre deux activités succéssives.

3.1. Description de l'expérience

L'expérience a impliqué 33 participants $(P_i, i = 1,..,33)$, agés entre 22 ans et 64 ans (14 femmes et 19 hommes), de tailles (164 cm à 191 cm) et de poids entre 56 kg et 112 kg, dans leurs domiciles pendant 6 mois.

Tous les participants ont été équipés d'un smartphone, d'une smart Watch et d'une smart TV. Six cameras IP ont été fixées à différents endroits de leurs domiciles afin de récolter les données relatives à l'activité physique des participants. Les vidéos ont été enregistrées et synchronisées sur un server local. Les trois objets connectés contiennent les capteurs renseignés dans le **tableau 1**.

Tableau 1 : les différents capteurs embarqués dans les trois objets connectés.

Smartphone	Smart Watch	Smart TV
Accéléromètre, gyroscope, magnétomètre	Altimètre baromètre	Accéléromètre, magnétomètre
Détecteur de proximité	Cardiofréquencemètre	gyroscope,
Capteur de luminosité ambiante	Accéléromètre	Microphone
Capteur d'empreinte digitale	Gyroscope	Surface tactile
Microphone	Microphone	

Tous les participants ont rempli un formulaire de consentement à travers lequel ils ont exprimé leurs accords sur le principe de la collecte et le traitement de données et les objectifs de leur participation à cette expérimentation.

Un système d'information a été proposé afin d'enregistrer les données provenant de tous les capteurs des trois objets connectés (**Figure 2**).

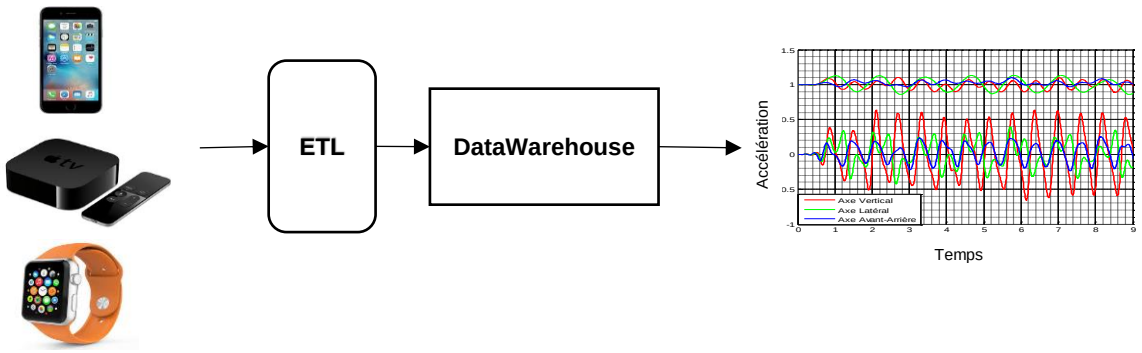


Figure 2 : Système d'information pour la collecte de données des activités physiques. Ce système est composé de trois éléments : (1) les trois objets connectés (smartphone,

smartwatch et smart TV), (2) les données extraites des capteurs des trois objets connectés vont être intégrées dans une base de données à l'aide d'un outil d'intégration (Extract, Load, Transform, ETL) et enfin (3) ces données vont être stockées dans un entrepôt de données (datawarehouse, en anglais) afin de faire des extractions et des visualisations des données.

La fréquence d'échantillonnage était de 8KHz pour les capteurs audio et de 128 Hz pour les autres capteurs. L'enregistrement journalier était ségmenté sur des plages de 3 heures et 50 minutes afin de permettre de recharger les batteries des trois objets connectés, particulièrement celle du smartphone qui avait une autonomie assez limitée.

Nous avons développé trois applications mobiles permettant de collecter les données des capteurs embarqués dans les objets connectés : une application sur le smartphone, une application sur la smart Watch et une application sur la smart TV. En particulier, l'application développée sur la smart TV permet, entre autre, d'accéder au contenu du programme TV, son résumé, son début et sa fin. Nous savons donc accéder à des contenus multimédias pour avoir quelques informations sur les programmes TV en cours.

Les données ainsi collectées ont été intégrées dans un entrepôt de données (datawarehouse). Le datawarehouse ainsi développé contient alors trois datamarts :

- Un datamart pour les données du smartphone.
- Un datamart pour les données de la smartwatch.
- Un datamart pour les données de la smart TV.

Au sein de chaque datamart, nous avons effectué des extractions afin de faire des traitements sur les données ainsi collectées.

Les vidéos enregistrées ont été labélisées avec le logiciel Elan Software [101] afin de reconnaitre toutes les activités élémentaires. Selon la présence ou pas de chevauchement entre deux activités successives, nous proposons à la fin de ce chapitre d'étudier l'impact du chevauchement entre deux activités sur la qualité de l'annotation.

Nous proposons, dans la partie suivante, d'effectuer une analyse préliminaire sur le corpus de données.

3.2. Pré-analyse du corpus

Les données ainsi collectées à partir des capteurs du **tableau 1** seront traitées afin de les rendre exploitables par les algorithmes d'apprentissage automatique. La durée totale de chaque activité physique élémentaire est représentée dans la **figure 3**.

Nous remarquons que la distribution temporelle de l'ensemble des activités physiques élémentaires n'est visiblement pas bien répartie par rapport au volume horaire alloué à

chacune. En effet, l'activité « Allongé » est majoritaire, alors que l'activité « marcher » reste minoritaire. Ainsi, les deux activités élémentaires « debout » et « assis » sont approximativement égales en termes de volume horaire. Une vue d'ensemble de la distribution des activités élémentaires pendant les jours de la semaine est représentée dans la **Figure 4**.

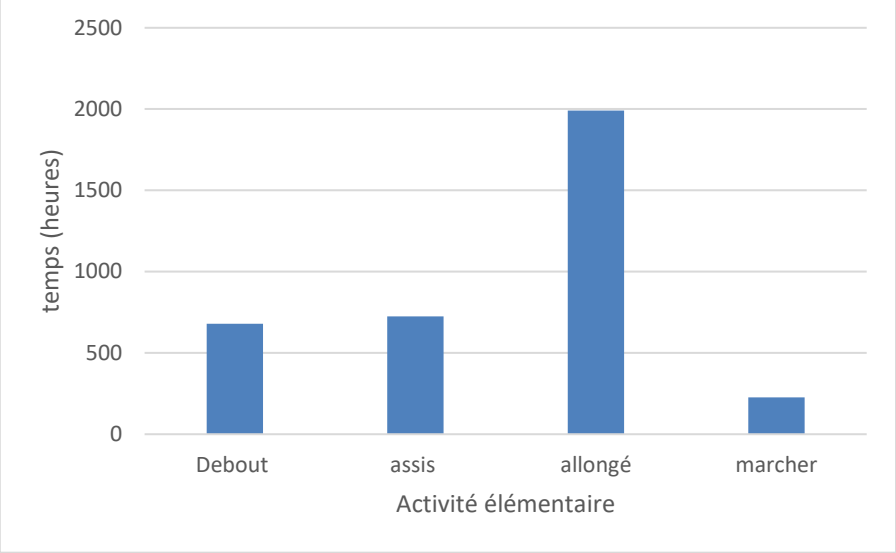


Figure 3 : Distribution temporelle des activités physiques élémentaires.

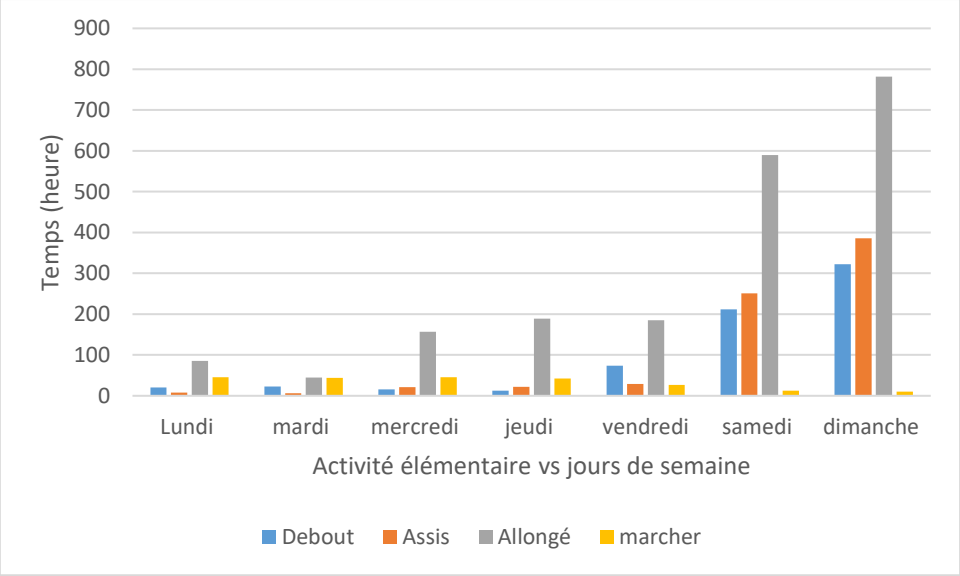


Figure 4 : distribution du temps d'activité élémentaire dans la semaine pour les 33 participants.

L'activité élémentaire « allongé » est globalement majoritaire pendant les journées de la semaine. Alors que les autres activités élémentaires sont beaucoup moins apparentes. Nous remarquons également que la tendance est différente selon les journées de la semaine, en particulier les weeks ends où les activités élémentaires « allongé », « assis », et « debout » sont plus importantes que l'activité « marcher ».

L'état de l'art a montré que la position du capteur ou de l'objet connecté et sa localisation (ou son portage) sur le corps influe considérablement la qualité de la reconnaissance [102]. Cependant, l'une des limitations soulevées par l'état de l'art est l'absence d'études approfondies considérants les objets connectés en état de « hors de portée », c'est-à-dire les états où les objets connectés ne sont pas portés par le participant. Par exemple : smartphone abandonné sur la table, etc. Les données correspondantes à l'état « hors de portée ou hors de portage» sont, dans la plupart des cas, supprimées ou négligées [103].

Suite à l'analyse des vidéos de collecte de données, nous avons identifié quatre cas de portabilité des objets connectés.

- Les trois objets connectés sont tous portés par les participants.
- Un des objets n'est pas porté par les participants.
- Deux objets ne sont pas portés par les participants.
- Trois objets ne sont pas portés par les participants.

La figure 5, représente la distribution de l'activité physique élémentaire en fonction de la portabilité des objets connectés.

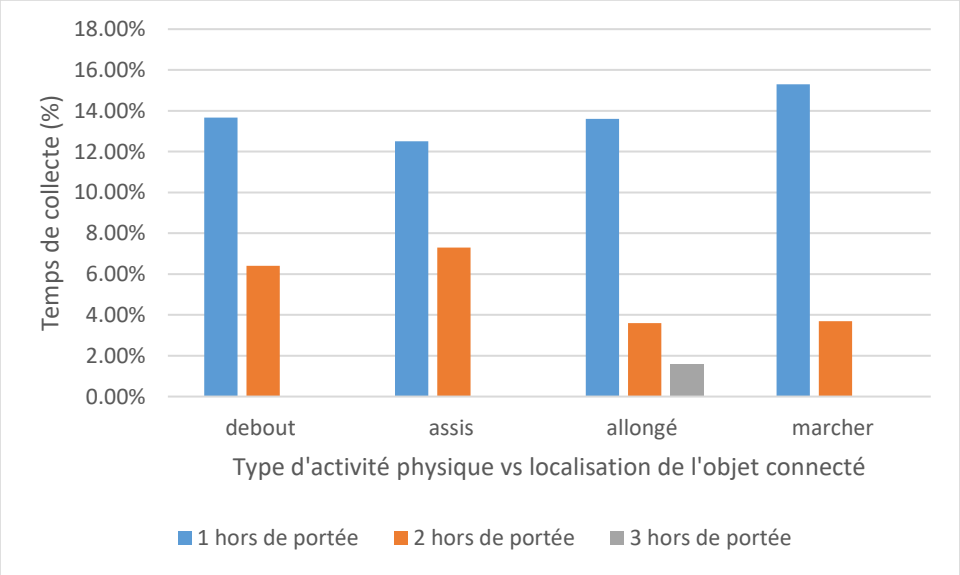


Figure 5 : Distribution du temps de portabilité des objets connectés par rapport aux activités physiques élémentaires.

Seule l'activité « Allongé » comporte au plus trois objets connectés en état de « hors de portée ». Les autres activités élémentaires peuvent avoir des portabilités allant de 1 objet connecté en état de « hors de portée » jusqu'à 2 objets connectés. Ainsi, les pourcentages de « hors de portage » sont relativement bas. En effet, le plus grand pourcentage où l'un des objets connectés est resté « hors de portée » est d'environ 16% du temps total

d’enregistrement. Toutefois, l’activité « allongé » est la seule où les trois objets connectés sont restés en état de « hors de portée ».

4. Extraction de descripteurs

Après avoir collecté et annoté les données, nous allons extraire les descripteurs les plus utilisés dans l’état de l’art [104]. Puis, nous mènerons une étude comparative afin d’en sélectionner le meilleur. Dans ce cas et comme nous l’avons défini dans la section1 de ce chapitre, seules les données des accéléromètres, des gyroscopes et des magnétomètres des trois objets connectés sont prises en compte.

Nous avons extrait les descripteurs des deux domaines temporel et fréquentiel. Les descripteurs extraits du domaine temporel sont les suivants :

- La moyenne (mean) [60-66]
- Ecart type (Std) [89-93].
- Zone d'amplitude du signal (Signal Magnitude Area) (SMA) représente l'amplitude d'un ensemble de données. Il est exprimé comme la somme de toutes ses valeurs [105].
- Corrélation entre les signaux inter-capteurs (Inter Sensor Signals Correlation) (ISSC), mesure la corrélation croisée entre l'axe d'un capteur pour un ensemble de données, un pour chaque axe.
- Coefficients de régression automatique (Auto-regression Coefficients ou ARC), représente les paramètres du modèle autorégressif construit à partir du signal. Un modèle autorégressif permet de reconstruire le signal sous la forme d'une combinaison linéaire de ses valeurs précédentes [93-114].
- L’énergie : désigne l'énergie du signal et est définie comme la somme de ses valeurs au carré. (Et) [122-126].Les descripteurs extraits du domaine fréquentiel sont les suivants :
- Entropie : représente l'entropie de Shannon. Elle est la quantité moyenne d'informations que nous recevons par événement (valeur du signal). Elle représente intuitivement la "surprise attendue" de la valeur, notée Entropy [221-223].
- L'asymétrie de fréquence (Frequency Skewness) : représente le biais, départ de l'horizontale symétrie d'un histogramme, du spectre de fréquence. c'est une mesure de la forme du spectre (FS) [226-229].

- L'énergie : désigne l'énergie du signal. Elle est définie comme la somme de ses valeurs au carré (E).
- DCT : La transformée en cosinus discrète. Elle est très utilisée en traitement d'images, possède une particularité de compression de données et fournie une décorrelation optimale pour un signal donné.
- TFD : La Transformée de Fourier discrète proche de la DCT sauf que le noyau de projection est une exponentielle complexe et crée donc des coefficients complexes

Afin d'extraire ces descripteurs, les signaux des accéléromètres, gyroscopes et magnétomètres doivent être ségmentés en utilisant des fenêtres de taille 512 points. Cette taille est majoritairement utilisée dans la plupart des travaux sur la reconnaissance de l'activité [106].

Afin d'expliquer les signes des différents axes des trois types de capteurs (accéléromètre, gyroscope et magnétomètre), nous proposons d'illustrer ce propos par un graphique des signaux accéléromètriques. Le meme raisonnement est à appliquer pour les signaux des gyroscopes et des magnétomètres.

D'après la position de l'accéléromètre embarqué dans le smartphone, la smart Watch ou la Smart TV, les signes de chaque axe sont pris de la manière suivante :

- Axe Vertical (Rouge): Positif dans le sens de la gravité, négatif vers le haut.
- Axe Latéral (Vert) : Positif vers la droite, négatif vers la gauche.
- Axe Avant-Arrière (Bleu) : Positif vers l'avant, négatif vers l'arrière.

Pour chaque graphique du signal accéléromètrique, l'axe des abscisses représente le temps (en secondes) et l'axe des ordonnées l'accélération (en multiples de g, la gravité). **La figure 6** illustre ce type de signaux.

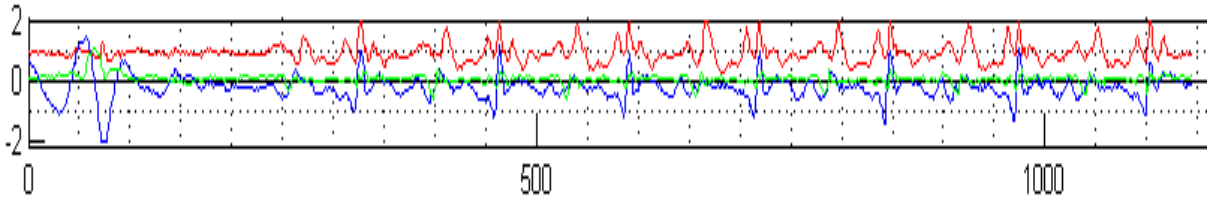


Figure 6 : signaux accéleromètriques représentant l'activité marcher.

Dans les sections 4.1 et 4.2 nous présenterons la démarche suivie pour découper les signaux ainsi que la taille de découpe optimale retenue.

4.1. Découpage et ségmentation des signaux

La première étape permettant l'exploitation des données de ces signaux correspond au découpage et ségementation de signaux des trois types de capteurs. Ainsi, les signaux des

accéléromètres, des gyroscopes et des magnétomètres des trois objets connectés seront découpés en portions ou découpes de tailles Δt (**Figure 7**).

Il existe deux méthodes permettant de retrouver la taille optimale Δt. Une première méthode consiste à faire un découpage manuel. La seconde méthode, consiste à effectuer un découpage dynamique [42].

L'avantage du découpage manuel est le gain en temps de calcul. En effet, la connaissance à priori (de l'état de l'art) de la taille Δt, permet de l'appliquer directement sans avoir à la rechercher. Réciproquement (la taille Δt n'est pas connue), il faudrait effectuer des tests en essayant plusieurs tailles afin de retrouver la taille optimale. Ce travail est très fastidieux à mettre en place avec un risque de divergence de la solution optimale. Les méthodes dynamiques permettent de retrouver la bonne taille Δt en procédant iterativement : Δt allant de 0 jusqu'à atteindre la taille maximale du signal. Toutefois, cette approche s'avère trop couteuse en temps de calcul [77-79].

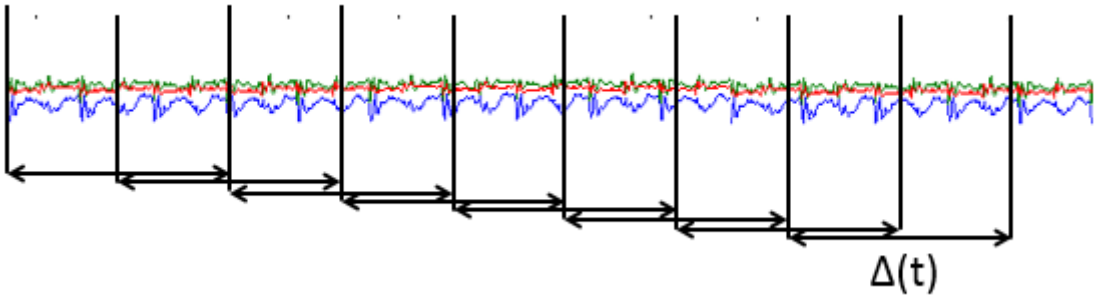


Figure 7 : Découpage des signaux de l'accéléromètre, gyroscope et magnétomètre.

Nous avons étudié les deux approches, et compte tenu des objectifs futurs, nous avons retenu la sélection manuelle discrète non dynamique des coupes, que nous allons détailler dans la partie suivante.

4.2. Taille d'une coupe

Nous avons étudié deux tailles (Δt) de coupe. La première est d'une longueur de 256 échantillons (2,13 secondes) et la seconde est de 512 échantillons (4,26 secondes) avec un chevauchement entre coupes. Nous ne pouvons pas aller au-delà de 512 points par fenêtre de peur de compromettre l'aspect temps réel de la réponse.

Nous avons commencé par étudier les descripteurs temporels ou fréquentiels permettant d'avoir les meilleures précisions de reconnaissance. Nous avons entrainé un SVM à noyau Gaussien en utilisant la validation croisée 10 fois [96]. Les résultats sont présentés dans la section 4.3

4.3. Résultats Globaux

4.3.1. L'influence du domaine du descripteur et le chevauchement entre les coupes

Afin de sélectionner le descripteur qui donne les meilleurs résultats de classification, nous avons entrainé un SVM à noyau Gaussien. Nous avons d'abord comparé les performances des descripteurs du domaine temporel et ceux du domaine fréquentiel en prenant en compte le chevauchement entre les coupes. Les résultats sont résumés dans les **figures 8 à 15.**

a) Configuratipon présentant un chevauchement :

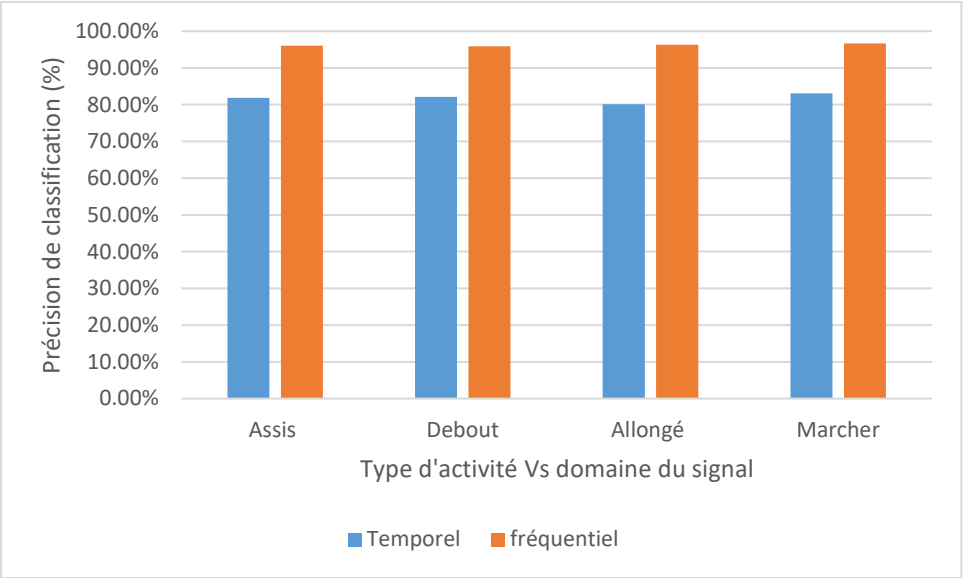


Figure 8 : Comparaison des précisions de classification des activités physiques élémentaires entre le domaine temporel et fréquentiel, en utilisant les données accéléromètriques. Ici, tous les descripteurs cités plus haut ont été concaténés et rangés selon leur type (fréquentiel ou temporel), avec chevauchement entre les coupes.

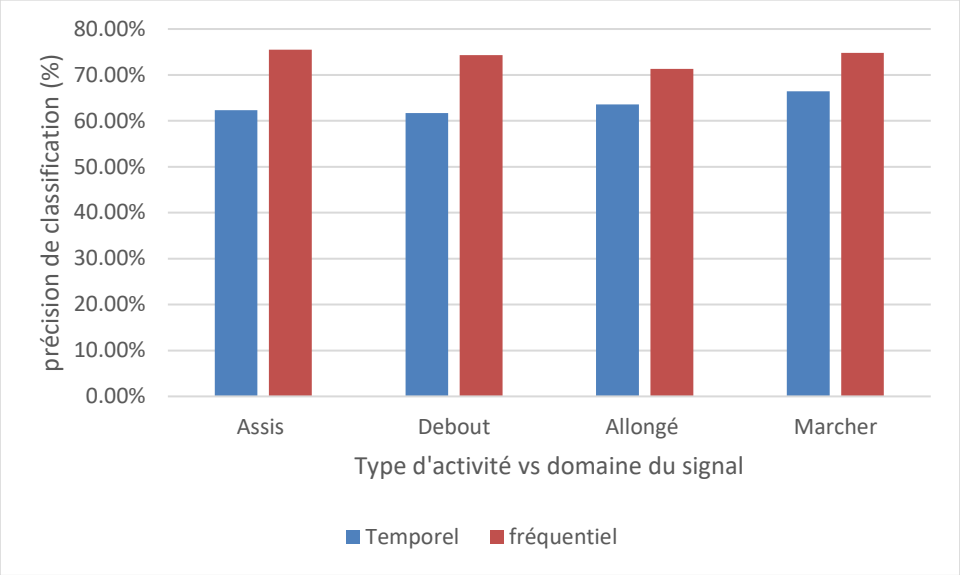


Figure 9: Comparaison des précisions de reconnaissance des activités physiques élémentaires dans les domaines temporel et fréquentiel en utilisant les données des signaux des trois gyroscopes, avec chevauchement entre les coupes.

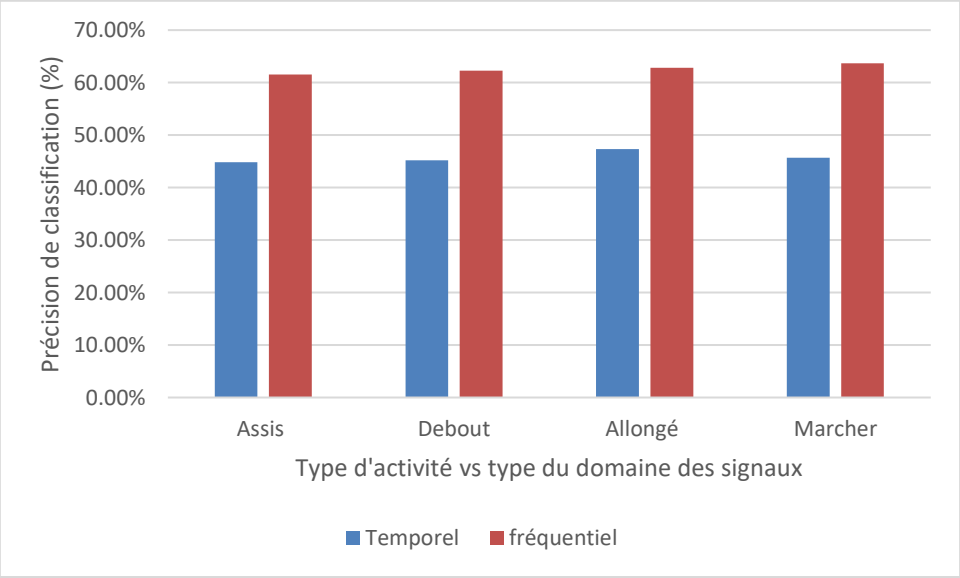


Figure 10: Comparaison des précisions de classification des domaines temporel et fréquentiel en utilisant les signaux des magnétomètres, avec chevauchement entre les coupes.

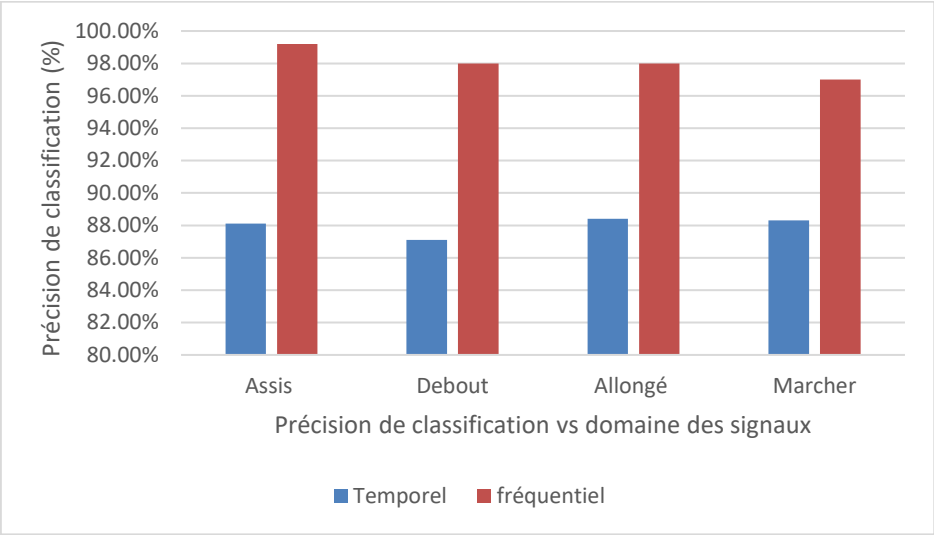


Figure 11 : Comparaison des précisions de classification des activités élémentaires entre les domaines temporel et fréquentiel en concaténant les signaux de tous les capteurs des trois objets connectés (accéléromètres, gyroscopes, magnétomètres), avec chevauchement entre les coupes.

Les précisions de classification des activités physiques élémentaires sont bien meilleures dans le domaine fréquentiel que dans le domaine temporel en chevauchant les différentes coupes.

b) Configuration ne présentant pas de chevauchement :

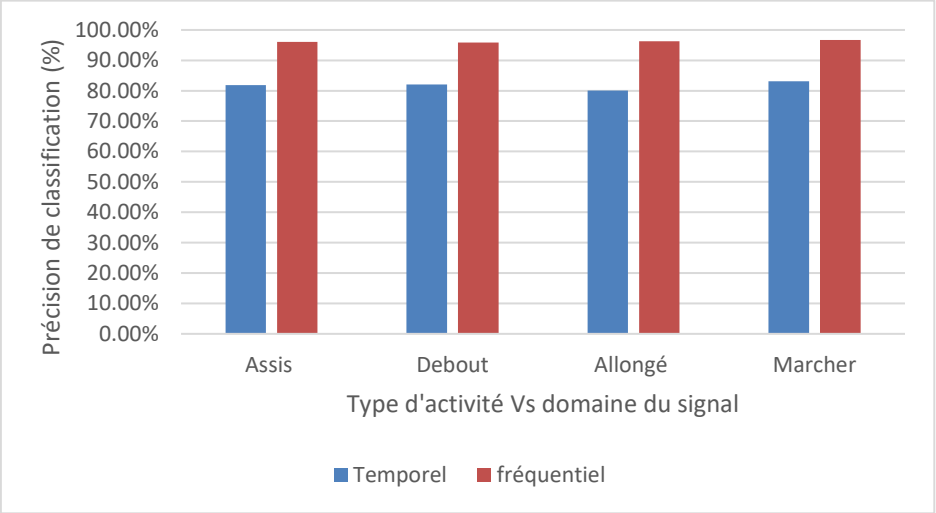


Figure 12: Comparaison des précisions de classification des activités physiques élémentaires entre le domaine temporel et fréquentiel. Ici, tous les descripteurs cités plus haut ont été concaténés et rangés selon leur type (fréquentiel ou temporel), sans chevauchement entre les coupes.

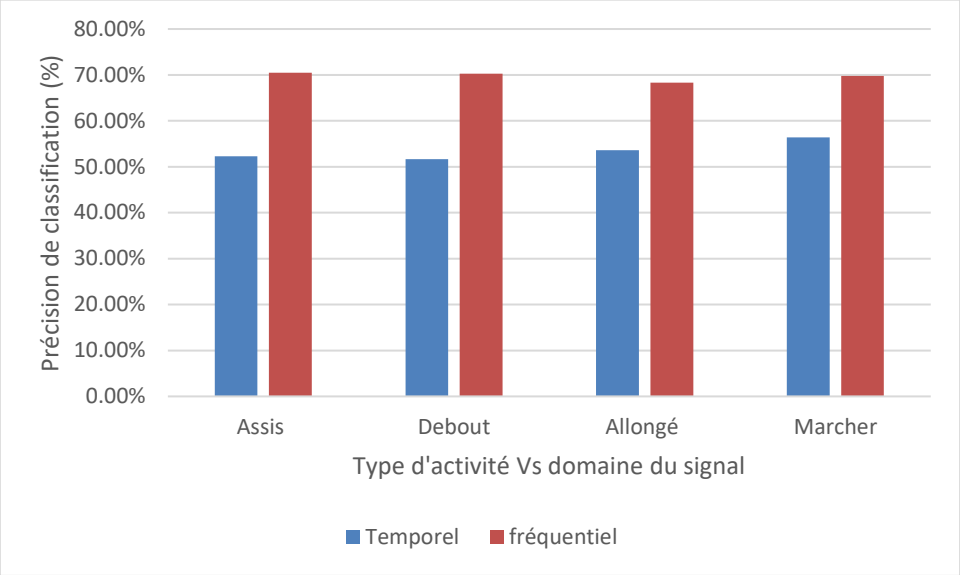


Figure 13: Comparaison des précisions de reconnaissance des activités physiques élémentaires dans les domaines temporel et fréquentiel en utilisant les données des signaux des trois gyroscopes, sans chevauchement entre les coupes.

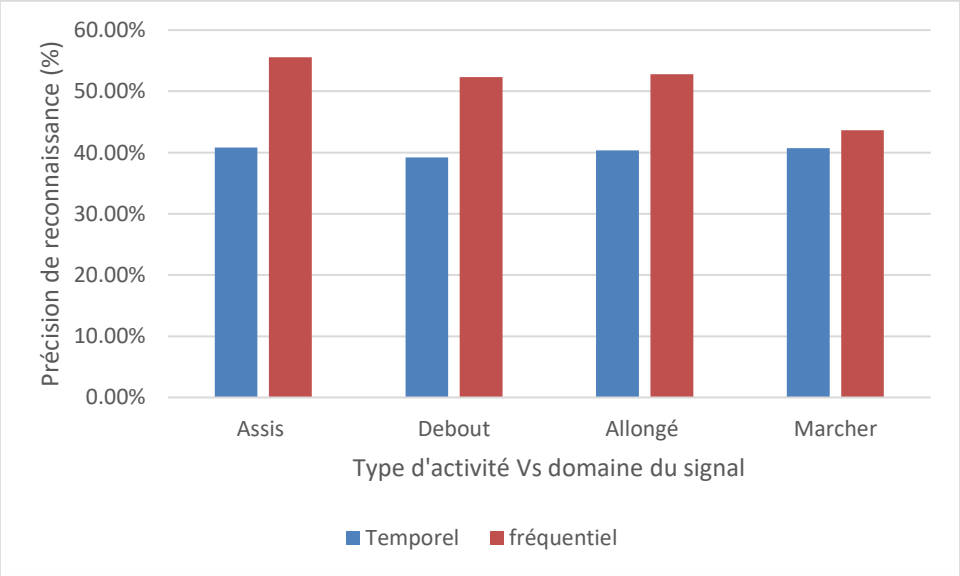


Figure 14: Comparaison des précisions de classification des domaines temporel et fréquentiel en utilisant les signaux des magnétomètres sans chevauchement entre les coupes.

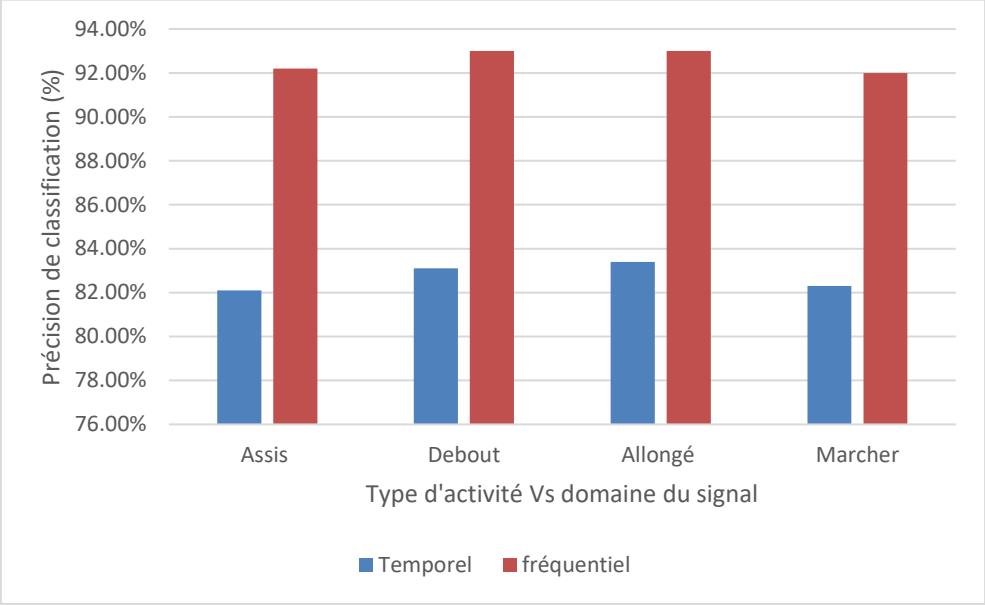


Figure 15 : comparaison des précisions de classification des activités élémentaires entre les domaines temporel et fréquentiel en concaténant les signaux de tous les capteurs des trois objets connectés (accéléromètres, gyroscopes, magnétomètres), sans chevauchement entre les coupes.

Les résultats montrent que la précision de classification des descripteurs du domaine fréquentiel sont meilleurs que celles du domaine temporel. De plus, la prise en compte du chevauchement entre les coupes s'avère efficace (en termes de précision). Les transitions entre les différentes coupes permettent d'améliorer la précision de classification, ce qui reste en adéquation avec l'état de l'art.

4.3.2. Etude comparative

Afin d'étudier la performance et l'efficacité d'utiliser un SVM à noyau Gaussien, nous avons entrainé les quatre autres algorithmes (MLP, DT, NB, RF) et comparé avec les précisions obtenues avec l'algorithme du SVM à noyau Gaussien. Les résultats de cette étude sont présentés dans la **figure 16.**

Les résultats montrent que l'algorithme du SVM à noyau Gaussien obtient de meilleurs résultats. Ainsi, le chevauchement entre les coupes s'avère mieux adapté à notre étude. Enfin, les descripteurs du domaine fréquentiel permettent d'aboutir à de meilleurs résultats de classification.

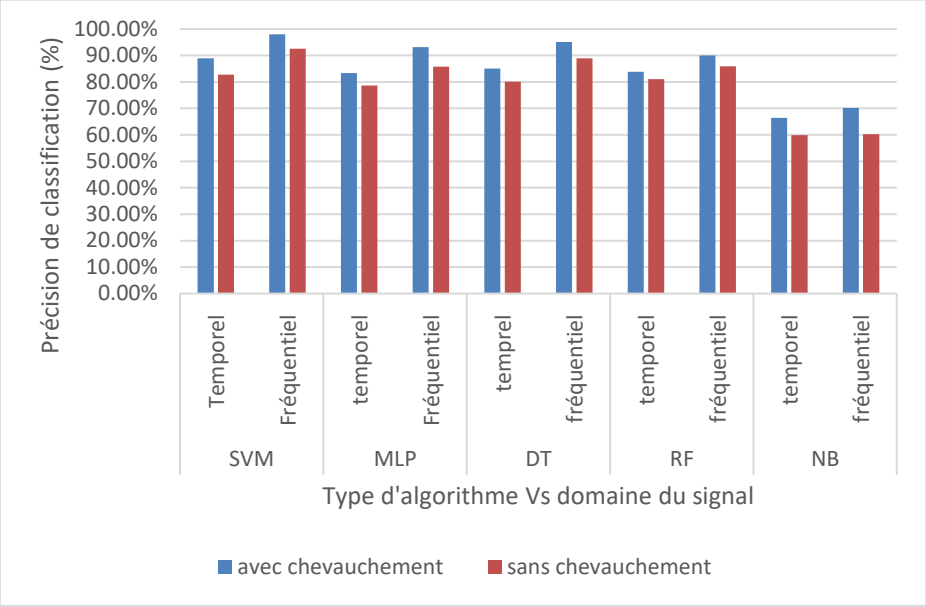


Figure 16 : Etude comparative entre les algorithmes d'apprentissage prenant en compte le type du domaine des descripteurs ainsi que la prise en compte du chevauchement entre les coupes.

Compte tenu des résultats de cette étude, nous décidons donc de poursuivre la sélection de descripteurs avec ceux du domaine fréquentiel. Ainsi, le modèle de classification choisi étant le SVM à noyau Gaussien. La **figure 17**, représente les résultats de classification des activités physiques élémentaires en utilisant les descripteurs du domaine fréquentiel ainsi choisis.

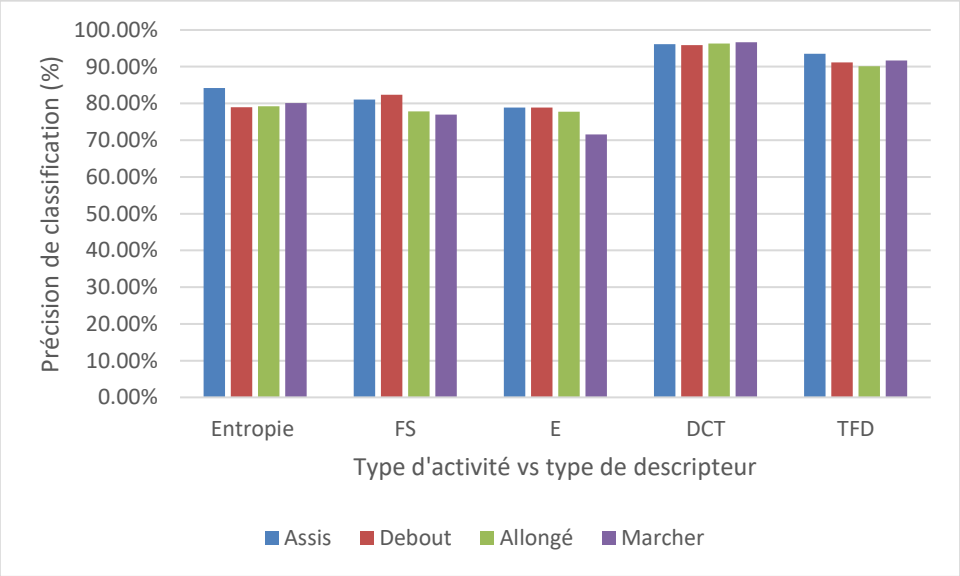


Figure 17: Précision de reconnaissance des activités physiques élémentaires en utilisant toutes les données des trois objets connectés (accéléromètres, gyroscopes, magnétomètres).

Comme le montre **la figure 17**, l'utilisation de la DCT comme principal descripteur s'avère efficace, une précision moyenne de 89% a été atteinte. Nous décidons de le conserver comme

principal descripteur des données de l'activité physique élémentaire. Nous souhaitons explorer ce descripteur afin de dégager les principaux paramètres permettant de l'utiliser efficacement. Nous proposons d'aborder notamment le nombre d'échantillons nécessaires au chevauchement entre deux coupes et la taille du descripteur requise permettant d'avoir une meilleure précision de classification.

4.4. La DCT comme principal descripteur

Nous avons montré dans la séction précédente que la DCT est le descripteur le plus adapté pour l'apprentissage des activités physiques élémentaires. Afin de mener une étude approfondie sur ce choix, nous proposons d'explorer l'influence de quatre principaux points :

- Le chevauchement entre deux coupes dans deux cas : en présence de chevauchement entre découpes ou en l'absence de celui-ci.
- La taille adéquate du descripteur
- Voir s'il y a possibilité de choisir un, deux ou tous les trois axes (x, y, z) des trois capteurs afin d'effectuer l'apprentissage du modèle.

Nous nous intéressons, dans un premier temps à l'étude de la taille de la coupe dans deux cas différents : i) avec ou ii) sans chevauchement. Les résultats de l'exploitation de la totalité des données des accéléromètres, des gyroscopes et des magnétomètres sont présentés dans le **tableau 2**.

Tableau 2 : Performances de reconnaissance des activités physiques élémentaires avec ou sans chevauchement entre les fenêtres de découpage.

Δ(t) / Activité	512		256	
	Sans chevauchement	**Avec chevauchement**	**Sans Chevauchement**	**Avec chevauchement**
Debout/autres	84%	90.2%	68%	74%
Assis/autres	83%	91%	67%	74%
Allongé/autres	85%	89%	65%	71%
Marcher/autres	85%	89%	60%	72%
Moyenne	84.25%	89.8%	65%	72.75%

Nous avons effectué une étude préliminaire sur la taille de la fenetre de chevauchement entre coupes. Nous avons retenu 256 comme taille optimale. Les performances de reconnaissance des activités physiques élémentaires sont plus élevées quand il y a chevauchement entre

découpes. Ainsi, avec $\Delta t = 512$ échantillons, les précisions de classification des activités physiques sont meilleures.

Afin de montrer l'efficacité du choix de cette taille ($\Delta t = 512$), nous avons fait varier Δt entre 128 et 512. Nous avons ensuite enregistré les performances de classification dans **les tableaux 3, 4 et 5**.

Tableau 3 : variation de la taille de la fenêtre de découpage des données des accéléromètres.

Δ(t) / activité	128	256	512
Debout/autres	57,59%	70%	88%
Assis/autres	56,61%	71%	88%
Allongé/autres	53,68%	70%	89%
Marcher/autres	51,57%	70%	90%
Moyenne	54.86%	70.25%	88.75%

Tableau 4 : variation de la taille de la fenêtre de découpage des données des gyroscopes.

Δ(t) / activité	128	256	512
Debout/autres	43,59%	50%	61%
Assis/autres	46,31%	52%	61%
Allongé/autres	43,17%	50%	61%
Marcher/autres	41,57%	50%	60%
Moyenne	43.66%	50.5%	60.75%

Tableau 5 : variation de la taille de la fenêtre de découpage des données des magnétomètres.

Δ(t) / activité	128	256	512
Debout/autres	33,19%	40%	50%
Assis/autres	32,31%	41%	51%
Allongé/autres	30,17%	42%	51%

Marcher/autres	30.22%	41.17%	50.84%
Moyenne	31.47%	41.04%	50.71%

La concaténation des vecteurs des descripteurs des données des trois types de capteurs nous donne les résultats présentés dans le **tableau 6.**

Tableau 6 : Performances du classifieur en fonction de la taille de chaque coupe, en concaténant toutes les données de tous les capteurs.

Δ(t) / activité	128	256	512
Debout/autres	59,69%	74%	90.2%
Assis/autres	59,61%	74%	91%
Allongé/autres	55,88%	71%	89%
Marcher/autres	54,57%	72%	89%
Moyenne	57,43%	72.75%	89.8%

Ces mesures ont été prises pour une taille du descripteur de 144 (48 pour chaque axe: X, Y et Z, de chaque capteur). Nous remarquons que la taille de chaque coupe ayant les performances optimales est de $\Delta t = 512$. Ainsi, les études suivantes exploitent des coupes de taille $\Delta t = 512$ avec chevauchement entre les coupes.

Nous propsosons dans la section suivante une évaluation de la taille du descripteur ayant permi d'obtenir les meilleurs résultats de classification.

4.4.1. Taille du descripteur

Afin de montrer la pertinence du choix de 48 premiers facteurs de la DCT comme principale taille du descripteur, nous proposons de faire varier celle-ci entre 2 et 64. Cet intervalle est est majoritairement utilisé dans la plupart des travaux.

La taille de chaque coupe étant fixée, nous allons réaliser une étude pour déterminer la taille du descripteur qui permet d'avoir de meilleures performances de classification.

L'une des caractéristiques de la DCT est le regroupement de l'énergie dans les coefficients de basses fréquences. Ainsi, un certain nombre de facteurs de la DCT (Δe) permet d'aboutir à une meilleure classification. Nous proposons, dans ce qui suit, de déterminer le meilleur (Δe), ou la taille du descripteur.

La figure 18 présente une vue d'ensemble du calcul des descripteurs, pour les signaux des accéléromètres, gyroscopes et magnétomètres.

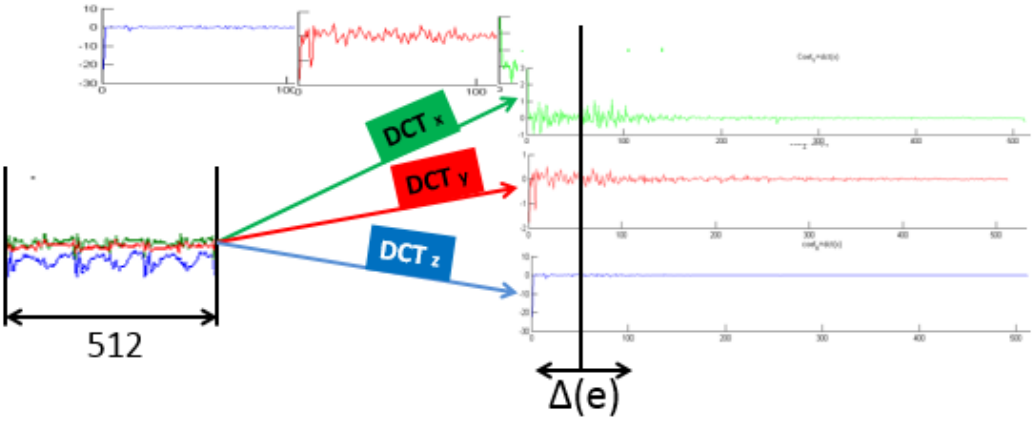


Figure 18 : Calcul des descripteurs (la DCT) pour chaque axe.

Les expériences réalisées pour avoir la taille du descripteur Δ(e) qui donne de meilleures performances sont résumées dans **les tableaux 7, 8, 9 et 10**.

Tableau 7 : Performances de classification en fonction de la taille des descripteurs pour les signaux des accéléromètres.

Δ(t) / Démarche	2	4	8	16	32	48	64
Debout/autres	43%	50%	63%	65%	80%	88%	87%
Assis/autres	41%	51%	60%	65%	81%	88%	86%
Allongé/autres	39%	53%	61%	69%	81%	89%	90%
Marcher/autres	41%	50%	60%	65%	82%	90%	90%
Moyenne	41%	51%	61%	66%	81%	**88.75%**	88.2%

Tableau 8 : Performances de classification en fonction de la taille des descripteurs pour les signaux des gyroscopes.

Δ(t) / Démarche	2	4	8	16	32	48	64
Debout/autres	32%	43%	53.4%	58.3%	70.3%	81.3%	74.3%
Assis/autres	31%	41.2%	50.4%	55.7%	71.32%	83.2%	75.9%
Allongé/autres	27%	43.3%	52.7%	59.8%	71.8%	85.3%	73.2%
Marcher/autres	30%	40.1%	51.8%	55.88%	72.9%	83.5%	70.8%
Moyenne	30%	41.9%	52.07%	57.42%	71.58%	**83.82%**	73.5%

Tableau 9 : Performances de classification en fonction de la taille des descripteurs des signaux des magnétomètres.

Δ(t) / Démarche	2	4	8	16	32	48	64
Debout/autres	23%	33%	43.45%	48.33%	60.31%	72.33%	70.3%
Assis/autres	21.5%	31.25%	40.43%	45.71%	61.33%	71.23%	70.9%
Allongé/autres	18.3%	33.37%	42.79%	49.89%	61.89%	73.33%	70.1%
Marcher/autres	22.9%	30.17%	41.84%	45.19%	62.92%	73.14%	68.3%
Moyenne	21.42%	31.94%	42.12%	47.28%	61.62%	**72.53%**	69.9%

La concaténation des vecteurs des descripteurs de tous les capteurs donne les résultats du **tableau 10**.

Tableau 10 : Performances de classification en fonction de la taille des descripteurs.

Δ(t) / Démarche	2	4	8	16	32	48	64
Debout/autres	44%	52%	66%	69%	81%	90.2%	90%
Assis/autres	44%	53%	61%	68%	83%	91%	90%
Allongé/autres	41%	55%	63%	70%	85%	89%	88%
Marcher/autres	43%	51%	63%	69%	85%	89%	90%
Moyenne	43%	52.75%	63.25%	69%	83.5%	**89.8%**	89.5%

Nous remarquons que les 48 premiers facteurs de la DCT permettent effectivement d'obtenir de meilleures performances de reconnaissance : en moyenne 89.8%.

Nous avons démontré, jusqu'à présent, que les performances délivrées par les accéléromètres étant supérieures aux performances de classification des gyroscopes et des magnétomètres. Toutefois, la concaténation des trois capteurs permet d'améliorer les résultats de classification. Dans l'objectif d'étudier la contribution de chaque capteur à la reconnaissance de l'activité physique élémentaire, nous nous somme focalisé sur l'apport de chaque axe (X, Y, Z) des trois capteurs en matière de précision de classification. La partie suivante propose de détailler cette étude.

4.4.2. Choix des axes

Nous avons prêté une attention particulière aux descripteurs construits à partir des trois axes, X, Y et Z mais aussi pour l'axe X et pour l'axe Y (et ainsi par déduction l'axe z). Les performances enregistrées dans chaque cas sont résumés dans les **tableaux 11, 12, 13 et 14**. A noter que ces performances citées dans ce tableau sont pour une taille de coupe de 512 avec chevauchement de 256 et une taille du descripteur de (Δe) égale à 144(48×3) pour chaque capteur.

Tableau 11 : Performance de classification suivant le choix des axes des signaux des accéléromètres.

Δ(t) / Démarche	X	Y	X,Y,Z
Debout/autres	58,19%	65,17%	88%
Assis/autres	54.01%	69,7%	88%
Allongé/autres	58,1%	65,05%	89%
Marcher/autres	52,11%	61,02%	90%
Moyenne	55.6%	65.23%	88.75%

Tableau 12 : Performance de classification suivant le choix des axes des signaux des gyroscopes.

Δ(t) / Démarche	X	Y	X,Y,Z
Debout/autres	70.11%	73.1%	75.1%
Assis/autres	70.99%	72.33%	76%
Allongé/autres	66.74%	72.04%	75%
Marcher/autres	69.69%	71.71%	74%
Moyenne	69.38%	72.29%	75.02%

Tableau 13 : Performance suivant le choix des axes des magnétomètres.

Δ(t) / Démarche	X	Y	X,Y,Z
Debout/autres	54.9%	55.05%	61.1%
Assis/autres	54.22%	55.8%	63%
Allongé/autres	54.54%	54.58%	65%
Marcher/autres	54.01%	60.66%	64%
Moyenne	55.02%	56.52%	63.27%

Tableau 14 : Performance de classification suivant le choix des axes en concaténant tous capteurs.

Δ(t) / Démarche	X	Y	X,Y,Z
Debout/autres	70,89%	81,7%	90.2%
Assis/autres	74%	81,7%	91%
Allongé/autres	70,82%	81,28%	89%
Marcher/autres	72,11%	81,02%	89%
Moyenne	71.95%	81.41%	**89.8%**

Le choix de la concaténation des axes X, Y et Z pour la formation du descripteur donne de meilleures performances.

Les DCT déterminées de chaque axe (X, Y, Z) sont concaténées avec les descripteurs des données des autres capteurs pour former des descripteurs qui seront ensuite stockés dans

une base de données (BDD). La base de données ainsi construite va servir pour l'étape de classification. Pendant cette étape, nous allons subdiviser notre ensemble BDD en ensemble d'apprentissage et ensemble de test de tailles égales avec un choix aléatoire des données d'apprentissage et de test (Cf. **figure 19**) :

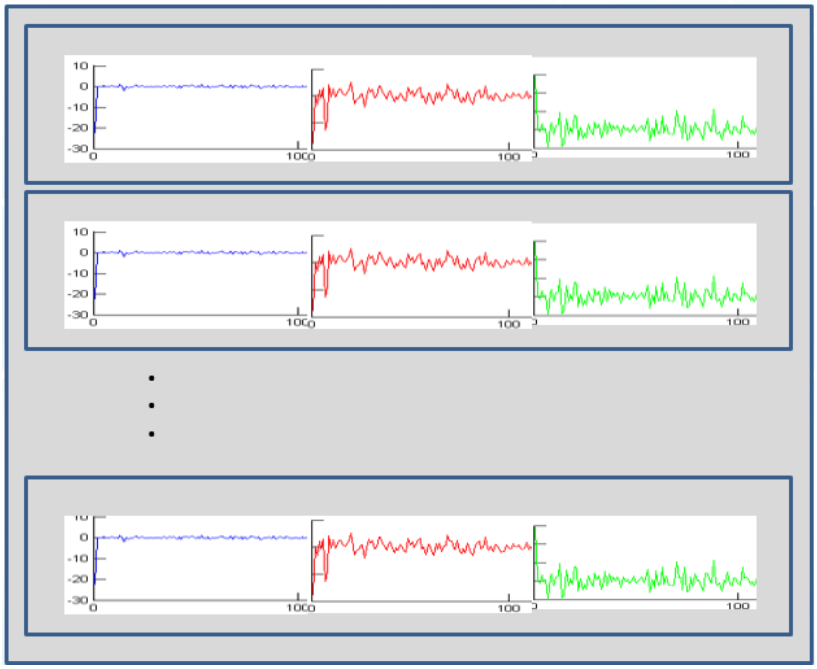


Figure 19 : Schéma illustratif de la concaténation des descripteurs extraits. Une ligne de cette figure représente la concaténation des descripteurs des trois axes d'un capteur. Toute la colonne représente la base de données totale (BDD).

4.4.3. Discussion

Nos résultats du choix de la DCT ainsi que de la méthode de classification sont comparables à ceux de l'état de l'art, tout particulièrement le travail référencé dans [107] qui sont les seuls, à notre connaissance, à avoir utilisé ce type d'approches pour le choix de la DCT comme principale descripteur de l'activité physique. Dans leur travail, les auteurs n'ont considéré qu'un seul accéléromètre triaxial posé dans la poche du pantalon. De plus, l'environnement de l'expérience est contrôlé ne reflètant pas la réalité du déroulement des activités. Ils ont atteint une précision de 97.51% de classification de quatre activités : debout, courir, sauter et marcher.

Les différents paramètres pris en compte, jusqu'à maintenant, pour le calcul de la DCT concerne essentiellement la fenêtre de découpage des signaux et la taille de descripteur qui sont les 48 premiers facteurs de la DCT. Nous proposons dans la dernière partie de discuter un aspect important : le chevauchement entre les activités successives.

4.5. Influence du chevauchement entre activités élémentaires

L'objectif de cette partie est d'étudier l'impact de la prise en compte du chevauchement entre deux activités successives sur la reconnaissance de l'activité physique.

L'état de l'art a montré que le déroulement naturel des activités physiques s'effectue continuellement. Une personne passant d'une position assise vers la position debout traverse une periode de « transition » entre deux activités [106-119]. Ainsi, la plupart des travaux négligent les données de transition lors du processus d'étiquetage des données d'activité et les considèrent comme étant du bruit [125-138]. D'autres travaux effectuent une ségmentation approximative des activités physiques en prenant en compte ces transitions. Ceci est justifié par l'absence d'indication précise sur les moments de début et de fin de l'activité ségmentée.

L'annotation des activités élémentaires de notre corpus a été faite sans prise en compte d'un certain niveau de chevauchement ou données de transition entre les activités. En effet, cette étape a été laissée à l'appréciation du participant et de l'annotateur des données, qui peut ou pas considérer le chevauchement. Ceci devient problématique si nous ne pouvons pas estimer le niveau de chevauchement entre deux activités successives, ce qui peut induire a des erreurs de classification et d'interprétation [195-218].

Nous proposons dans cette partie une étude permettant d'effectuer une meilleure ségmentation de deux activités succéssives en prenant en compte deux niveaux de ségmentations : (i) sans chevauchement entre les deux activités succéssives et (ii) avec chevauchement entre deux activités succéssives. Nous proposons dans cette partie une approche permettant de sélectionner dynamiquement le niveau de chevauchement afin d'avoir les meilleures performances de classification.

4.5.1. Configuration ne présentant pas de chevauchement entre activités succéssives

Nous commençons par évaluer notre modèle SVM à noyau Gaussien sur des données d'activités physiques élémentaires sans prise en compte du chevauchement. Puis, nous proposons de sélectionner des pourcentages différents des fenêtres de chevauchement entre deux activités physiques élémentaires. Nous avons entrainé un SVM à noyau Gaussien sur les données des activités élémentaires avec une validation croisée 10 fois. Les résultats de classification sont montrés dans **les tableaux 15** et **16**.

Tableau 15: Précision de reconnaissance des activités élémentaires sans chevauchement entre activités.

	Debout	Assis	Allongé	marcher
SVM	90%	91%	91%	90%

Pour une analyse plus détaillée, nous proposons de calculer la matrice de confusion de ces activités. Les résultats sont présentés dans le **tableau 16.**

Tableau 16 : Matrice de confusion entre les activités physiques élémentaires.

	Debout	Assis	Allongé	Marcher
Debout	100%	17%	1%	12%
Assis	15%	100%	1%	8%
Allongé	14%	12%	100%	8%
Marcher	4%	3%	3%	100%

Nous remarquons que certaines activités sont confondues. Par exemple l'activité debout est confondue avec l'activité Assis (17%), Allongé (1%) et Marcher (12%). L'activité "Assis » est confondue avec l'activité « debout » (15%), « Allongé » (1%) et « Marcher » (12%). L'activité « Allongé » est confondue avec l'activité « Debout » (14%), l'activité « Assis » (12%) et l'activité « Marcher » (8%). Enfin, l'activité « Marcher » est confondue avec l'activité « Debout » (4%), l'activité « Assis » (3%) et l'activité « Allongé » (3%).

Les activités « debout », « assis » et « allongé » sont souvent confondues entre elles avec des pourcentages assez variables.

4.5.2. Configuration présentant un chevauchement entre activités succéssives

Afin de prendre en compte ce niveau de chevauchement, nous avons mené une étude sur plusieurs tailles à prendre en compte.

Nous avons choisi de tester 10 valeurs différentes de tailles de chevauchement entre deux activités élémentaires successives**. La figure 20,** présente les résultats obtenus.

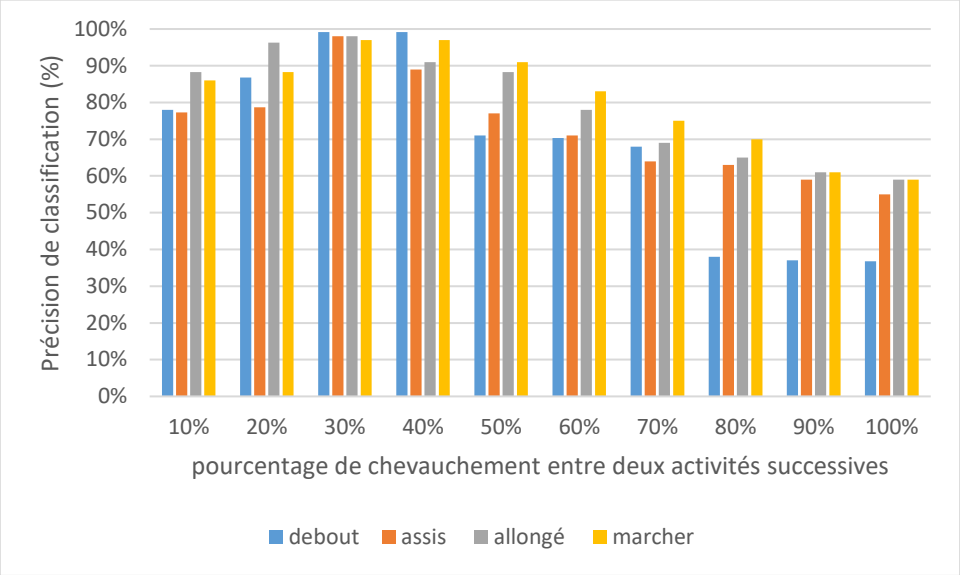


Figure 20 : Evolution du taux de chevauchement entre deux activités successives en fonction de la précision de classification.

Nous remarquons que plus le taux de chevauchement augmente plus la précision de reconnaissance augmente, jusqu'à atteindre une valeure maximum. Une fois cette valeure atteinte, la précision chute. Nous constatons que la valeure optimale est de 40% des données de chevauchement. A partir de cette valeure optimale, nous remarquons que les performances de classification se dégradent. Ceci peut être expliqué par le fait que l'algorithme entre dans une phase de sur-apprentissage où les performances se voient dégradées. D'où l'intérêt de prendre 40% comme taux de chevauchement entre deux activités successives.

5. Discussion

La non prise en compte du chevauchement permet d'avoir des résultats de classification élévés d'environs 90% de précision de classification. Toutefois, ceci cause entre autre une perte de l'information relative aux transitions, si nous les supprimons, comme c'est cas de plusieurs travaux. A contrario, la prise en compte des chevauchements permet d'atteinte une meilleure précision de classification (91% en moyenne). Ceci peut etre expliqué par l'apport de l'information de transition entre deux activités succéssives.

L'approche de ségmentation dynamique de deux activités succéssives s'avère efficace. En effet, elle permet de choisir avec précision les moments de débuts et de fin de chaque activité et ne pas avoir à effectuer plusieurs tests.

L'utilisation de la DCT avec un paramétrage adéquat a permis d'avoir de meilleurs résultats de classification. Notamment le travail référencé dans [99,107] qui ont obtenu une précision de l'ordre de 88% pour classifier des activités physiques dans un environnement contrôlé.

L'état de l'art a montré que les signaux accélérométriques étant très corrélés [108]. De plus, l'ajout d'autres signaux acceléromètriques peut rendre ce caractère très persistant en particulier dans le cas d'un chevauchement d'activités physiques [107] [109]. De plus, ceci peut causer plusieurs problèmes au procéssus de reconnaissance de l'activité. En effet, comme plusieurs travaux l'ont montré, l'apprentissage des données d'activités, avec un réseau de neurones, précisément, constitué principalement de données accélérométriques et éventuellement gyroscopes et magnétomètres a besoin d'un minimum de régularisation afin d'éviter d'éventuel problème de disparition/explosiion des gradients, sur-apprentissage, etc (dropout, dans la plupart des travaux), qui sont les principaux problèmes. L'état de l'art a montré que ceci peut retarder l'entrainement des modèles d'apprentissage et d'influencer négativement les résultats de classification, précisément dans les modèles d'apprentissage profonds (Cf. Chapitre 5).

Nos travaux montrent qu'il serait préférable d'avoir un descripteur permettant de décorréler les signaux, afin de permettre d'extraire efficacement des descripteurs représentatifs de l'activité tout en gardant le caractère compressif de données. D'autre part, le fait que les signaux soient corrélés ne donne pas forcement une meilleure précision de classification. Ainsi, une autre étude [110] a montré que ce caractère influence fortement la performance et la complexité de calcul des algorithmes d'apprentissage automatique. La DCT répond parfaitement à ces caractéristiques ce qui explique sa capacité à influencer positivement la performance de classification. De plus, l'état de l'art a montré que la DCT possède la particularité de regroupement de l'énergie [111-112], dans les coefficients basses fréquences grâce à son approximation de la transformée de Karhunen-Loève [113] donc de l'analyse en composantes principales, ce qui explique notre choix de la taille du descripteur calculée qui résulte de ce regroupement de l'énergie dans les coefficients basse fréquence.

Nous avons experimenté une méthode de classification non linéaire en utilisant un perceptron multicouche. Les résultats de cette étude sont présentés dans l'annexe B.

Conclusion

Dans ce chapitre, nous avons introduit le concept d'activité physique élémentaire, qui représente les activités : « debout », « assis », « allongé » et « marcher » indépendamment de tout contexte et évènement extérieur.

Nous avons proposé un nouveau corpus de données des activités de 33 personnes, utilisant 3 objets connectés du quotidien. Quelques résultats statistiques ont été présentés sur ce corpus.

Nous avons pu mettre en œuvre une méthodologie permettant d'extraire, tester et analyser la DCT. Nous avons ensuite proposé une approche pour choisir les meilleurs paramètres de la DCT permettant d'avoir les meilleurs résultats de reconnaissance de l'activité physique.

Ce chapitre a permis de mettre la lumière sur la problématique du chevauchement entre deux activités élémentaires successives. En effet, notre étude a montré l'intérêt du chevauchement entre deux activités succéssives afin de pouvoir les distinguer les instants de débuts et fin de chaque activités. Deplus, ceci permet d'avoir les meilleures performances de reconnaissance en prenant en compte les transitions entre deux activités.

A travers cette étude, nous avons pu mettre en place une méthodologie permettant d'obtenir une meilleure précision de classification des activités élémentaires. Toutefois, cette approche reste limitée. En effet, l'état de l'art a montré que les activités physiques sont, en général, effectuées séquentiellement, l'une à la suite de l'autre, ainsi que parallélement. Par exemple, une personne assise, peut parler au téléphone en regardant la TV en meme temps. L'activité « assis » est effectuée parallement aux activités « parler et téléphone » et « regarder la TV ». Le fait de parler au téléphone et de regarder la TV permet d'apporter certaines informations appelées : le contexte dans lequel se déroulent les différentes activités. De plus, les objets connectés utilisés contiennent d'autres types de capteurs qui n'ont pas été explorés dans cette étude. Nous proposons d'aborder ce sujet dans le chapitre 3.

Chapitre 3

Reconnaissance de l'activité physique basée sur le contexte

1. Introduction …… ..... …94
2. Etat de l'art ……..…96
3. Activités de premier niveau ..... ….99
3.1. Statistiques descriptives sur le corpus d'activités ..... …99
3.2. Annotation des données ……..... …99
3.3. Pré-analyse du corpus …. ..... …100
4. Processus de classification des activités physiques. …..... …..101
5. Activités physiques de deuxième niveau …..... …106
5.1. Annotation …..... …106
5.2. Résultats …..... …106
6. Activités physiques de troisième niveau : détection de l'attention envers la TV …… .. …109
6.1. Segmentation des individus…….. …..112
6.2. Descripteur de l'attention envers la TV……..... …112
7. Synthèse des chapitres 2 et 3…… 113
8. Limites des études effectuées dans les chapitres 2 et 3… ..... …116
Conclusion…..... …127

Chapitre 3

1. Introduction

L'activité physique est, en général, liée à l'environement dans lequel elle se déroule. Par exemple, le fait d'etre assis en regardant la TV est contextuellement différent du fait d'etre assis en mangeant. Du point de vu technique de détection de l'activité physique, l'état de l'art a montré que les modèles de reconnaissance de ces deux types d'activités sont identiques. Ainsi, la sortie d'un tel type de modèle appliqué à ces deux types d'activité serait l'activité « assis ». En revanche, ce type de protocole de traitement ne nous donne aucune information sur l'environement ou le contexte dans lequel se déroulent les deux activités citées ci haut.

La prise en compte du contexte dans lequel se déroule l'activité physique nécéssite l'incorporation d'autres sources de données (ex.capteurs) et d'autres techniques de traitements efficacent. En effet, la détection de l'activité « assis en mangeant » nécéssite de détecter d'abord l'activité « assis », comme nous l'avons fait dans le chapitre 2, puis détécter le fait que la personne est « entrain de manger ». Il y a donc deux niveaux de granularité de détection : un premier niveau, qualifié de bas niveau, consiste à détecter l'activité physique « assis » et le deuxième niveau, qualifié de haut niveau, consiste à détecter l'activité « entrain de manger ». Cependant, si nous souhaitons pousser le raisonnement et savoir s'il regarde la TV avec attention, il faudrait introduire un troisième niveau de granularité. L'association entre les trois niveaux de granularité est appelée : activité physique basée sur le contexe.

Dans le cadre de nos travaux, nous proposons une architecture à trois niveaux de granularité de reconnaissance des activités physiques.

- **Un premier niveau** (bas niveau) concerne toutes les activités que les participants effectuent pendant la durée d'enregistrement (6 mois). Ces activités peuvent etre élémentaires et des activités plus complexes dépendantes d'un contexte (debout en parlant au téléphone, assis en mangeant, etc).
  Le procéssus de reconnaissance des activités de ce premier niveau concerne la reconnaissance des activités physiques en prenant en compte toutes les données de l'ensemble les capteurs des trois objets connectés. C'est donc une reconnaissance d'activités que nous qualifions d'ordinaires, comme « debout en parlant au téléphone », « assis en regardant la TV », etc. Le processus de reconnaissance utilisera l'annotation déjà faite sur les activités élémentaires (Cf. Chapitre 2) en ajoutant les autres données des autres capteurs. Puis, une extraction des principaux descripteurs sera faite sur l'ensemble des données. Enfin, une classification par apprentissage automatique de l'ensemble de ces activités sera effectuée.

- **Un deuxième niveau** (niveau intermédiaire) concerne la reconnaissance des activités résultantes de l'interaction avec l'écran de la Smart TV. Ce niveau d'activité prend en compte le type d'activité effectuée pendant l'interaction avec la smart TV (Allumer la TV en étant assis, manger en augmentant le volume de la TV, etc) en ajoutant une information supplémentaire qu'est la phase de visionnage d'un programme TV. En effet, nous avons souhaité ségmenter les phases de visionnage d'un programme TV en trois phases : avant, pendant et après avoir visionné un programme TV. Cette information supplémentaire permet de classer les activités effectuées pendant ces trois phases en un niveau supérieur : Le deuxième niveau.

  Le processus de reconnaissance des activités physiques est similaire au processus effectué dans le premier niveau. La différence se situe au niveau de la prise en compte du déroulement de l'activité physique (avant, pendant et après).
- **Un troisième niveau** (de haut niveau) : l'interaction avec l'ecran de la smart TV permet de donner des informations sur l'attention de l'utilisateur envers la TV. Ce type d'activités nécéssite non seulement de savoir que le visionneur est entrain de visionner un programme TV, et donc pendant le visionnage d'un programme TV. De plus, cela nécéssite aussi de savoir s'il est attentionné ou pas pendant cette phase. C'est donc une information supplémentaire apportée aux activités de deuxième niveau. Nous appelons ce type d'activités : activité de troisième niveau.

La figure 1 donne une vue d'ensemble du processus de traitement et des objectifs visés par ce chapitre.

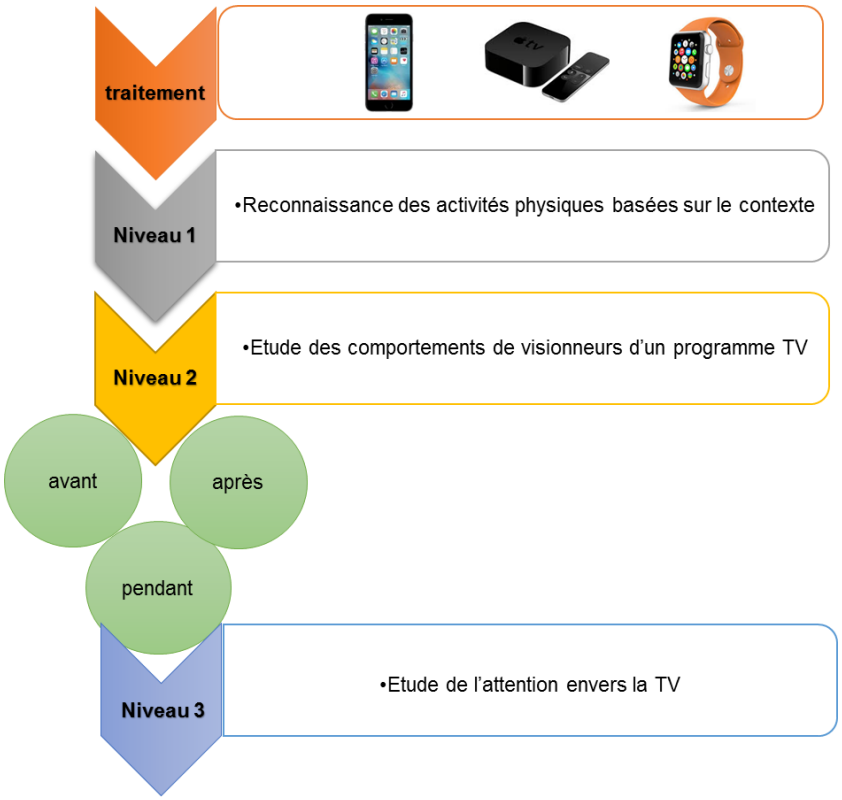


Figure 1 : Vue d'ensemble du processus de traitement et reconnaissance des différents niveaux d'activités basées sur le contexte.

Nous aborderons dans la partie suivante l'état de l'art sur la prise en compte de la notion de contexte pour la reconnaissance des activités physiques et les différents travaux connexes.

2. Etat de l'art

Plusieurs travaux ont abordé le concept de l'activité basée sur le contexte. Certains travaux définissent le contexte en faisant référence à l'activité physique (gestuelle, activité élémentaire, comportement, posture, corpulence, etc.), couplée avec la notion d'usage de l'objet connecté [119]. La notion d'usage a été définie comme étant l'utilisation d'une ou plusieurs fonctionnalités propres à l'objet connecté. Par exemple, le smartphone pour téléphoner, ou la smartwatch pour répondre aux appels téléphoniques ou consulter sa fréquence cardiaque [120]. Ceci étant sans faire abstraction de l'activité physique effectuée. Un cadre experimental a été proposé dans [71] pour la reconnaissance de certaines activités basées sur le contexte en utilisant plusieurs smartphones. Ils ont considéré que les activités physiques basées sur le contexte dépendent des activités effectuées en prenant en compte les usages des différents smartphones ainsi que les façons de les tenir et de s'en servir (téléphoner en mettant l'appareil en mode haut-parleur ou en approchant le téléphone de la tete, etc.).

L'état de l'art a montré que, quel que soit l'état physique d'un individu, il est nécessairement subdivisible en plusieurs activités (gestuelles, activités élémentaires ou basiques, etc.) [114]. Par exemple, l'activité « travailler sur PC », inclus la gestuelle de la main et la manière de taper sur un clavier, l'activité « assis », l'activité « assis en parlant » (s'il parle en même temps), etc. [57]. Cet ensemble d'activités n'est pas reconnu de la même manière par la plupart des travaux. En effet, l'activité « travailler sur PC » est parfois reconnue comme simplement « assis », parfois « assis en mettant les mains sur la table », etc.[115]. Alors que toutes ces activités font référence à une seule activité avec des niveaux de granularité différents. Par exemple : « travailler sur PC », peut contenir l'activité « assis » et l'activité « taper sur un clavier ». Ainsi, elle peut contenir l'activité « attentionnée ou regarder le clavier en écrivant »[116]. Ce sont donc trois niveaux de granularité différents. Nous proposons dans le tableau 1 un aperçu des principaux travaux sur les systèmes de reconnaissance de l'activité physique. Nous avons recensé le nombre d'activités détectées, le type de capteurs utilisés, les descripteurs séléctionnés et la précision de reconnaissance atteintes. Enfin, nous nous sommes intéréssés à savoir quels ont été les contextes dans lequels se sont déroulées les différentes activités ainsi que les niveaux de granularité abordés dans chaque étude.

Tableau 1 : Synthèse de l'état de l'art sur les principaux systèmes de reconnaissance de l'activité

Ref.	Activités (#)	capteurs	descripteurs	précision	Contexte	Niveau de granularité
Ermes[37]	5	Acc (poignet, genou, taille)	TD,FD	94%	Comportement dans une maison intelligente	1 ou indéfini
eWatch[59]	6	Acc, ENV (poignet)	TD,FD	94%	Comportement dans une maison intelligente	1 ou indéfini
Tapia[78]	30	HRM, Acc (5 places)	TD,FD,HB	86% 56% 56%	Comportement dans une maison intelligente	1 ou indéfini
Vigilante [77]	5	Acc, VS (poitrine)	TD,FD,TF	96.8% 92.6%	Comportement dans une maison intelligente	1 ou indéfini
Kao[57]	7	Acc (poignet)	TD,LDA	94.71%	Comportement dans un laboratoire	1 ou indéfini
Brezmes[22,23]	5	Acc (téléphone)	TD,FD	80%	Omportement dans un laboratoire	1 ou indéfini
COSCAR[71]	20	GPS, acc (watch, phone)	TD	93%	Comportement dans un laboratoire	1 ou indéfini
ActiServ[26,28]	11	Acc (phone)	$\overline{y}$,σ_y^2	71%-98%	Comportements dans une maison intelligente	1 ou indéfini
[128]	Attention envers la TV	Eye tracker	Statistiques	70-78%	Comportement devant la TV	1 ou indéfini

L'état de l'art a montré que la reconnaissance des activités physiques basées sur le contexte nécessite l'utilisation de plusieurs types de capteurs et objets connectés. La détection des activités telle que l'attention envers la TV utilisent des objets non adaptés dans des environnements contrôlés sans prise en compte de l'activité physique inhérente à l'état de l'attention envers la TV [128]. Ainsi, ces travaux ne prennent pas en compte les activités physiques induites par l'interaction des participants avec la TV (assis, debout, en manipulant un objet connecté, en parlant avec quelqu'un, etc.), ce qui limite la compréhension de l'activité effectuée. Un exemple illustratif de ce propos est données en annexe A.

Enfin, l'état de l'art a montré que la plupart des travaux sur la reconnaissance des activités physiques basées sur le contexte n'ont pas abordé suffisamment la notion de granularité dans les activités physiques, ce qui limite grandement leurs compréhensions en pariculier lorsqu'il s'agit d'activités complexes (assis en regardant la TV et en mangeant, etc) ou tout simplement d'étudier un phénomène sous jacent à la reconnaissance de l'activité (Cf Annexe A).

Afin de répondre à ces limitations, nous proposons dans ce chapitre, une architecture à trois niveaux pour la reconnaissance de l'activité physique basée sur le contexte. Nous considérons la reconnaissance des activités utilisant que les données capteurs embarqués dans ces objets connectés et leurs usages, appartenant au premier niveau (bas niveau). La reconnaissance des activités ou comportements des visionneurs d'un programme TV pendant les phases de visionnage de la TV, comme des activités appartenant à un niveau plus élevé, le deuxième niveau. En effet, ces activités impliquent une interaction avec l'écran de la smart TV, en se focalisant sur les activités physiques effectuées pendant ces trois phases. Enfin, la détéction de l'attention envers la TV comme étant des activités de troisième niveau.

Nous proposons, dans la partie suivante, de reconnaitre les activités physiques de premier niveau, ainsi définies.

3. Activités de premier niveau

Nous rappelons que les activités physiques du premier niveau concernent toutes les activités physiques effectuées par les 33 participants pendant 6 mois. Deux points essentiels différencient entre les activtés du premier niveau et les activités élémentaires. Premièrement, les activités élémentaires sont au nombre de quatre : debout, assis, allongé et marcher. Les activités basées sur le contexte sont au nombre de 18 (voir section suivante). Deuxièmement, les activités physiques basées sur le contexte sont détectées en utilisant la totalité des capteurs des trois objets connectés, alors que les activités élémentaires sont détectées en utilisant trois types de capteurs : les accéléromètres, les gyroscopes et les magnétomètres.

Nous commencerons par établir quelques statistiques descriptives sur le corpus de données. Puis, nous proposons une nouvelle approche d'annotation des activités physiques du premier niveau. Ensuite, un ensemble de descripteurs à donner en entrées des algorithmes d'apprentissage va etre extrait.

3.1. Statistiques descriptives sur le corpus d'activités

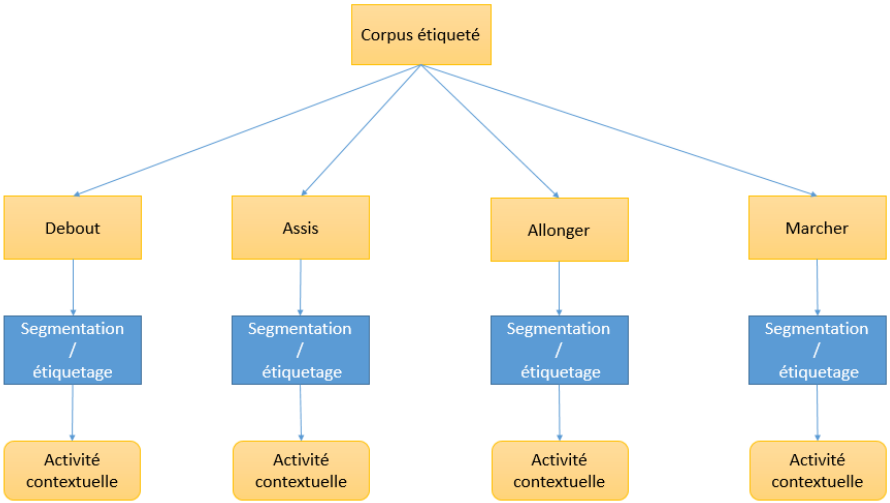
Nous rappelons que l'environnement principal du déroulement des activités a impliqué 33 participants chez eux pendant 6 mois. Les participants interagissent avec la Smart TV tout le long de cette expérience.

3.2. Annotation des données

L'étiquetage du corpus a été effectué par trois personnes différentes parmi les 33 participants. Afin de pouvoir effectuer l'étiquetage/annotation des données du corpus, nous avons observé le comportement de tous les participants pendant la durée de l'expérience.

Le corpus de données a déjà été étiqueté dans le chapitre 2, en quatre étiquettes : debout, assis, allonger et marcher. Il a subi un deuxième étiquetage plus fin. Au sein de chaque activité physique élémentaire, nous avons rajouté des étiquettes correspondantes au contexte du déroulement de l'activité.

En effet, prenons un exemple d'un participant dont le signal accélérométrique entre les instants t_i et t_{i+1} ait été étiqueté comme étant debout. Supposons que pendant cet intervalle de temps, le participant parle au téléphone, puis regarde sa montre pour voir sa fréquence cardiaque, etc. L'intervalle $[t_i, t_{i+1}]$ sera subdivisé en plusieurs sous intervalles correspondants aux différents comportements du participant entre les instant t_i et t_{i+1} (parler au téléphone, regarder sa montre, etc.). Ainsi, l'opération d'étiquetage se termine à la fin de l'activité physique élémentaire en cours. **La figure 2** donne une vue d'ensemble explicative de notre approche d'étiquetage. **Le tableau 2** représente l'ensemble des activités basées sur le contexte étiquetées.



Figue 2 : Démarche d'étiquetage des activités physiques basées sur le contexte.

Tableau 2 : Ensemble des activités physiques basées sur le contexte détectées.

N° activité	Activités
#1	Allumer la TV {debout, assis, allonger, marcher}
#2	Manipuler IPhone en étant {debout, assis, allongé}
#3	Manger en étant assis
#4	Zapper en parlant et assis
#5	Zapper en parlant et debout
#6	Zapper en parlant et allongé
#7	Zapper en manipulant l'IPhone
#8	Manger en parlant avec quelqu'un
#9	Manger en marchant
#10	Manger étant debout
#11	Zapper étant debout
#12	Zapper étant assis
#13	Zapper étant allongé
#14	Rester sur la même chaine de TV en travaillant sur PC
#15	Augmenter/baisser le volume
#16	Rester sur la même chaine TV en manipulant l'IPhone
#17	Eteindre la TV en étant allongé
#18	Eteindre la TV étant assis/debout/marcher

3.3. Pré-analyse du corpus

Afin d'avoir une idée de la distribution des activités physiques basées sur le contexte par rapport aux activités élémentaires, nous proposons de représenter graphiquement la répartition de ces activités par rapport à chaque activité physique élémentaire dont elle est issue (voir **figure 3**).

Nous remarquons que les tendances sont assez variables selon les différentes activités physiques élémentaires. En effet, l'activité #1 (Allumer la TV en étant soit debout, assis, entrain

de marcher ou allongé) est beaucoup plus présente dans l'activité physique élémentaire (Allonger), avec 85%, que dans d'autres activités physiques élémentaires : marcher, assis et debout avec 1%, 12% et 2% respectivement. Alors que l'activité #14 (rester sur la même chaine en travaillant sur PC) est majoritairement effectuée en étant assis.

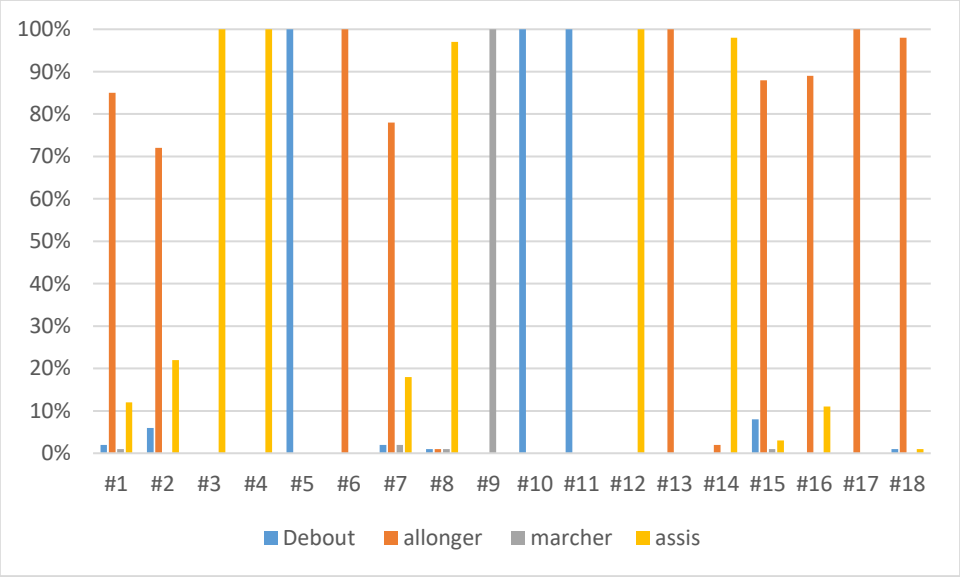


Figure 3 : Distribution des activités physiques basées sur le contexte aux seins des activités élémentaires.

4. Processus de classification des activités physiques

Afin de reconnaitre les activités physiques basées sur le contexte listées dans le tableau 1, l'état de l'art suggère de considérer l'aspect portabilité des trois objets connectés [102-106]. En effet, comme l'environnement de l'expérience est totalement non contrôlé, les participants n'ont subi aucune recommandation. Ainsi, ils peuvent, naturellement déplacer l'un ou tous les objets connectés d'une position à une autre pendant l'expérience. Quatre principaux positionnements des objets connectés sur le corps ont été récoltés :

- Tous les objets connectés sont portés par les participants.
- Un des objets n'est pas porté par les participants.
- Deux objets ne sont pas portés par les participants.
- Trois objets ne sont pas portés par les participants.

La figure 4, représente la distribution des activités basées sur le contexte en fonction de la portabilité des objets connectés.

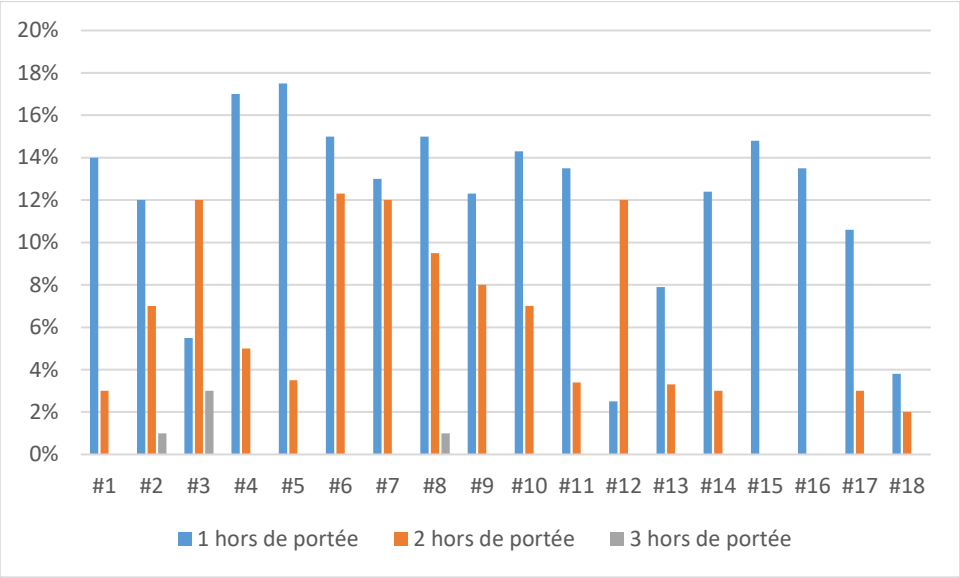


Figure 4: Distributions des activités basées sur le contexte selon la portabilité des objets connectés.

Les trois objets connectés ne sont que très rarement « hors de portée » simultanément. Ainsi, les durées de « hors de portage » sont relativement petites, de l'ordre de moins de 20% du temps d'enregistrement.

4.1. Extraction de descripteurs

Pour rappel, nous avons extrait les descripteurs renseignés dans le **tableau 3**. Ces descripteurs sont les plus utilisés dans l'état de l'art (Cf. chapitre 1).

Pour les données des accéléromètres, gyroscopes et magnétomètres, nous avons gardé la DCT comme principal descripteur. Ce descripteur a montré son efficacité en classification d'activités physiques élémentaires (Cf. Chapitre 2).

Tableau 3 : Les principaux descripteurs extraits de tous les capteurs des trois objets connectés.

Capteur	Descripteur extrait
Accéléromètre, gyroscope, magnétomètre	La DCT
Détecteur de proximité	Moyenne, écart type
Capteur de luminosité ambiante	Moyenne, écart type
Capteur d'empreinte digitale	Moyenne, écart type
Microphone	DCT
Cardiofréquencemètre	La variation de la fréquence cardiaque
Surface tactile	Moyenne, écart type

4.1.1. Résultats : Reconnaissance de l'activité physique élémentaire

Nous avons vu que l'annotation des activités basées sur le contexte est issue de celles des activités physiques élémentaires. Nous proposons donc de reconnaitre d'abord les activités élémentaires, en utilisant tous les capteurs des trois objets connectés, puis, de discuter les résultats sur les activités basées sur le contexte.

Les descripteurs ainsi extraits ont été concaténés pour former un vecteur final prêt pour l'apprentissage. Comme nous avons vu dans le chapitre 2, nous avons cinq algorithmes à entrainer: MLP (Perceptron Multicouche avec 1 couche d'entrée correspondant aux descripteurs extraits, deux couches cachées et enfin une couche de sortie avec quatre neurones correspondant aux quatre activités physiques debout, assis, allongé et marcher, SVM (support vector machines, avec un noyau Gaussien), RF (forets aléatoires), Naive Bayes (NB) et l'algorithme des arbres de décision (DT). La méthode de 10 fold cross validation a été adoptée pour la phase de validation. **Le tableau 4** montre les résultats obtenus.

Tableau 4 : Résultats de classification des activités physique élémentaires en utilisant toutes les données de tous les capteurs des trois objets connectés.

	MLP	¨SVM	RF
Debout	95.78%	98.13%	98.14%
Assis	96.64%	99.4%	98.27%
Marcher	96.59%	99.04%	98.32%
Allonger	95.11%	99.17%	98.17%

Nous remarquons une amélioration des résultats de classification des activités physiques élémentaires. L'algorithme du SVM avec un noyau Gaussien a permis d'atteindre 99% de précision de classification, en moyenne. Ceci est en adéquation avec l'état de l'art. En effet, l'ajout de données des autres capteurs permet d'améliorer la précision de classification. Nos résultats sont comparables à ceux de l'état de l'art, notamment le travail de (Inoue et al., 2018)[130] ayant utilisé un modèle basée sur les réseaux de neurones et ont obtenu une précision de classification de 95% en utilisant seulement un smartphone. Ils ont utilisé les données brutes de l'accéléromètre alors que l'environnement était contrôlé.

Afin d'expliquer ce résultat, nous détaillons dans **le tableau 5** les précisions de classification des données de chaque objet connecté.

Tableau 5 : Précision de classification des activités physiques élémentaires par rapport à chaque objet connecté.

	Smartphone	Smartphone, smart Watch	Smartphone, smart Watch et TV
Debout	95.17%	98.16%	99.13%
Assis	97.01%	97.99%	99.4%
Marcher	97.8%	98.01%	99.04%
Allongé	97.99%	98.4%	99.17%

Nous remarquons que l'ajout des données de chaque objet connecté apporte de plus en plus de la précision. Les données de la smartwatch et de la smart TV ont contribué à ces résultats.

En effet, l'activité debout, par exemple, se voit passer d'une précision de reconnaissance de 95.17% à 98.16% en ajoutant les données de la smartwatch. Alors que l'ajout des données de la Smart TV fait grimper la précision jusqu'à 99.13%. L'écart de précision entre ce qu'apport la smartwatch comme information est plus important que l'information apportée par la Smart TV. Ceci peut s'expliquée par le fait que la smartwatch est en permanence sur le bras des participants et que l'algorithme d'apprentissage utilisé ayant bien appris avec ce type de données.

4.1.2. Résultats : Reconnaissance des activités du premier niveau

Après avoir montré que les activités élémentaires sont bien détectées en ajoutant les autres données des trois objets connectés, nous proposons de poursuivre la reconnaissance des activités physiques, en reconnaissant toutes les activités issues de chaque activité physique élémentaire (**Figure 5**). Nous rappelons que pour détecter les activités physiques du premier niveau, nous nous servons de la ségmentation déjà faite sur les activités physiques élémentaires. Quatre étiquettes ont déjà été établies : debout, assis, allongé et marcher. Chaque étiquette représente « une plage horaire » ou une fenetre temporelle. Au sein de chaque fenetre temporel, nous procédons à une seconde ségmentation permettant de détecter les activités physiques (Figure 5).

Dans cet objectif, nous avons donc extrait des descripteurs de toutes les données des capteurs des trois objets connectés. Les descripteurs ainsi extraits, ont été concaténés pour former un seul vecteur prêt pour l'apprentissage. Nous avons retenu à nouveau l'algorithme du SVM avec un noyau Gaussien pour la reconnaissance des activités basées sur le contexte.

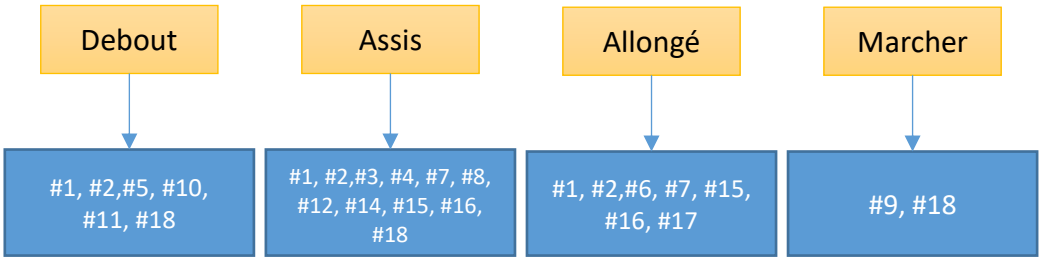


Figure 5 : Vue d'ensemble sur la déduction de l'activité basée sur le contexte à partir des activités physiques élémentaires.

Le modèle SVM à noyau Gaussien a été entrainé sur la totalité des données disponibles. La **figure 6** présente les résultats de classification obtenus.

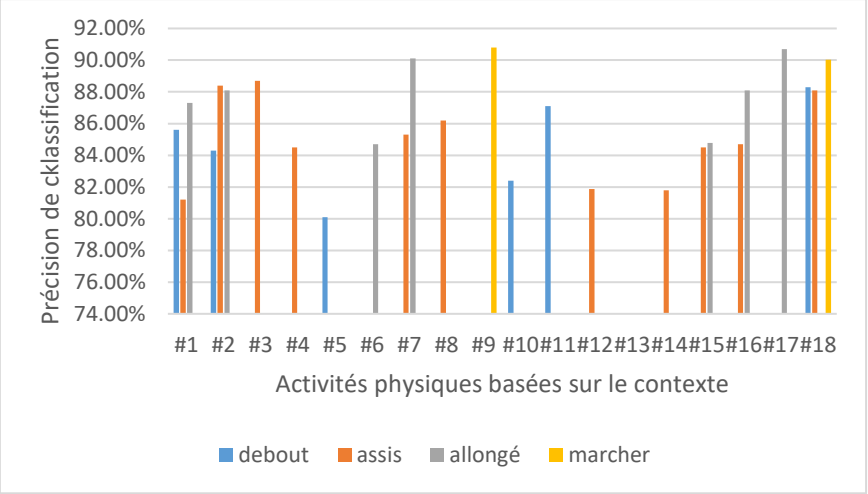


Figure 6 : Précisions de classification des activités basées sur le contexte au sein de chaque activité élémentaire.

Les résultats montrent que'au sein de chaque activité élémentaire, les activités physiques basées sur le contexte sont bien classées. Ainsi, en connaissant les activités physiques basées sur le contexte, nous serons amenés à reconnaitre l'état physique du participant, c'est-à-dire, s'il est debout, assis, allongé ou entrain de marcher. Cette classification est rendue possible grâce à l'information apportée par chaque objet connecté. En effet, nous pourrons regarder de plus près l'impact de l'information apportée par chaque objet connecté sur le taux de reconnaissance des activités basées sur le contexte. Les résultats de cette étude est donnée dans l'histogramme de la **figure 7**.

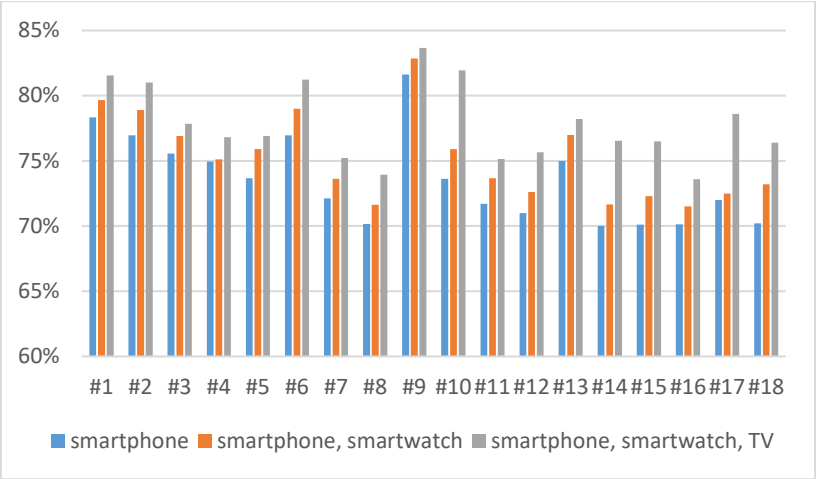


Figure 7 : Résultats de reconnaissance des activités physiques basées sur le contexte. Le but de ce graphique est de montrer la contribution de chaque objet connecté à la précision de reconnaissance des activités physiques basées sur le contexte. Le fait de considérer que le Smartphone seul, la smartwatch seule ou la Smart TV seule ou bien les prendre tous ensemble influ sur la précision de classification.

En comparant les résultats des **figures 6 et 7**, nous remarquons que le taux de détection des activités physiques basées sur le contexte est nettement moins élevé que celui obtenu pour les activités élémentaires. Ce résultat est du à: premièrement, le fenêtrage des données prises pour l'apprentissage est plus important dans le cas de l'activité élémentaire que dans le cas des activités physiques basées sur le contexte. Ceci permet à l'algorithme d'avoir plus de données d'apprentissage pour l'activité élémentaire que pour les activités basées sur le contexte, ce qui influe sur la qualité de l'apprentissage. Deuxièmement, certaines activités basées sur le contexte ne sont pas bien standardisées pour pouvoir les prendre comme référence pour l'apprentissage. Par exemple, l'activité « manger en parlant avec quelqu'un », peut être effectuée de diverses manières et ne sont pas toutes étiquetées entant que « manger en parlant avec quelqu'un». Troisièmement, la limitation de l'ensemble de données d'apprentissage pour certaines activités basées sur le contexte. Ce dernier problème est central : la limitation de l'apparition de certaines activités de contexte pousse les algorithmes d'apprentissage à apprendre par cœur ces données et sont donc incapables de bien généraliser.

Enfin, la contribution des trois objets connectés à détecter des activités basées sur le contexte est très visible.Le fait de prendre les trois objets connectés tous ensemble, a permis d'atteindre une meilleure performance de reconnaissance (**Figure 7**).

5. Activités physiques de deuxième niveau : comportement des visionneurs d'un programme TV

Les activités physiques basées sur le contexte définies dans la séction 4 ne donnent pas d'informations sur le moment de l'interaction avec la Smart TV. En effet, nous savons, d'après le procéssus de classification des activités du premier niveau, que les activités ont été effectuées dans les domiciles des participants. Les activités physiques telles que #8 Manger en parlant avec quelqu'un ne donne aucun renseignement sur la phase de visionnage de la TV dans laquelle elle s'est déroulée. Nous souhaitons donc savoir si cette activité s'est déroulée avant, pendant ou après avoir visionné un programme TV. C'est donc une information supplémentaire par rapport aux activités physiques du premier niveau. Nous avons donc qualifié ce type d'activités de **deuxième niveau**.

5.1. Annotation

Nous nous intéressons à présent à reconnaitre leurs comportements durant trois phases de visionnage d'un programme TV.

Nous commençons par annoter les données des activités suivant ce nouvel objectif. **Le tableau 6** représente les résultats de l'étiquetage.

Tableau 6 : Etiquetage des activités physiques basées sur le contexte suivant la phase de visionnage d'un programme TV.

N° activité	Activités	Avant	Pendant	Après
#1	Allumer la TV {debout, assis, allonger, marcher}	✓		
#2	Manipuler IPhone en étant {debout, assis, allongé}	✓		✓
#3	Manger en étant assis	✓	✓	✓
#4	Zapper en parlant et assis			✓
#5	Zapper en parlant et debout		✓	✓
#6	Zapper en parlant et allongé		✓	
#7	Zapper en manipulant l'IPhone	✓	✓	
#8	Manger en parlant avec quelqu'un		✓	
#9	Manger en marchant	✓	✓	
#10	Manger étant debout	✓	✓	
#11	Zapper étant debout	✓	✓	
#12	Zapper étant assis	✓	✓	

#13	Zapper étant allongé	✓	✓	
#14	Rester sur la même chaine de TV en travaillant sur PC		✓	
#15	Augmenter/baisser le volume	✓	✓	
#16	Rester sur la même chaine TV en manipulant l'IPhone	✓		✓
#17	Eteindre la TV en étant allongé	✓		✓
#18	Eteindre la TV étant assis/debout/marcher			✓

Nous remarquons qu'il y a bien des activités spécifiques à chaque phase de visionnage (Allumer la TV {debout, assis, allonger, marcher}, par exemple). Toutefois, d'autres activités appertiennent aux différentes phases (Zapper étant assis, par exemple).

Nous souhaitons savoir s'il existe une activité basée sur le contexte pouvant caractériser une phase particulière de visionnage d'un programme TV, telle qu'une personne qui regarde les informations télévisées en étant assise en mangeant. Nous cherchons à savoir si cette activité caractérise une phase de visionnage particulière (phase pendant le visionnage) et si on peut l'associer à un programme TV particulier (ici le journal télévisé).

Pour cela, nous avons appliqué un algorithme d'apprentissage automatique (SVM) pour classifier les trois phases de visionnage, après avoir extrait les mêmes descripteurs que précédemment et effectué les mêmes paramétrages. Les résultats sont détaillés dans la partie suivante.

5.2. Résultats

Dans cette séction, nous présentons les résultats de reconnaissance des activités physiques du deuxième niveau. Nous commençons par une étude sur la séparabilité des trois phases de visionnage d'un programme TV. Celle-ci consiste à montrer qu'il existe des activités pouvant distinguer chaque phase de visionnage d'un programme TV.

5.2.1. Distinction entre les trois phases de visionnage d'un programme TV

La première étape consiste à étudier la séparabilité des trois phases de visionnage. Ceci peut aider à savoir s'il y a des comportements pouvant les distinguer. L'entrainement de l'algorithme du SVM à noyau Gaussien en utilisant la totalité des données d'activités donne les résultats de la **figure 8.**

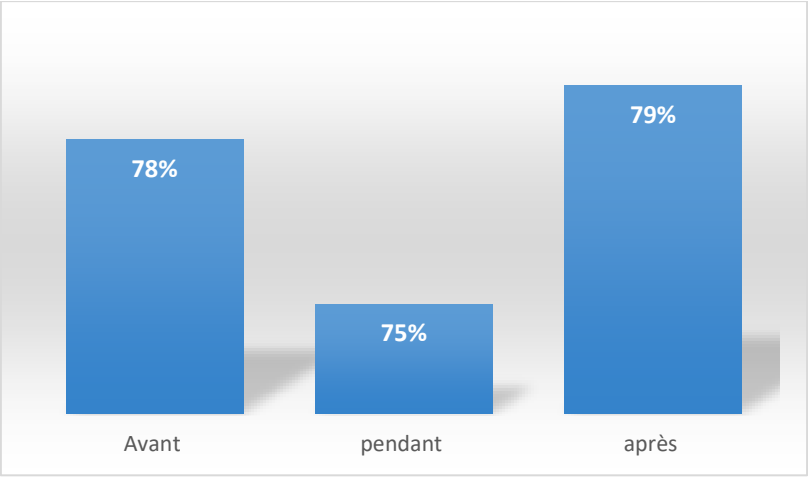


Figure 8 : Précision de reconnaissance des phases de visionnage d'un programme TV.

Les trois phases de visionnage d'un programme TV sont clairement distinguables (entre 75% et 78%). Afin de mieux comprendre ce résultat, nous avons représenté la matrice de confusion des différentes activités (**tableau 7**).

Tableau 7 : Matrice de confusion de classification des trois phases de visionnage d'un programme TV.

	Avant	Pendant	Après
Avant	1	0.33	0.14
Pendant	0.26	1	0.28
Après	0.07	0.35	1

La phase « avant » est parfois confuse avec les phases « pendant » et « après » avec 33% et 14% respectivement. De même, la phase « pendant » est confuse avec les phases «avant » et « après » avec 26% et 28% de précision respectivement. Enfin, la phase « après » est confondue avec les phases « avant » et « pendant » à 7% et 35% respectivement.

Il y a donc des confusions plus marquées par rapport à la phase « pendant ». Ceci peut s'expliquer par le fait que cette phase est intermédiaire et que naturellement les comportements, à la limite des phases ou surfaces de décision, peuvent être similaires. Afin de comprendre ces résultats, nous avons calculé la matrice de confusion des activités physiques basées sur le contexte selon les trois phases de visionnage (**le tableau 8**).

Tableau 8 : Matrice de confusion pour la reconnaissance des activités physiques basées sur le contexte.

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	#17	#18
#1	1	0.01	0.03	0.07		0.07	0.03	0.03	0.01	0.02	0.01	0.02	0.03		0.01	0.01		0.01
#2	0.01	1	0.01	0.02		0.03	0.01	0.02		0.01	0.01		0.02		0.01	0.01	0.01	0.01
#3	0.01	0.03	1	0.03	0.02	0.06	0.03	0.01	0.01	0.02		0.02	0.01	0.01		0.033		
#4	0.06		0.01	1		0.01	0.03	0.02	0.01	0.02	0.03	0.01		0.01	0.01	0.01		0.02
#5	0.01			0.03	1	0.01	0.03	0.04		0.01	0.04	0.01	0.03	0.04	0.04	0.02	0.04	0.01
#6		0.01		0.06	0.03	1	0.02		0.02		0.01		0.01	0.01		0.01	0.02	0.01
#7	0.02		0.06				1			0.03		0.03						
#8		0.01		0.05				1			0.01		0.02		0.01	0.03		0.03
#9						0.01			1			0.01			0.03		0.01	
#10		0.03			0.01			0.01		1								0.01
#11				0.01		0.01			0.01		1		0.03	0.02	0.01		0.03	
#12		0.01	0.01									1						
#13							0.06			0.03			1		0.01			0.01
#14		0.03		0.01					0.03		0.01			1			0.01	
#15						0.01							0.03		1			
#16					0.03				0.01							1		
#17	0.01		0.01					0.01				0.01		0.04			1	
#18					0.03					0.01	0.03	0.01			0.04	0.01	0.04	1

Il y a effectivement confusion entre plusieurs activités, en particulier, les activités pouvant arriver simultanément dans deux phases comme : #3 et #4 : manger en étant assis et zapper en étant assis. La phase 2 comprend ainsi, une phase critique où les activités basées sur le contexte se chevauchent.

5.2.2. Résultats de classification des activités de deuxième niveau

Nous présentions ci-dessous les résultats de classifications des activités de deuxième niveau. Les séctions 5.2.2.1, 5.2.2.2 et 5.2.2.3 présentent les résultats de classification de ces activités dans les phases avant, pendant et après, respectivement.

5.2.2.1. Phase « avant »

La figure 9 présente l'ensemble des résultats de classification des activités physiques de deuxième niveau pendant la phase « avant ». Nous remarquons qu'il y a effectivement certaines activités bien classées, comme par exemple #1, #2, #3, #8 et #17. D'autres activités sont moins bien classifiées, comme par exemple, les activités #4, #5, #6 et #7. La précision moyenne de classification est de 77.3%.

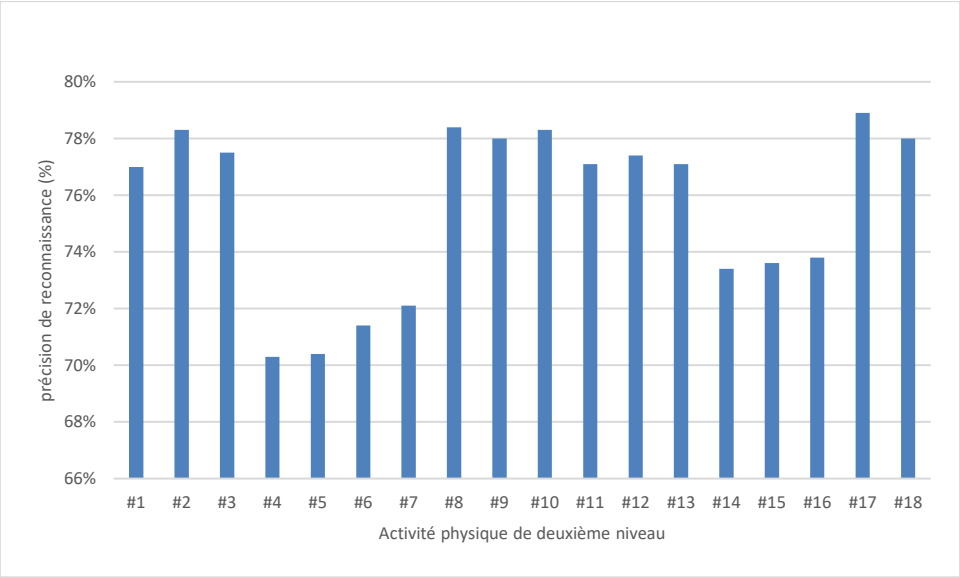


Figure 9 : Précision de reconnaissance des activités physiques pendant la phase « avant »

5.2.2.2. Phase « pendant »

La figure 10 présente les résultats de classification des activités physiques de deuxième niveau dans la phase « pendant ». Ici encore, certaines activités restent moins bien classifiées (par ex. #3, #5, #6 et #7). Le reste des activités sont bien classifiées. La précision de classification moyenne dans cette phase atteint 74.89%.

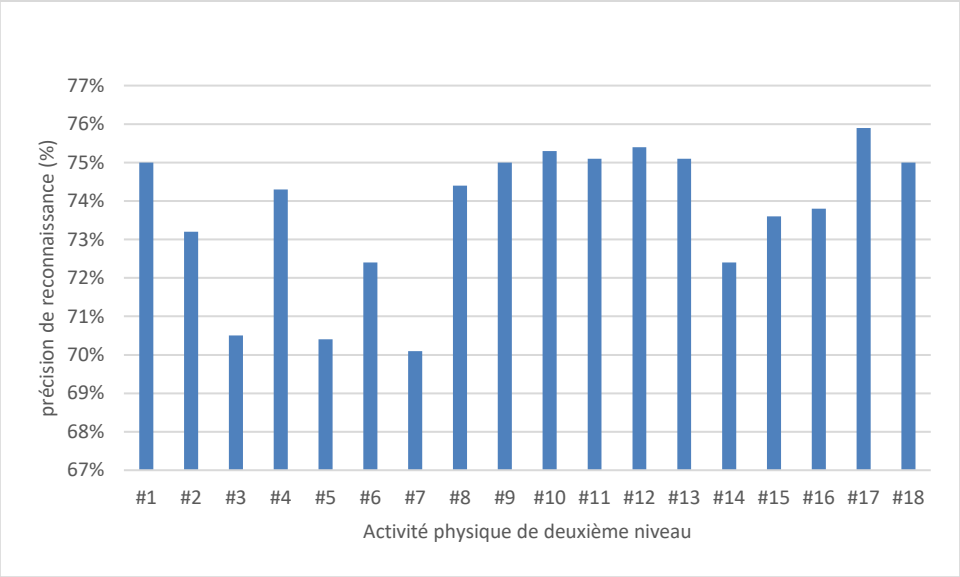


Figure 10 : Précision de reconnaissance des activités physiques pendant la phase « pendant»

5.2.2.3. Phase « après »

La figure 11, présente les résultats de classification des activités physiques de deuxième niveau, pendant la phase « après ». Dans cette phase, les activités physiques sont relativement bien classifiées. La précision moyenne de classification est de 78.3%.

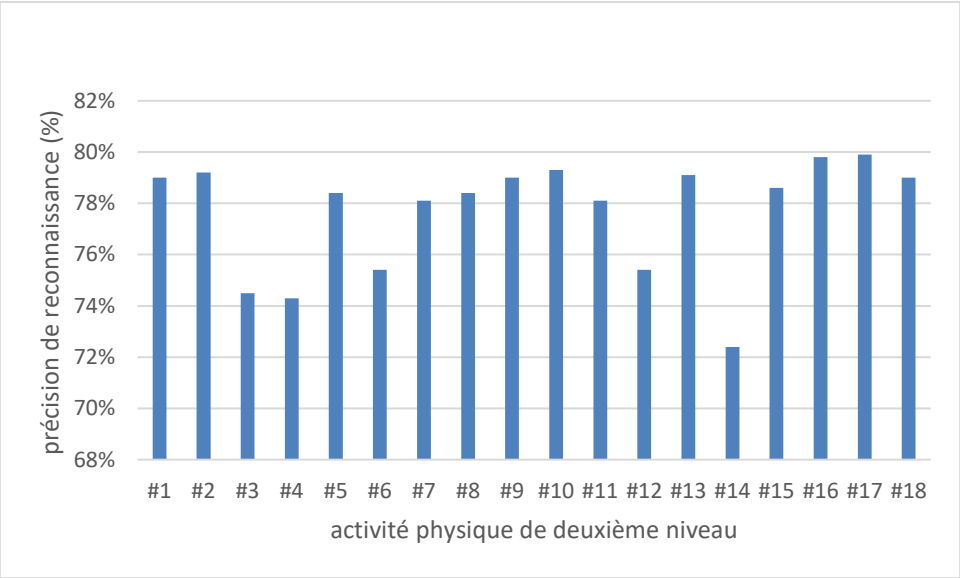


Figure 11 : Précision de reconnaissance des activités physiques pendant la phase « après»

Les précisions de reconnaissance des activités physiques de deuxième niveau sont moins bonnes que celles du premier niveau. Ceci peut etre expliqué par le fait que les activités physiques de deuxième niveau aient « une contrainte » de classification de plus par rapport aux activités du premier niveau. En effet, les activités du deuxième niveau ont été annotées suivant trois phases de visionnage d'un programme TV. Ceci peut limiter grandement l'ensemble d'apprentissage et influe négativement sur les résultats de classification. Prenons l'exemple de l'activité #9 : manger en marchant. Celle-ci a été détectée avec une précision de 84% dans le cas de la reconnaissance des activités physiques du premier niveau (Figue 7). Dans les phases «avant », « pendant » et « après », elle a été détéctée avec 78%, 75% et 79% de précision de classification, respectivement. Afin d'expliquer cette différence, nous prenons un exemple illustratif. Supposons que nous avons 10 activités ayant l'étiquette #9. Cette étiquette correspond à l'activité #9 du premier niveau. Lors du passage aux activités du deuxième niveau, celles-ci auront des étiquettes différentes. Supposons que sur les 10 étiquettes, trois d'entre elles seront attribuées à la phase « avant », trois seront attruées à la phase « pendant » et enfin les quatre restantes seront attribuées à la phase « après ». La reconnaissance de cette activité dans les trois phases implique

l'apprentissage du modèle SVM dans les trois phases. Sachant que le nombre d'activités de type #9 a diminué dans chaque phase. L'ensemble d'apprentissage s'est donc rétréci. Ce qui rend l'apparition de cette activité dans chaque phase de plus en plus rare. La généralisation de l'algorithme d'apprentissage sera donc de moindre qualité. D'où la diminution des précisions de classification.

6. Activités physiques de troisième niveau : détection de l'attention envers la TV

La détection des activités physiques de deuxième niveau consistait à reconnaitre les activités physiques basées sur le conetxte du pemier niveau en ajoutant une information supplémentaire sur les phases de visionnage d'un programme TV. Les activités physiques basées sur le contexte du deuxième niveau ne peuvent pas nous renseigner sur l'attention de l'utilisateur envers la TV. En effet, cette nouvelle information consitute un niveau supplémentaire de reconnaissance des activités physiques basées sur le contexte.

Afin de reconnaitre l'attention des participants envers la TV, nous nous intéressons tout particulièrement à la phase « pendant le visionnage d'un programme TV ». En effet, cette phase caractérise le visionnage « effectif ou réel » des participants, alors que les autres phases de visionnage sont plutôt des phases préliminaires, pour la phase avant, et finale, pour la phase après. Nous pouvons ainsi analyser l'attention à la TV pendant cette phase car elle peut etre une « bonne candidate » à contenir des informations pouvant détecter l'attention à la TV. **La figure 12** illustre ce propos.

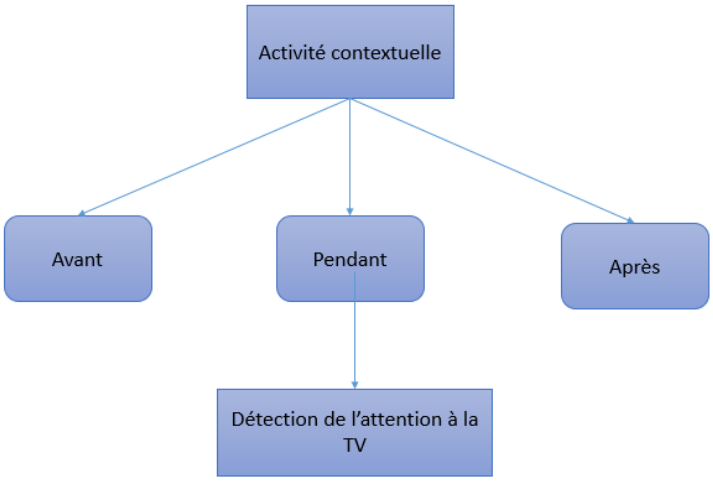


Figure 12 : Etude de l'attention envers un programme TV en utilisant les trois objets connectés.

Cette étude s'inspire des travaux du suivi à domicile des personnes agées [357-358] dans un environement intelligent. Les auteurs ont conçu un système composé de plusieurs capteurs

attachés dans une pièce. L'objectif était de détecter des états de détrèsses chez une personne agées, en partcuiler la chute. Ils ont développé un système de prédiction stipulant que si une personne se trouvant dans la cuisine en étant allongée à 3 heures du matin alors il s'agit d'une chute. Un système permettant de distinguer si la personne allongée est entrain de dormir ou pas a été proposé en se basant sur les informations renvoyées par chaque capteur. Des microphones ont aussi servi à reconnaitre des caracteristiques d'une chute en reconnaissant les voix de la personne agée appelant au secours ou en criant simplement suite à une chute.

De manière analogue, nous avons étudié l'attention de l'utilisateur envers la TV en se basant sur l'utilisation de trois systèmes : (1) distinction entre sommeil et allongé. (2) distinction entre les différentes activités physiques et (3) distinction entre les différentes voix des locuteurs de la maison afin de savoir s'il parle avec quelqu'un en regardant la TV ou pas.

Dans la littérature, nous avons relevé le problème de la détection des activités physiques basées sur le contexte avec des objets connectés intrusifs. Un travail de recherche ayant effectué l'étude de l'attention envers la TV a été illustrée dans la partie 1. Ce dernier a traité le problème de détection avec un traqueur des yeux. Il a ensuite utilisé une technique basée sur le vote majoritaire pour déterminer la signature de l'attention envers la TV [131].

Mise à part l'aspect intrusif de l'expérience menée, la reconnaissance de ces activités n'a pas pris en considération l'activité physique inhérente à cet état d'attention envers la TV. En effet, les activités ou états physiques des participants ne sont pas connues. Nous ne savons pas ce que les participants faisaient pendant qu'ils sont face à l'écran.

Nous proposons, dans cette partie, une nouvelle approche d'analyse et détection de l'attention envers la TV en utilisant les trois objets connectés. L'approche proposée a l'avantage d'utiliser des objets connectés du quotidien et moins intrusifs que ceux déjà explorés dans l'état de l'art. Ainsi, notre approche est décrite dans les points suivants :

- Ségmentation (profiling ou profilage) de l'ensemble des participants afin de dégager des groupes d'individus ayant des comportements similaires ; suite à cela, trois différentes bases de données de comportements sont générées.
- Les groupes ainsi détectés dans l'étape précédente vont permettre de construire une architecture de détection de l'attention envers la TV ; chaque composante fournira une signature partielle d'attention envers la TV.
- Les signatures sont concaténées et classifiées pour détecter l'attention envers la TV.

La figure 13 représente une vue d'ensemble des traitements effectuées afin de reconnaitre l'attention à la TV.

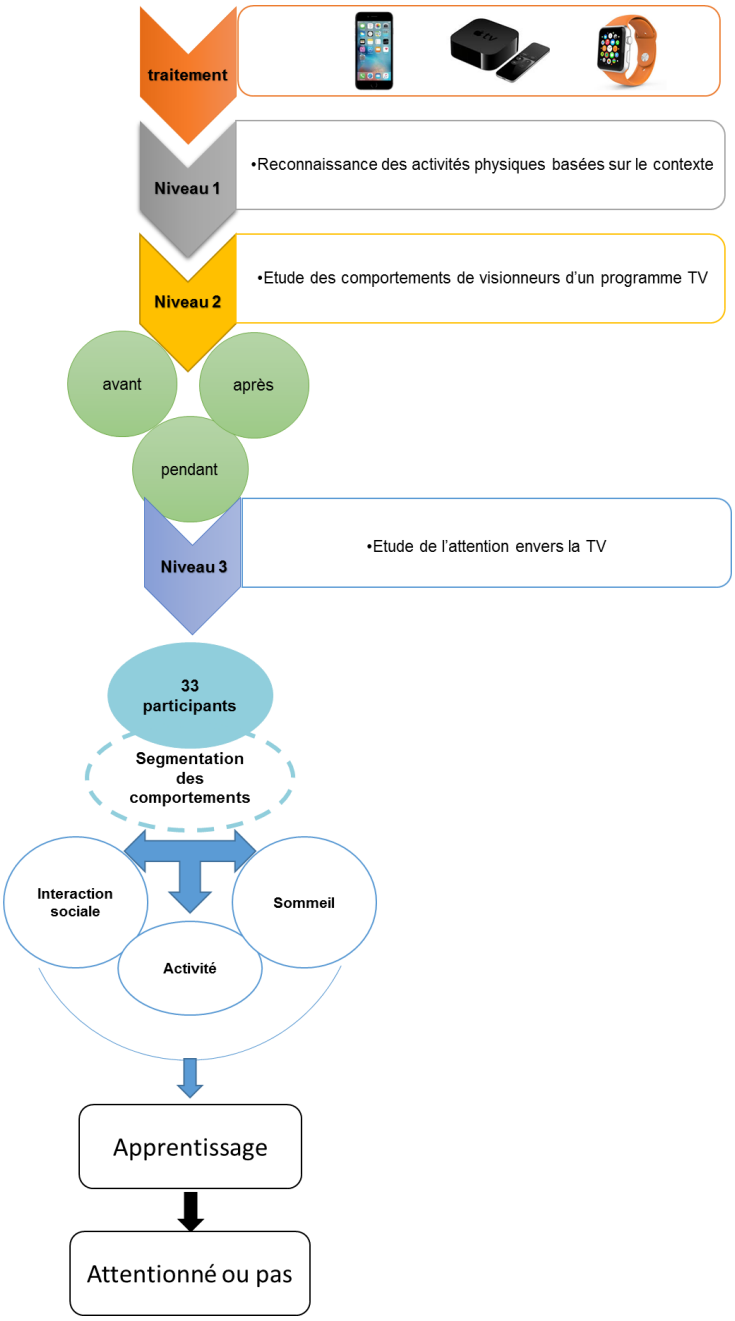


Figure 13 : Vue d'ensemble du procéssus de reconnaissance des activités physiques de Troisième niveau.

Nous proposons de détailler les traitements effectués dans les sections suivantes.

6.1. Segmentation des individus

Le but de la ségementation des comportements des participants est de dégager des groupes de comportements permettant de décrire le mieux l'attention en vers la TV. La ségmentation des comportements des individus suit un algorithme de clustering hiérarchique. Nous commençons par une ségmentation non hiérarchique afin d'initialiser les groupes de comportements. Puis une étape de ségmentation hiérarchique afin d'utiliser les groupes

définis dans l'étape précédente. Enfin une ségmentation non hiérarchique sur les s groupes déterminés dans l'étape précédente sera faite. Le détail de l'algorithme est donné comme suit :

(1) L'objectif est d'obtenir rapidement une partition de m (m grand) en k classes, avec k supérieur au nombre de classes finales *s* souhaité, grâce à une méthode de segmentation non hiérarchique. On prend par exemple k=100, tel que k $\ll m$. Cette segmentation n'est pas utilisable directement : les groupes sont nombreux et proches les uns des autres. Toutefois, ils ont l'avantage d'être très homogènes et de contenir des individus qui n'ont pas à être séparés. On peut répéter cette étape pour rechercher des formes fortes.

(2) On peut désormais utiliser un algorithme de segmentation hiérarchique pour regrouper les k groupes de l'étape (1) (chaque groupe a pour poids la somme des poids qu'il représente). L'étape d'agrégation va permettre de reconstituer des groupes (clusters) qui ont été inutilement fragmentés lors de l'étape précédente, en sélectionnant un nombre de classes optimal (s). Il convient d'utiliser le critère de Ward pour la constitution de cet arbre, afin de tenir compte des masses des classes.

(3) Enfin, un clustering non hiérarchique final est réalisé, pour s groupes définis lors de l'étape (2). Comme pour l'étape (1), on peut répéter cette étape pour rechercher des formes fortes.

Nous avons, avant tout, commencé par omettre l'étape (1) car nous avions supposé ne pas connaitre le nombre de classes. Puis, afin de verifier la pertinence du résultat, nous avons executé toutes les étapes de 1) à 3).

6.2. Le descripteur de l'attention envers la TV

Le profilage appliqué dans la partie précédente a permis de dégager trois grands groupes de données:

- Un groupe caractérisant l'activité physique élémentaire et basée sur le contexte.
- Un groupe caractérisant des individus en position allongée/dormir/en état de sommeil : les données de ce groupe sont soit des données représentant l'activité allongée, soit les données représentant une personne en état de sommeil.
- Un groupe caractérisant les interactions sociales permettant l'identification des personnes présentes avec la personne effectuant l'expérience.

Grace à cette architecture de données, le modèle de reconnaissance de l'attention envers la TV va etre entrainé de la manière suivante :

Le système « sommeil » sur la figure 14 va distinguer si le participant dort ou est juste allongé. Le système « interaction sociale » va distinguer si la personne est seule ou elle parle avec quelqu'un. Le système « activité physique » sert de validation. C'est-à-dire qu'il permet juste de détecter les activités physiques et de comparer avec les activités détectées dans le système

« sommeil ». L'information renvoyée par les trois systèmes va alimenter l'algorithme de reconnaissance de l'attention. Par exemple, si la personne est allongée et ne dort pas et ne parle avec personne et que le participant augmente le volume de la TV à l'aide de la touche « + », il y a de fortes chances qu'il soit attentionné envers la TV.

En se basant sur cette structure de trois catégories de groupes, nous avons établi une série de traitements afin de détecter des signatures de l'attention envers la TV.

Le schéma de la **figure 14** indique les trois principales parties : (1) reconnaissance de l'activité physique, (2) reconnaissance du locuteur et enfin (3) reconnaissance du sommeil.

Les données étiquetées ont été donc ségmentées afin d'avoir des groupements de données homogènes.

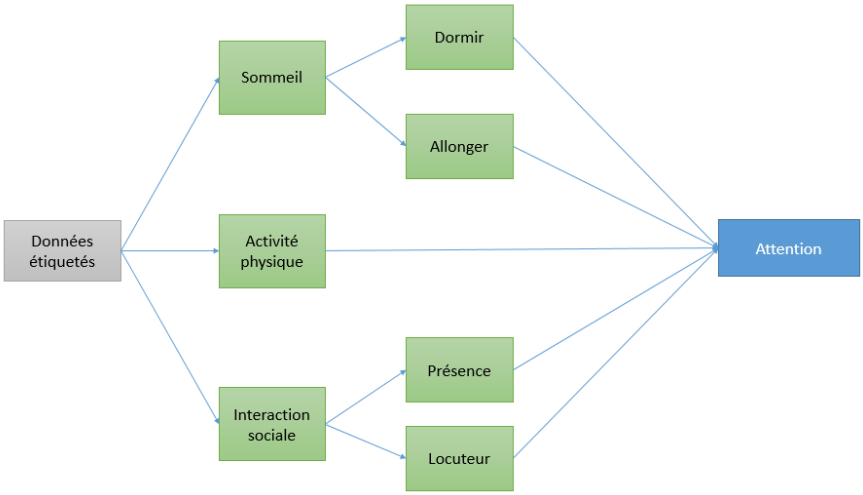


Figure 14 : Système de détection de l'attention envers la TV.

Nous proposons dès à présent d'étudier les trois systèmes ainsi définis.

6.2.1. Système de reconnaissance du locuteur

Les trois objets connectés embarquent des microphones. Nous avons donc décidé de bénéficier de la fonctionnalité offerte par ces capteurs pour détecter les différents locuteurs de la maison, entre autre celle du porteur des objets connectés.

Les systèmes d'identification automatique de locuteurs offrent des performances acceptables lorsqu'ils sont utilisés avec une parole propre (nettoyée) [132]. Cependant, ils deviennent pratiquement instables lorsqu'ils fonctionnent dans des environnements bruyants comme c'est le cas des objets connectés. Afin de calculer les descripteurs pour la reconnaissance de l'attention envers la TV, nous proposons d'extraire un ensemble de descripteurs déjà bien utilisés dans l'état de l'art comme une combinaison de la méthode de prédiction linéaire perceptive de la transformation spectrale relative (RASTA-PLP) et des coefficients

conventionnels de fréquence fréquentielle de Mel (MFCC), à partir des signaux des microphones des trois objets connectés, (le détail de ce calcul est décrit en Annexe C).

L'extraction de descripteurs est la première étape de la tâche de reconnaissance automatique du locuteur. Elle vise à extraire des caractéristiques de la forme d'onde de la parole pour représenter la signature de la voix du locuteur. Puisque la parole est un signal non stationnaire, les paramètres des descripteurs doivent être estimés sur des intervalles de 10 ms à 20 ms [166, 208], dans lesquels la parole est considérée comme stationnaire. Les principaux types de techniques de traitement utilisés sont: le codage prédictif linéaire (LPC) [133], la prédiction linéaire perceptuelle (PLP) [134], Mel Frequency Cepstral Coefficients (MFCC) [135], les coefficients de céphalométrie fréquentielle de Gammatone (Gammatone Frequency Cepstral Coefficients) (GFCC) [136] et RASTA-PLP. La **figure 15** représente une vue d'ensemble du processus de traitement et d'identification du locuteur. Nous nous sommes donc inspirés des travaux référencés dans [134-135]. Nous avons combiné les trois descripteurs PLP, RASTA PLP et MFCC car cette combinaison avait donné de bonnes perfromances de reconnaissance.

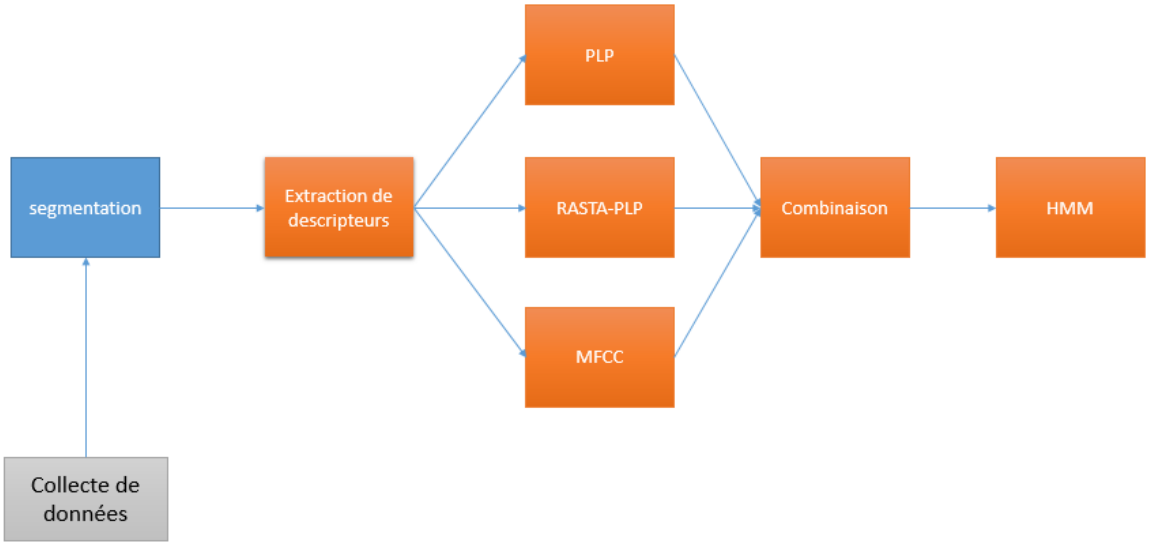


Figure 15 : Système d'extraction de descripteurs pour la reconnaissance du locuteur.

Dans cette partie, nous nous sommes intéressés tout particulièrement à trois types de descripteurs déjà utilisés dans l'état de l'art : MFCC et RASTA-PLP, deux des plus populaires descripteurs très largement utilisés dans l'état de l'art. Nous avons choisi d'étudier le RASTA-PLP combiné avec la méthode conventionnelle MFCC pour caractériser les voix du locuteur. Le descripteur ainsi extrait est donné en entré du HMM pour la classification. Les résultats et détails de cette étude sont donnés en Annexe C.

6.2.2. Système de reconnaissance de l'état du sommeil

Notre descripteur final de l'attention à TV devrait inclure non seulement la signature de l'activité physique du locuteur mais aussi de l'état du sommeil (allongé simplement ou allongé par ce que la personne dort).

Pour cela, nous nous sommes concentrés sur le capteur du cardiofréquencemètre embarqué dans la smartWatch.

En effet, en plus des signatures déjà calculées pour les deux précédents systèmes, la smartWatch peut nous renseigner si la personne dort ou pas. Ainsi, couplée avec les informations reçues du système de reconnaissance de l'activité physique, ce système pour nous confirmer si la personne est allongée ou pas. Pour cela, nous avons calculé une signature bien connue de l'état de l'art à savoir la variabilité de la fréquence cardiaque (VFC) [137] [138].

Le signal de la variabilité de la fréquence cardiaque (VFC) sera dérivé des signaux du cardiofréquencemètre de l'AppleWatch. Ce signal sera prétraité pour une analyse ultérieure, comme présenté dans un travail similaire [139].

L'analyse de la variabilité de la fréquence cardiaque est un outil très puissant et non invasif déjà largement utilisé dans l'état de l'art pour diverses applications, principalement pour l'analyse du sommeil [140]. Notre méthode comprend les étapes suivantes: (1) l'acquisition des données, (2) l'étape de prétraitement, (3) l'étape d'extraction des caractéristiques et (4) l'étape de classification.

Le prétraitement des données VFC a été une étape cruciale pour extraire avec précision les caractéristiques. Les mesures du VFC dans le domaine temporel et dans le domaine fréquentiel ont été calculées à partir de l'intervalle RR [141]. Différentes architectures de réseau de neurones ont été testées.

Enfin, nous avons sélectionné le meilleur réseau de neurones basé sur des matrices de confusion. Les résultats vont permettre de classer les états de sommeil / éveil en fonction de la VFC. La **figure 16** donne une vision globale de la démarche suivie.

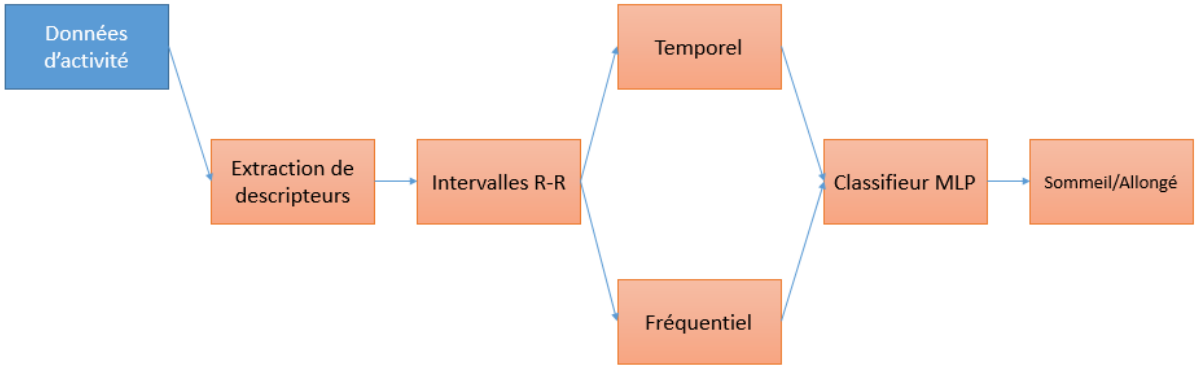


Figure 16 : Système de reconnaissance de l'état du sommeil exploitant la variabilité de la fréquence cardiaque.

Le signal sera prétraité de la manière suivante : l'enregistrement du cardiofréquencemètre qui consiste à identifier d'éventuels artefacts, exclure les intervalles RR non normales. Après prétraitement des intervalles RR, une analyse plus poussée est possible sur les données NN (normales à normales) obtenues.

Les signatures des trois systèmes sont concaténées et données en entrées de plusieurs algorithmes d'apprentissage automatiques. **Le tableau 9** résume les résultats de classification.

Tableau 9 : Résultats de classification de l'attention envers la TV en utilisant quatre algorithmes d'apprentissage : Machine à vecteurs de support (SVM), Un réseau de neurones multicouches (MLP), un classifieur naif de Bayes (NB) et un algorithme des forets aléatoires (RF).

	SVM	*MLP*	*NB*	*RF*
Système 1	*85.5%*	*70.1%*	*55%*	*70.8%*
Système 2	*73.1%*	*66.6%*	*57.5%*	*66.05%*
Système 3	*75.4%*	*69.47%*	*60.04%*	*70.1%*
Concaténation	*70.21%*	*64.41%*	*61.01%*	*69.02%*

La prédiction de l'attention envers la TV est relativement bien détectée. L'algorithme du SVM donne quant à lui de meilleurs résultats de classification.

La contribution de la détection du sommeil ainsi que le locuteur ont permis d'améliorer la précision de classification.

Notre approche a certe permis d'avoir une bonne précision de reconnaissance de l'attention envers la TV, les résultats de classification sont comparables à ceux de l'état de l'art,

notamment le travail de Chamaret et al [129] ayant obtenu une précision de 90% de reconnaissance dans un environnement contrôlé en utilisant un traqueur des yeux.

7. Synthèse des chapitres 2 et 3:

Dans les chapitres 2 et 3 nous avons réalisé des études afin de reconnaitre des activités physiques élémentaires et des activités physiques basées sur le contexte. Ces études ont permis de proposer un nouveau corpus de données des activités physiques réalisées dans un environement non controlé et en utilisant trois objets connectés du quotidien. Nous avons monté que la DCT est le meilleur descripteur pour les activités physiques élémentaires.

La prise en compte des chevauchements entre activités a été explorée. Nous avons montré que 40% des données de chevauchement étaient nécéssaires afin d'atteindre une meilleure performance de reconnaissance, ce qui n'a pas été suffisamment exploré par les outils conventiennels de reconnaissance de l'activité physique.

La prise en compte des sources de données supplémentaires telles que les données audio, boutons, etc, a permis de faire emmerger la notion d'activité physique basée sur le contexte. Afin de reconnaitre ce type d'activité, une architecture à trois niveaux de granularité à été proposée.

Contrairement à d'autres approches déjà existantes dans l'état de l'art, notre approche de reconnaissance de l'activité physique basée sur le contexte s'inspire des étiquettes déjà attribuées pour la reconnaissance des activités élémentaires afin de reconnaitre les activités basées sur le contexte. Ce nouveau protocole de ségmentation et étiquetage des données des avtivités baséés sur le contexte, permet de retrouver les activités physiques élémentaires dont les activités basées sur le contexte sont issues. Une telle stratégie a permis de reconnaitre des activités telles que l'attention envers la TV en utilisant des objets non intrusifs.

8. Limites des études effectuées dans les chapitres 2 et 3 :

La reconnaissance des activités physiques élémentaires et basées sur le contexte fait intervenir des données multicapteurs. Les descripteurs extraits à partir de ces différentes sources ont été concaténés afin de les donner en entrées des algorithmes d'apprentissage. L'état de l'art a montré que ce type d'approche ne permet pas d'entrainer efficacement les modèles d'apprentissage. En effet, chaque source de données apporte sa propre information qui peut avoir un impact significatif sur l'algorithme d'apprentissage. Le fait de concaténer les sources fait perdre cette caracteristique importante dans le procéssus de reconnaissance de l'activité (lenteur, redondance, etc)

Les données capteurs sont par définition imparfaites. Elles sont entachées d'erreurs, de bruits, d'incertitude, etc. Ainsi, les modèles d'apprentissage automatique ne prennent pas cette contrainte en compte, ce qui limite grandement leur utilisation.

L'utilisation d'algorithmes d'apprentissage automatique a donné de bons résultats de reconnaissance des activités élémentaires et basées sur le contexte. Toutefois, les approches utilisées nécessitent une étape de séléction de descripteur et réduction de la dimensionnalité des descripteurs. Cette étape est très fastidieuse à mettre en place. Une des solutions proposées dans l'état de l'art est l'utilisation d'approches a base de réseaux de neurones profonds. En effet, celles-ci ont la capacité d'apprendre une grande quantité de données sans avoir à séléctionner des descripteurs ni à réduire la dimentionnalité de celles-ci. Toutefois, ce type d'approches souffre de plusieurs problèmes : lenteur du procéssus d'apprentissage, explosion et disparition du gradient, etc.

Afin de répondre à ces limitations, nous proposons d'aborder la théorie de Dempster Shafer pour la fusion de données multicapteurs, afin de prendre en compte l'incertitude des données provenant des différentes sources.

Nous proposons également d'explorer de nouvelles voies de recherche dans l'utilisation des réseaux de neurones profonds pour la reconnaissance de l'activité physique. Ainsi, la suite de ce manuscrit va etre subdivisée en deux chapitres :

- **Le chapitre 4 :** Nous verrons comment remédier au problème de fusion de données multi capteurs et comment prendre en compte l'imperfection de ce type de données. Nous proposerons de nouvelles approches permettant d'impélmenter efficacement ce modèle appliqué à la reconnaissance de l'activité physique.
- **Le chapitre 5 :** Nous implémenterons différentes approches basées sur l'utilisation de modèles d'apprentissage en profondeur (avec et sans sélection de descripteurs). Nous proposerons également une étude exploratoire et comparative permettant d'optimiser l'entrainement des réseaux de neurones profonds. Enfin, nous introduirons une heuristique permettant d'améliorer les résultats ainsi obtenus.

Conclusion

Dans ce chapitre, nous avons proposé une définition des activités physiques basées sur le contexte. De part la complexité du processus de reconnaissance de ce type d'activités, nous avons proposé une architecture montrant trois niveaux de granularité pour la reconnaissance des activités physiques. Une méthodologie permettant d'extraire des descripteurs à partir des

signaux de capteurs embarqués dans trois objets connectés : un smartphone, smart Watch et une smart TV a été également explorée.

Nous avons exploré la puissance des algorithmes de ségmentation afin de proposer une méthode permettant de ségmenter les comportements des différents individus et nous a donc permis de dégager des groupes homogènes de comportements qui ont permis d'extraire des caractéristiques pour la détection de l'attention à la TV.

Les comportements des visionneurs d'un programme TV ont été étudiés. Grace à la ségmentation en trois phases de ces comportements, nous avons pu établir une architecture pour détecter l'attention envers la TV. Ainsi, l'attention à la TV a pu être détectée sans utilisation de capteurs intrusifs, dans un environnement non contrôlé et enfin en utilisant l'information sous-jacente à l'activité basée sur le contexte.

Dans le chapitre suivant, nous proposons de nouvelles approches pour fusionner les données des différents capteurs afin de prendre en considération le contexte et l'incertitude sur les données provenant des différents capteurs.

Chapitre 4

Prise en compte de l'imperfection des données pour la reconnaissance de l'activité physique

1. Introduction……119

2. Etat de l'art……120

3. Vue d'ensemble de la DST pour la reconnaissance de l'activité 126

4. Résultats……134

4.1. Activités physiques élémentaires.....134

4.2. Activités physiques basées sur le contexte.....135

Conclusion…… 156

Chapitre 4

1. Introduction

Les données provenant des différents capteurs sont, par définition, imparfaites. Elles peuvent être source d'erreurs, d'incertitude, etc. La classification des activités physiques par apprentissage automatique ne prend pas en compte cet aspect. Par ailleurs, la non prise en compte de l'imperfection peut induire à des erreurs de reconnaissance voire d'interprétation.

Parmi les différentes approches proposées dans l'état de l'art, nous pouvons trouver les méthodes basées sur les fonctions de croyances, par exemple la théorie de Dempster-Shafer (DST). Celles-ci, à la différence d'autres approches (par ex. logique floue, théorie des probabilités ou des possibilités, ect.), permettent de modéliser les imperfections liées à différents niveaux d'incertitudes et les incorporer dans la prédiction finale de l'activité.

La mise en oeuvre de l'algorithme de Dempster-Shafer s'effectue en trois étapes : (1) l'extraction des descripteurs, (2) le calcul des fonctions de masse et enfin (3) la fusion des fonctions de masses à l'aide de la règle de combinaison de Dempster-Shafer et la prédiction de l'activité.

Touefois, l'implémentation de cette théorie pose deux principales problématiques : i) l'absence d'une approche générique de calcul des fonctions de masse et ii) la fusion des fonctions de masse nécessite un coût de calcul élevé et un temps de calcul long. Ce problème limite grandement son application pour la reconnaissance de l'activité physique.

Deux approches permettant de calculer les fonctions de masse ont été explorées ; la première consiste à subdiviser dynamiquement l'espace des caractéristiques en plusieurs cellules. La deuxième approche consiste à utiliser des classifieurs. Les fonctions de masse ainsi calculées seront fusionnées en utilisant la règle de combinaison de Dempster-Shafer.

Avant de fusionner les fonctions de masses, celles-ci doivent etre approximées afin d'optimiser les calculs induits par la fusion. Une nouvelle méthode basée sur le coefficient de corrélation sera utilisée, ce qui peut réduire le coût de calcul de la combinaison de preuves efficacement. Enfin, l'approximation des fonctions de masse génère parfois une perte d'information. Nous proposons à la fin de ce chapitre une étude comparative entre l'approche proposée et celles étudiées dans l'état de l'art, en mettant l'accent sur la perte d'information générée par l'approximation des fonctions de masse.

Dans la partie suivante, nous proposons d'aborder quelques travaux sur la reconnaissance des activités physiques avec la théorie de Dempster-Shafer.

2. Etat de l'art

La théorie de Dempster-Shafer est de plus en plus appliquée pour représenter et manipuler l'information contextuelle sous incertitude dans un large éventail de systèmes de reconnaissance de l'activité.

Dans le domaine de la reconnaissance de l'activité physique, la DST a été utilisée principalement afin de représenter les informations contextuelles inhérentes aux activités physiques [142]. Par exemple, Milaghardan et al., 2018 [143] se sont intéressés à détecter des trajectoires des participants et leurs associer des degrés d'incertitude en utilisant différents types de capteurs (accéléromètres, gyroscopes, caméras, etc), fixés à l'environnement.

D'autres travaux ont utilisé la DST dans le but d'étudier l'impact des erreurs induites par des capteurs potentiellement défaillants sur la précision de reconnaissance de certaines activités physiques élémentaires. De plus, ils ont proposé une étude de cas montrant que la DST peut incorporer l'incertitude dérivée des erreurs du capteur déduire ainsi l'activité. Une précision de 69.4% a été atteinte [144].

Enfin, nous avons recueilli les principaux travaux sur la reconnaissance de l'activité physique ayant utilisé la DST. Le **tableau 1** suivant résume les principales caractéristiques extraites de ces études, notamment : le nombre d'activités détectées, la précision de classification globale, la méthode de calcul de la fonction de masse et enfin le temps et la complexité de calcul.

Tableau 1 : Caractéristiques des études réalisées autour de la méthode de Dempster-Shafer pour la reconnaissance de l'activité physique.

	Nombre d'activités détectées	Précision globale	Fonction de masse	Temps et complexité de calcul	Environnement
[145]	4	88.2%	Vote majoritaire	Polynomiale	Contrôlé
[146]	3	79%	Vote majoritaire	Pas d'indication	Contrôlé
[147]	7	70%	Réseau de neurones	Pas d'indication	Contrôlé
[148]	5	70%-80%	Vote majoritaire	Polynomiale	Contrôlé
[149]	3	71%-80%	Méthode du rejet	Pas d'indication	contrôlé

Jing et al [145] [144] ont utilisé des caméras attachées à pluieurs endroits d'une maison. Les participants étaient équipés d'accéléromètres fixés aux niveaux des genoux.Ils ont proposé d'utiliser la DST pour détecter les activités suivantes : ouvrir le robinet, ouvrir la porte, allumer

la lumière et tirer la chasse, dans un environement controlé. Ils ont proposé une structure de combinaison de la fonction de masse à trois couches : une couche représentant les différents capteurs utilisés, une deuxième couche considère les caracteristiques à extraire des signaux et enfin une dernière couche concerne les activités inférées. Cette structure peut être utilisée pour combiner les fonctions de masse dérivées des capteurs. Suite à l'évaluation de la méthodologie proposée, ils ont démontré que la théorie des preuves de Dempster-Shafer peut incorporer l'incertitude dérivée des erreurs du capteur et inférer ensuite l'activité. Une précision de 88,2% à été atteinte.

Sebbak et al [146] ont proposé une autre façon d'appliquer la DST en utilisant des informations de capteurs discrets binaires pour la reconnaissance d'activité sous incertitude. Leur technique est basée sur une méthode permettant de convertir et d'agréger les données brutes capturées, à l'aide d'un réseau de capteurs sans fil (caméras, accéléromètres, gyroscopes et lumière ambiante). Ils ont proposé une technique de résolution de conflit pour optimiser la prise de décision. Ils ont utilisé une technique de séléction et normalisation des éléments focaux dans le cadre de discernement, ce qui leur a permis d'optimiser les calculs. Pour la validation de l'approche, ils ont utilisé un ensemble de données collectées à l'aide de capteurs déployés dans une maison intelligente, dans des conditions contrôlées. Une précision de 79% a été atteinte pour détecter trois activités : préparer le diner, préparer le petit déjeuner et utiliser les toilettes.

McKeever et al [147] ont proposé d'inclure d'avantage l'information temporelle pour la reconnaissance des activités physiques à l'aide de la DST. Ils ont proposé un cadre de raisonnement probant qui incorpore des connaissances temporelles (à quelle heure s'est réalisée telle activité ?, quelle moment se réalisera ?, etc). Ils ont évalué l'efficacité de leur approche à l'aide d'un ensemble de données collectées dans une maison intelligente. Une précision de la reconnaissance des activités : quitter la maison, toilette, douche, dormir, petit déjeuner, diner, boire : était de 70%.

Perez et al [148] ont proposé une approche de combinaison de classificateurs basée sur l'utilisation de la DST et la combinaison de plusieurs classifieurs en cascade avec rejet de certains résultats des classifieurs ainsi combinés. Cette architecture fonctionne comme une méthode de vote majoritaire, elle prend en compte les résultats renvoyés par plusieurs classifieurs et retient le nombre maximum de précision. Les résultats ont montré l'intérêt de l'approche basée sur la fonction de croyance, qui permet d'obtenir une plus grande précision par rapport aux autres stratégies de combinaison (Mixed-based classifers, en anglais). Les résultats expérimentaux révèlent également que la crédibilité d'une source est définie en

fonction de sa performance sur l'ensemble d'apprentissage. L'approche proposée est robuste à des sources peu fiables et à un petit nombre de sources que d'autres approches combinées.

Bien que la DST soit largement utilisée dans de nombreux domaines, il y a tout de même des difficultés quant à son déploiement dans des systèmes tels que les objets connectés ou portables. En effet, l'un de ses principaux problèmes est que lorsque le cadre de discernement contient un nombre important d'éléments focaux, le calcul et la fusion des fonctions de masses requiert un cout et temps de calcul assez important. Pour surmonter ces inconvénients, de nombreuses méthodes ont été proposées, elles sont subdivisées en trois catégories :

La première catégorie modifie la règle de combinaison de Dempster-Shafer. Dans cette catégorie, les méthodes réduisent la combinaison de calcul des preuves. Plus précisément, ces méthodes réalisent le calcul rapide en approximant la règle de combinaison pour réduire les calculs inutiles. Par exemple, les travaux dans [150] [151] introduisent un algorithme optimal pour la règle de combinaison de Dempster-Shafer.

La deuxième catégorie [152] réduit le coût de calcul en utilisant des techniques de Monte-Carlo [153] qui donne une approximation et une valeur approchée des fonctions de masse, ces techniques ont l'avantage d'etre précises mais très couteuses en temps et complexité. Toutefois, à notre connaissance, elles n'ont jamais été appliquées à la reconnaissance de l'activité physique.

La troisième catégorie consiste à supprimer certains éléments focaux pour obtenir un cadre de discernement plus simple. Notre approche entre dans le cadre de cette catégorie. En effet, la principale cause du grand cout de calcul est que le nombre d'éléments focaux augmente exponentiellement avec la croissance du cadre de discernement, l'approximation de la fonction de masse s'est avérée efficace pour réduire le temps de calcul [154]. Dans cette catégorie, le processus d'approximation est divisé en deux étapes :

Tout d'abord, certains éléments focaux sont supprimés selon des critères spécifiques de différentes méthodes. De plus, l'attribution de masse des éléments supprimés est affectée aux éléments focaux restants selon différentes stratégies d'attribution. Par exemple, Tessem [155] a proposé la méthode d'approximation *k-l-x* [156]. Cette méthode consiste à supprimer les éléments focaux de plus petite valeur de masse et à normaliser tous les éléments restants. En outre, Lowrance [157] a proposé une méthode de récapitulation (Summarization method, en anglais) [158], qui supprime également les éléments focaux de plus petite valeur de masse comme dans la méthode *k-l-x*. Les affectations de masse des éléments focaux supprimées sont accumulées et affectées à leur ensemble d'union. De plus, Yang [159] a proposé une approximation de fusion au niveau du rang [160]. Dans cette méthode, les éléments focaux sont supprimés en fonction de la masse et de la cardinalité du cadre de discernement. Ensuite,

tous les éléments restants sont normalisés. Ainsi, cette catégorie inclut également la méthode D1 [156], l'approximation de non-redondance [161], l'approximation itérative basée sur la distance de preuve (distance of evidence, en anglais) [162] [163], etc. utilisant le coefficient de corrélation [164], qui appartient à la troisième catégorie.

Dans ce contexte, beaucoup de travaux ont été proposés dans la littérature nous les résumons dans les quatre points suivants:

a) *k-1-x* approximation [155]. Il s'agit d'une méthode classique pour supprimer les éléments focaux. La fonction de masse est approximée en supprimant un certain nombre d'éléments focaux de la fonction, laissant ceux avec les plus grandes masses. Ensuite, la fonction de masse est normalisée. Pour la fonction approximée, au plus k éléments focaux et au moins *l* éléments focaux peuvent être laissés. La masse des éléments focaux restants est supérieure à *1-x.*

b) Approximation récapitulative (Summarization approximation, en anglais) [157]. Cette méthode est similaire à *k-l-x*. La seule différence est que la masse d'éléments focaux supprimés est affectée à leur ensemble d'union. Par rapport à *k-l-x,* les informations sur les éléments focaux supprimés ne sont pas complètement abandonnées.

c) Approximation basée sur la fusion au niveau du rang (Rank-level fusion, en anglais) [159]. Cette méthode peut à la fois tenir compte de la cardinalité des éléments focaux et des valeurs d'assignation de masse. Elle fonctionne de la manière suivante : tout d'abord, tous les éléments focaux d'une fonction de masse originale (avec L éléments focaux) sont triés en fonction des valeurs d'assignation de masse (dans l'ordre croissant). Le vecteur de rang obtenu est le suivant :

$$r_m = [r_m(1), r_m(2), \dots, r_m(L)]$$

où $r_m(i)$ indique la position de rang des ième éléments focaux (i = 1, ..., L) dans la fonction d'origine en fonction de l'assignation de masse. Ainsi, tous les éléments focaux d'une fonction de masse originale sont triés en fonction des cardinalités (dans l'ordre décroissant). Le vecteur de rang peut être obtenu par l'équation suivante:

$$r_c = [r_c(1), r_c(2), \dots, r_c(L)]$$

Où $r_c(i)$ indique la position de rang des ièmes éléments focaux dans la fonction d'origine selon la cardinalité. Puis, en fonction de la fusion des rangs, nous pouvons obtenir un rang fusionné par l'intermédiaire de l'équation ci-dessous :

$$\boldsymbol{r_f = [r_f(1), r_f(2), \dots, r_f(L)]}$$

Tel que $r_f(i) = \alpha . r_f(i) + (1 - \alpha) . r_c(i)$ et α ∈ [0,1] est utilisé pour montrer la préférence de deux critères différents. Un tel rang fusionné peut être considéré comme un critère plus

complet contenant à la fois l'information des assignations de masse et la cardinalité. Enfin, trouver l'élément focal avec la plus basse valeur de $r_f(i)$ et le retirer de l'assignation (ou la masse) d'origine. Puis, renormaliser les éléments focaux restants. Enfin, répéter les étapes ci-dessus jusqu'à ce que k éléments focaux soient retirés.

d) Distance d'approximation basée sur des preuves (*Distance of evidence based approximation*) [162]. Cette méthode propose une approche d'approximation itérative basée sur la minimisation d'une distance. Elle fonctionne de la manière suivante : tout d'abord, un élément focal A_i est enlevé de la fonction d'assignation de masse originale m (avec L éléments focaux). Ensuite, une normalisation de la masse de l'élément A_j restant est effectuée pour générer une nouvelle fonction m_i' . Puis, la distance entre m_i' et m désignée par d (i) est calculée. Ainsi, l'élément focal avec la plus petite masse est retiré. Les masses restantes vont etre normélisées afin d'obtenir de nouvelles masses approximatives. Enfin, l'ensemble des étapes ci-dessus est répété jusqu'à ce qu'il reste k éléments focaux.

L'état de l'art montre trois limitations : 1) il n'y a pas de méthode générique de calcul des fonctions de masse. 2) Quand un grand nombre d'éléments focaux est présent dans le cadre de discernement, précisément pour les activités basées sur le contexte, la fusion des fonctions de masses requiert un temps de calcul élevé et une complexité de calcul importante. 3) Les approches proposées dans l'état de l'art sont essentiellement basées sur la suppression d'éléments focaux. Celles-ci ne permettent pas de mesurer les similarités entre les masses d'origine et celles approximées. Ceci ne permet pas de quantifier l'information perdue suite à cette approximation.

Afin de répondre à ces limitations, nous proposons dans ce chapitre d'explorer la théorie de Dempster-Shafer pour la reconnaissance des activités élémentaires et celles basées sur le contexte. Nous commençons par établir un premier modèle pour la reconnaissance des activités élémentaires. Nous proposerons deux approches permettant de calculer les fonctions de masses : une première approche basée sur la subdivision de l'espace des descripteurs en cellules. La seconde approche consiste à utiliser des classifieurs. Les fonctions de masses seront fusionnées à l'aide de la règle de combinaison de Dempster-Shafer.

Une fois le modèle de Dempster-Shafer, pour la reconnaissance de l'activité physique, établi nous l'utilisons comme base pour proposer une architecture pour la reconnaissance des activités physiques basées sur le contexte. L'ajout d'informations telles que les signaux audio, etc. implique l'agrandissement du cadre de discernement, qui va contenir toutes les autres activités physiques basées sur le contexte générées dans le chapitre 3, ce qui implique une complexité supplémentaire de calcul induite par les combinaisons des différents éléments

focaux dans le cadre de discernement. Nous proposons une approche permettant d'optimiser ce calcul en explorant la puissance des fonctions de corrélation.

Le schéma de la **figure 1** illustre l'ensemble des objectifs de ce chapitre.

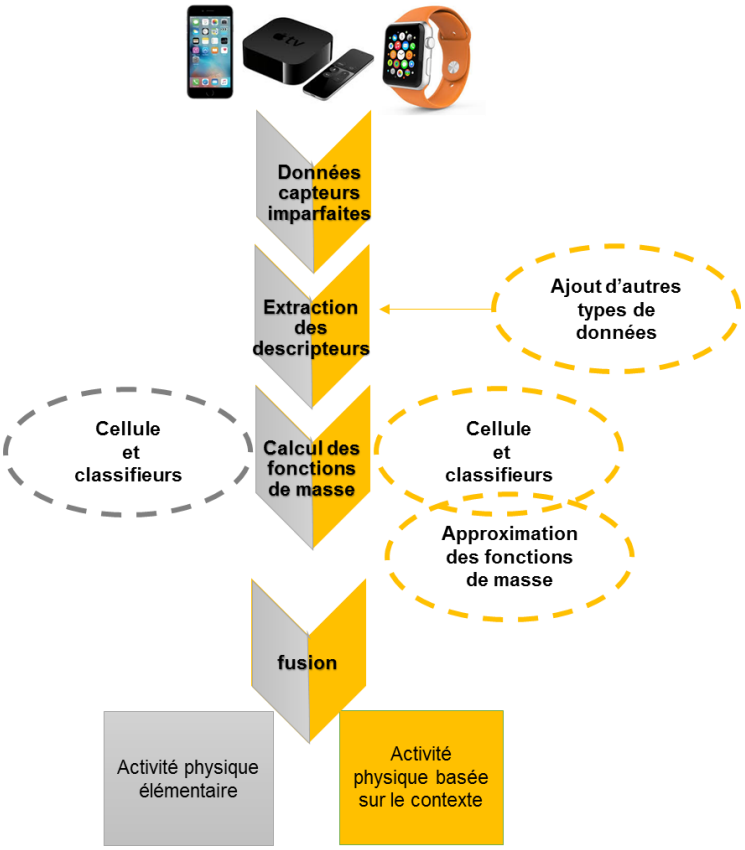


Figure 1 : Le procéssus de traitement proposé.

Une fois les données collectées des trois objets connectés, l'étape d'extraction de descripteurs permet d'extraire l'ensemble des descripteurs proposés dans le chapitre 2 et le chapitre 3. Une fois les descripteurs extraits, nous proposons deux approches pour calculer les fonctions de masse : i) une approche basée sur la subdivision de l'espace des descripteurs en cellules, et ii) une autre approche consiste à utiliser des classifieurs. Enfin, les masses ainsi calculées vont être fusionnées à l'aide de la règle de combinaison de Dempster-Shafer pour reconnaitre les activités physiques élémentaires. Afin de reconnaitre les activités physiques basées sur le contexte, nous poursuivons la même procédure : extraction de descripteurs, calcul des fonctions de masses. Une fois les fonctions de masses déterminées, nous appliquons un algorithme basée sur l'utilisation de la fonction de corrélation, qu'on va définir à la fin de ce chapitre, afin d'alléger les calculs en approximant les masses ainsi calculées. Les masses approximées seront fusionnées en utilisant la règle de combinaison de Dempster-Shafer pour prédire les activités physiques basées sur le contexte (Figure 1).

3. Vue d'ensemble de la DST pour la reconnaissance de l'activité

Dans cette partie, nous proposons de présenter les différents paramètres et mode de fonctionnement de la DST pour la reconnaissance de l'activité physique. Nous présenterons, ensuite, les différentes implémentations liées au calcul des fonctions de masse.

La théorie de Dempster-Shafer est un algorithme de fusion de capteurs qui donne une réponse à une question donnée, i.e. quelle activité est en train d'effectuer un participant ? La forme prise par la réponse est une distribution de probabilité parmis toutes les réponses possibles. Pour cette raison, il est très différent des algorithmes de classification utilisés dans le domaine de l'apprentissage automatique. Un schéma synthétique du principe de l'algorithme de la DST est présenté dans la **figure 2**:

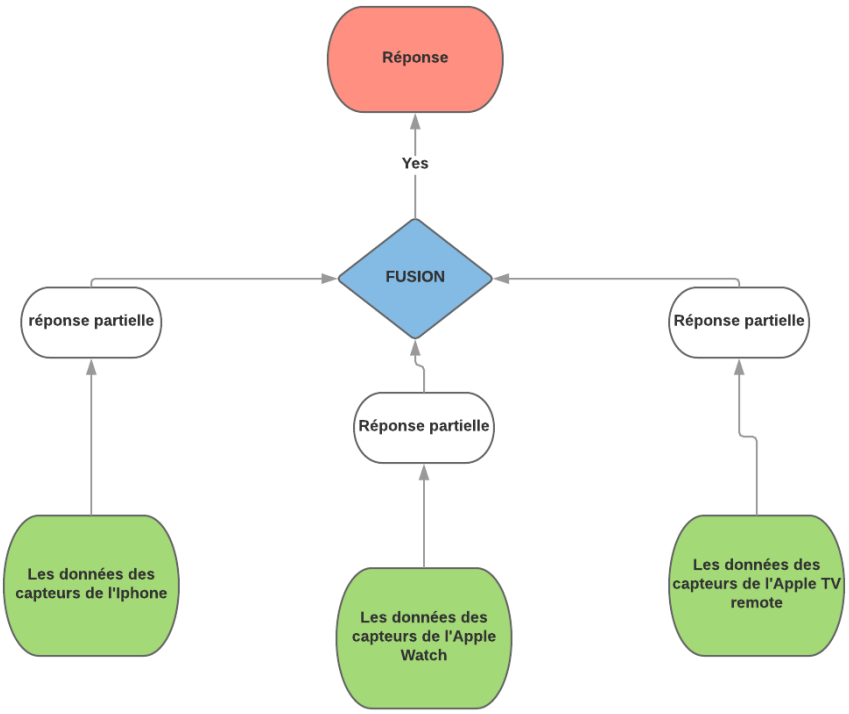


Figure 2 : Le processus de fusion des données de capteurs des trois objets connectés (Smartphone, SmartWatch, Smart TV).

L'objectif ici est de faire usage de tous les capteurs du smartphone, de la smartwatch et de la smart TV pour faire une prédiction plus fiable. Par exemple, le capteur de détection de lumière, le capteur de proximité et le gyroscope pourraient être utilisés pour déterminer la position du Smartphone (est-ce dans la main ou dans la poche). Les données du gyroscope et les données du magnétomètre pourraient également être utilisées pour aider la prédiction faite par l'accéléromètre si nous tenons compte de leur propre manque de fiabilité, ce qui est rendu possible par la DST.

Comme le montre la **figure 2** ci-dessus, chaque capteur donne une réponse à la question initiale qui a sa propre incertitude. Nous fusionnons ensuite les réponses partielles ensembles ce qui nous donne une réponse globale qui prend en compte les incertitudes impliquées des réponses partielles. La réponse issue de la fusion est supposée être plus fiable que les réponses partielles prises séparément.

Dans notre contexte de reconnaissance de l'activité physique basée sur le contexte, ou élémentaire, nous pourrions avoir le réseau de preuve suivant (**Figure 3**) :

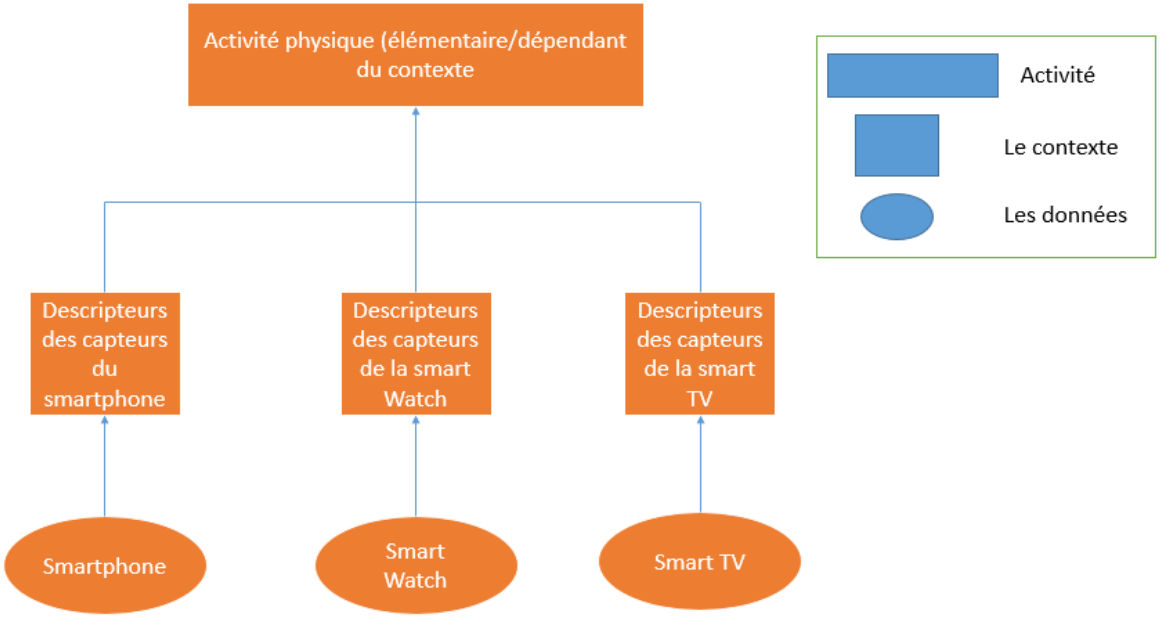


Figure 3 : Réseau d'ontologie reprenant l'univers de la théorie de Dempster-Shafer.

Le réseau contient trois niveaux : (1) le premier niveau concerne les différentes sources de données disponibles, dans notre cas, c'est l'ensemble des capteurs des trois objets connectés (Cf. Chapitre 2 et chapitre 3). (2) le deuxième niveau est celui des descripteurs extraits que nous pouvons calculer et extraire et enfin (3) le troisième niveau est celui de l'activité cible prédite : basée sur le contexte ou élémentaire.

L'implémentation de la DST nécéssite le calcul des éléments suivants :

a) Le cadre de discernement :

Nous définissons le cadre de discernement comme un ensemble de réponses possibles mutuellement exclusives à partir desquelles nous allons faire une prédiction.

Soit Θ notre cadre de discernement et dénote les différentes activités que l'utilisateur peut effectuer.

Θ = {« debout », « manger en étant assis », …}

Nous appelons un élément focal toute combinaison d'activités dans le cadre de discernement. Nous avons alors un total de $2^{\#\theta}$ éléments focaux.

b) Pièce d'évidence :

La valeur donnée par un capteur ou l'une des caractéristiques dérivées de sa valeur, parce qu'elle peut ne pas être exploitable telle quelle, est appelée élément de preuve ou pièce d'évidence. Cet élément de preuve donne une réponse subjective à la question posée, dans notre cas, la question est "quelle activité fait l'utilisateur?" en utilisant les trois objets connectés.

Les données fournies par ces capteurs sont des données temporelles. Ils ne sont pas très exploitables dans leur état brut. Des caractéristiques telles que "l'écart-type" sont extraites de ces données brutes et peuvent être considérées comme des éléments de preuve. Chacun d'eux donne une réponse différente possible à la question. Comme vu précédemment dans les chapitres 2 et 3, la DCT est retenue comme principal descripteur pour les activités physiques élémentaires et les descripteurs vus dans le chapitre 3 sont utilisés aussi pour l'activité basée sur le contexte.

c) Fonctions de masse :

La réponse de l'élément de preuve à une question donnée est représentée comme une fonction de masse. Une fonction de masse est une distribution de probabilité sur tous les éléments focaux.

$$\boldsymbol{m}: \mathbf{2}^{\#\boldsymbol{\theta}} \rightarrow [\mathbf{0}, \mathbf{1}]$$

Cette fonction de masse, également appelée affectation de croyance de base ou assignation de probabilité de base, a deux propriétés : 1) la masse de l'ensemble vide est égale à zéro et 2) la somme des masses est égale à 1, comme le montre l'équation ci dessous:

$$\boldsymbol{m}(\emptyset) = \mathbf{0}\ \boldsymbol{et} \sum_{\boldsymbol{A} \subseteq \mathbf{2}^{\#\boldsymbol{\theta}}} \boldsymbol{m}(\boldsymbol{A}) = \mathbf{1}$$

Les données que nous avons enregistrées au cours des expériences ont reçu une activité tirée du cadre de discernement. Cependant, nous considérons que les données des activités élémentaires et les activités basées sur le contexte, ces données ne prennent pas en compte tous les éléments focaux. Par conséquent, leur croyance de base peut être fixée à 0 et nous pouvons simplifier la propriété indiquée ci-dessus.

$$\sum_{\boldsymbol{A} \subseteq \boldsymbol{\theta}} \boldsymbol{m}(\boldsymbol{A}) = 1$$

d) Croyance et plausibilité :

La croyance mesure la force d'une preuve en faveur d'une réponse et est définie comme la somme des masses qu'elle entoure.

$$\boldsymbol{bel(A)} = \sum_{\boldsymbol{B|B \subseteq A}} \boldsymbol{m(B)}$$

Comme décrit et expliqué précédemment, seules les masses d'activité unique sont affectées aux valeurs différentes de 0. Nous avons donc cette simplification:

$$\boldsymbol{bel\ (A)} = \begin{cases} \boldsymbol{m(A)}, & \boldsymbol{si\ A \in \theta} \\ \boldsymbol{0}, & \boldsymbol{sinon} \end{cases}$$

Définissons la plausibilité de la réponse comme la mesure dans laquelle l'élément de preuve en faveur de la réponse opposée laisse place à la croyance dans la réponse.

$$\boldsymbol{pl(A) = 1 - bel(\overline{A})\ or\ pl(A)} = \sum_{\boldsymbol{B|B \cap A \neq 0}} \boldsymbol{m(B)}$$

Avec :

$$bel(A) < P(A) < pl(A)$$

e) La règle de combinaison de Dempster-Shafer

Le but de la théorie de Dempster-Shafer est de combiner les croyances de base fournies par des éléments de preuve afin de réduire l'incertitude de la réponse à la question posée.

La masse résultante donnée par la fusion de deux éléments de preuve est donnée par:

$$\boldsymbol{m_{1,2}(A) = m_1 \oplus m_2(A)} = \begin{cases} \boldsymbol{k^{-1}} * \sum_{\boldsymbol{B \cap C \neq A}} \boldsymbol{m_1(B) * m_2(C)}, & \boldsymbol{si\ A \neq \emptyset} \\ \boldsymbol{0}, & \boldsymbol{si\ A = \emptyset} \end{cases}$$

Avec k dénotant le nombre de conflits entre éléments de preuve:

$$k = \sum_{B \cap C \neq \emptyset} m_1(B) * m_2(C)$$

Nous pouvons fusionner plus de deux éléments de preuve avec:

$$\boldsymbol{m_{1 \oplus} m_2} \oplus \ldots \oplus \ldots = (\big((\boldsymbol{m_1} \oplus\ \boldsymbol{m_2}) \oplus \ldots\big) \oplus \ldots)$$

Étant donné que seules des activités uniques sont fournies avec une masse non nulle, nous pouvons simplifier la formule ci-dessus:

$$m_{1,2}(A) = \begin{cases} k^{-1} * m_1(A) * m_2(A), & si\ A \in \theta \\ 0, & sinon \end{cases}$$

Avec $k = \sum_{A \in \theta} m_1(A) * m_2(A)$

Nous allons présenter, dans la partie suivante, l'implémentation de la DST, nous présenterons étape par étape les différentes structures. Ainsi, nous mettrons le point sur (i) le calcul de la fonction de masse et (ii) l'approximation de l'assignation de probabilité de base afin d'alléger le temps de calcul.

3.1. Implémentation

Nous présenterons ci-dessous les différentes étapes et traitements adoptés pour la prédiction des activités physiques élémentaire et celles basées sur le contexte (**Figure 4**).

Notre implémentation de la DST est composée de trois étapes différentes: (a) l'extraction des caractéristiques à partir des données brutes des capteurs, (b) le calcul des fonctions de masse et (c) la fusion des fonctions de masse.

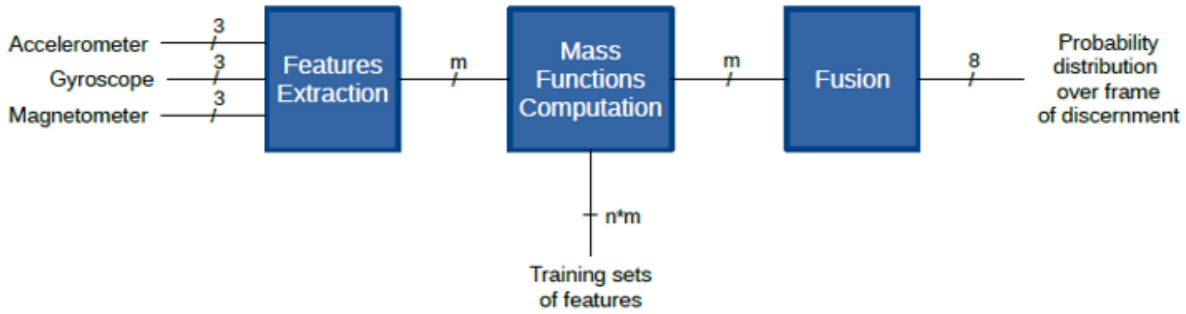


Figure 4 : Flux de traitement de la théorie de Dempster-Shafer. Ici nous présentons seulement le cas de la reconnaissance de l'activité physique élémentaire. Pour la reconnaissance des activités basées sur le contexte, le flux de traitement est identique.

Nous proposons d'illustrer cette implémentation dans un cadre général de la reconnaissance de l'activité physique. Nous montrons les résultats obtenus pour la reconnaissance des deux types d'activités élémentaires et basées sur le contexte.

Chaque capteur est décrit par ses valeurs générées suivants ses axes. Un nombre arbitraire de caractéristiques est extrait de ces valeurs. Cet ensemble de caractéristiques est comparé à des ensembles d'entités d'apprentissage, parce que leur activité correspondante est connue, résultant en un ensemble de fonctions de masse. Toutes ces fonctions de masse sont fusionnées pour donner une réponse à la question posée. La réponse prend la forme d'une distribution de probabilité sur l'espace des activités possibles.

a) Extraction des descripteurs :

L'extraction des caractéristiques à partir de données brutes est faite en trois étapes différentes (**Figure 5**): le filtrage des données, la création de fenêtres et le calcul des caractéristiques.

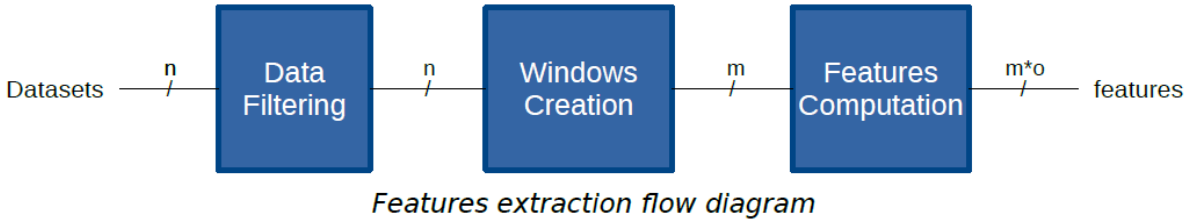


Figure 5 : Extractions de descripteurs dans la théorie de Dempster-Shafer.

L'extraction des descripteurs des activités élémentaires s'appuie sur la procédure présentée dans le chapitre 2. Pour les activités physiques basées sur le contexte, nous avons utilisé les descripteurs décrits dans le chapitre 3 (**Tableau 2**).

Tableau 2 : Ensemble des descripteurs extraits des différents capteurs embarqués dans les trois objets connectés.

Capteur	Descripteur extrait
Accéléromètre, gyroscope, magnétomètre	DCT
Détecteur de proximité	Moyenne, écart type
Capteur de luminosité ambiante	Moyenne, écart type
Capteur d'empreinte digitale	Moyenne, écart type
Microphone	• DCT coefficients • Zero-crossings
Cardiofréquencemètre	La variation de la fréquence cardiaque
Surface tactile	Moyenne, écart type

b) Calcul des fonctions de masse :

Le calcul des fonctions de masse est très important car il produit les réponses partielles, les réponses pour les descripteurs, à partir desquels nous obtiendrons la réponse globale. La qualité de la réponse partielle aura un impact sur la réponse globale, il est donc important de le faire d'une manière optimale.

Pour rappel, une fonction de masse pour un descripteur donné est une distribution de probabilité sur les éléments focaux, simplifiée pour les éléments du cadre de discernement ici, qui sont les différentes activités. De plus, la fonction de masse doit incorporer l'incertitude de la réponse, qui est représentée par la distribution de probabilité (**Figure 6**).

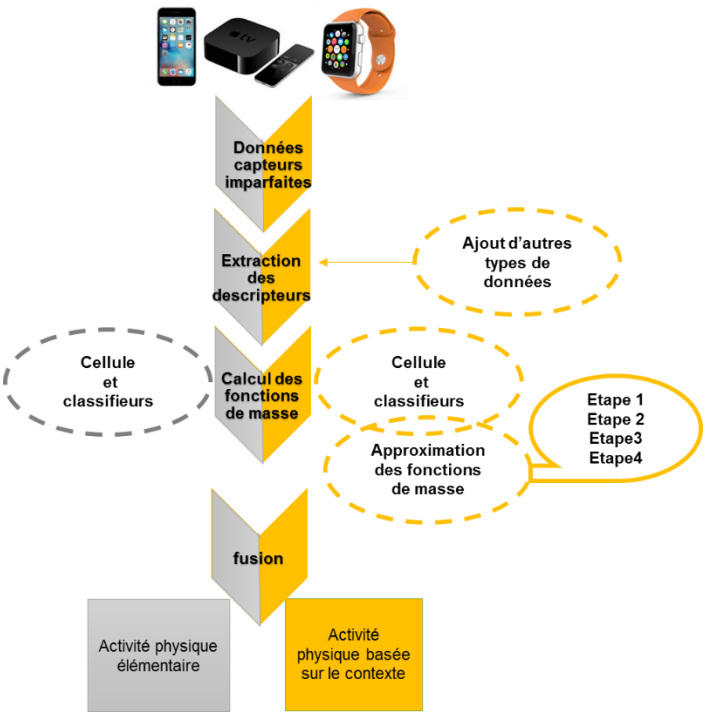


Figure 6: Démarche suivie pour le calcul et l'approximation des fonctions de masse.

L'état de l'art a montré qu'il n'existe pas de méthode générique pour calculer les fonctions de masse. Elles peuvent être dérivées de la connaissance d'experts, de statistiques ou de regroupement d'informations floues.

Ici, nous allons développer deux approches différentes pour calculer les assignations de probabilité de base. L'utilisation de connaissances spécialisées n'est pas une bonne approche car il est impossible de connaître l'activité simplement en connaissant la valeur de l'écart-type de l'axe «x» de l'accéléromètre, par exemple.

Notre première contribution consiste à diviser chaque espace de données des descripteurs en cellules qui ont toutes une distribution de probabilité spécifique sur les activités. Ces distributions de probabilité sont calculées statistiquement. La deuxième conribution consiste à prédire l'activité en utilisant des classificateurs et à ajouter de l'incertitude à la prédiction en fonction de l'incertitude du modèle.

b.1) Approche avec les cellules

Deux types différents de cellules ont été considérés: (i) les cellules à plage fixe et (ii) les cellules à taille fixe.

Les cellules à plages fixes voient leurs plages de valeurs fixées à un certain nombre en fonction du nombre de cellules. Par exemple, pour des valeurs comprises entre 0 et 1 et avec

un nombre de cellules égal à 10, la première cellule donnera des valeurs de 0 à 0,1, la seconde de 0,1 à 0,2, et ainsi de suite.

Une critique immédiate que nous pouvons donner aux cellules à plages fixes est le fait que cette méthode ne prend pas en compte la distribution des valeurs dans l'espace des valeurs des descripteurs.

Les cellules de taille fixe visent à résoudre ce problème. Toutes les cellules auront le même nombre de valeurs. Par exemple, supposons que nous disposions de 10 valeurs entre 0 et 0.1 et 10 valeurs entre 0.1 et 1. Si nous choisissons arbitrairement d'avoir 2 cellules différentes, la première cellule donnera des valeurs de 0 à 0.1 alors que la seconde cellule donnera des valeurs de 0.1 à 1. Les deux auront 10 valeurs dedans. Cette méthode permet ainsi de prendre en compte la distribution des valeurs sur l'espace des valeurs.

Après avoir divisé l'espace de valeur en cellules, nous pouvons calculer statistiquement la distribution de probabilité des valeurs sur les différentes activités.

b.2) Approches avec les classifieurs

L'utilisation de cellules comme décrit ci-dessus peut avoir des limites. Nous devons d'abord fixer un nombre arbitraire de cellules. Par ailleurs, même si des cellules de taille fixe ont aidé à améliorer la prise en compte la distribution des valeurs sur l'espace des valeurs, certains conflits peuvent survenir dans des zones très denses. La ségmentation des données peut être considérée comme un moyen de surmonter ces limites.

Cependant, la ségmentation ne prend pas en compte l'incertitude de la prédiction requise dans la DST. Ainsi, nous devons calculer la distribution de probabilité sur les activités et non pas une prédiction de l'activité.

Afin d'incorporer l'incertitude à la prédiction faite par le classificateur, nous pouvons estimer la précision du modèle en testant la prédiction de tout l'ensemble des caractéristiques les unes par rapport aux autres. En suivant cette démarche, nous pouvons calculer la précision de prédire une certaine activité, que nous pouvons maintenant utiliser pour ajouter de l'incertitude à notre prédiction.

Avant de faire une prédiction, nous devons évaluer l'incertitude associée à nos données d'entraînement. Pour ce faire, pour chaque donnée d'entraînement, nous prédirons son activité sur la base d'un modèle construit en utilisant les données d'entraînement restantes. Nous enregistrerons les prédictions ainsi que le nombre et le type d'activités prévues. Ceci nous permet d’obtenir une table de confusion pour chaque caracteristique. Une fois la table de

confusion construite, nous transformons ces valeurs en probabilités qui représentent les incertitudes en réalisant une prédiction pour un descripteur donné.

Une question se pose: comment choisir notre classificateur ? Pour identifier le (s) meilleur (s) classificateur (s), nous avons étudié le choix et la sélection de descripteurs.

Nous avons mené des études permettant de choisir les meilleurs descripteurs permettant d'aboutir à de meilleurs résultats de classification (Cf. Annexe D).

Ces études ont montré qu'il n'était pas possible de choisir un descripteur particulier. Ainsi, tous les descripteurs apporteraient une contribution à la précision finale. Ensuite, nous avons choisi d'étudier les performances de cinq classificateurs différents sur cette fonctionnalité: Naïve Bayes, k-Plus proches voisins, Arbre de décision, Machine à vecteurs de support, et Forêt aléatoire. Les résultats de ces études sont représentés dans la section 4.

c) Fusion :

Une fois que toutes les fonctions de masse ont été calculées, nous devons les fusionner. Pour ce faire, nous utiliserons la règle de combinaison simplifiée décrite précédemment:

$$m_{1,2}(A) = \begin{cases} (1-k)^{-1} * m_1(A) * m_2(A), & si\ A \in \theta \\ 0, & sinon \end{cases}$$

Avec $k = \sum_{B \neq C, B, C \in \theta} m_1(B) * m_2(C)$

La démarche ainsi définie est appliquée pour les deux types d'activités physiques que nous avons développées. La section suivante présente les résultats obtenus.

4. Résultats

4.1. Activités physiques élémentaires

Dans cette séction, nous commençons par présenter les résultats de calcul des fonctions de masse, puis les résultats relatifs à la prédiction des activités physiques.

4.1.1. Calcul des fonctions de masse

La figure 7 présente les résultats de calcul des fonctions de masse en utilisant la méthode des classifieurs (DST-classifieurs). Nous remarquons que l'algorithme des forets aléatoires permet d'avoir de meilleurs résulats de classification. Ce modèle sera donc retenu pour le calcul de la fonction de masse et la prédiciton de l'activité physique après la fusion par la règle de combinaison de Demspter-Shafer.

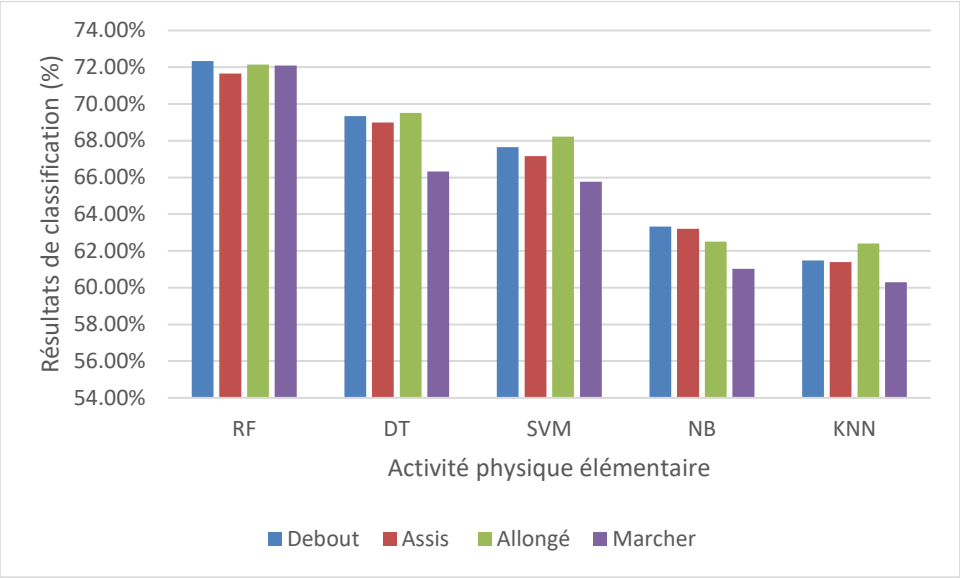


Figure 7 : Résultats de calcul des fonctions de masse par la méthode des classifieurs notée DST-classifieurs. Cinq algorithmes ont été testé : RF : forets aléatoires, DT : arbres de décision, SVM : machine à vescteurs de support, NB : naif de Bayes et KNN : K plus prochs voisins.

Les résultats de calcul des fontions de masse par la méthode des cellules à taille fixes (DST-Cel-TF) et cellules à plages fixes (DST-Cel-PF) sont présentés en annexe D.

Notre étude a montré que la méthode des classifieurs est plus précise que celle avec les cellules. En effet, le fait d'incorporer des classifieurs au lieu d'une ségmentation manuelle en céllules, a permi d'apprendre des modèles pouvant prédire les bonnes fonctions de masse à incorporer pour la prise de décision.

4.1.2. Prédiction de l'activité physique élémentaire

La **figure 8** présente les taux de détection des activités physiques élémentaires par rapport aux deux approches de calcul des fonctions de masse (DST-classifieurs, DST-Cel-TF et DST-cel-PF).

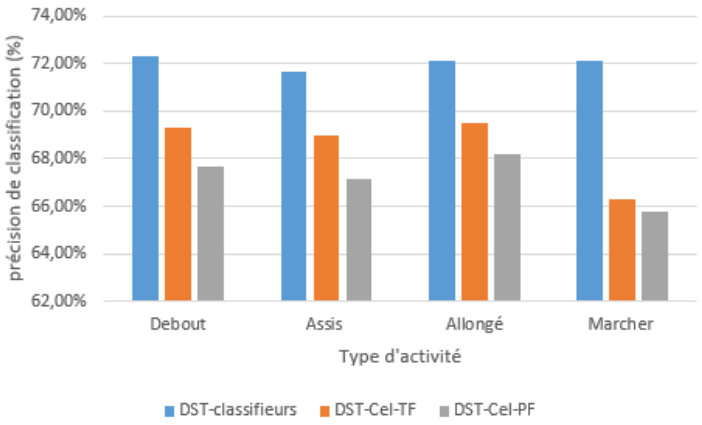


Figure 8 : Taux de détection des activités physiques élémentaires par application de la DST où (DST-classifieurs) représente la DST avec classifieurs, (DST-Cel-TF) représente la DST avec cellules à tailles fixes et enfin (DTS-Cel-PF) représente la DST avec cellules à plages fixes.

L'approche de calcul des fonctions de masse par classifieurs DST-classifieurs donne de meilleurs résultats par rapports aux deux autres méthodes. En effet, le calcul des fonctions de masse par classification permet de mieux regrouper les données homogènes (ou de même type). A contrario, les méthodes des cellules étaient basées sur la catégorisation « manuelle » sans prise en compte d'une quelconque structure. Ce qui était d'ailleurs prévisible d'après les résultats de classification trouvés dans le chapitre 2.

4.2. Activités physiques basées sur le contexte

Il s'agit ici de prendre en considération l'ensemble des données capteurs des trois objets connectés. L'ensemble Θ contiendra l'ensemble des activités basées sur le contexte répertoriées dans le chapitre 3.

La reconnaissance de l'activité physique basée sur le contexte a nécessité l'adaptation du mode de calcul au niveau de la fusion des fonctions de masse ainsi calculées. En effet, une fois les fonctions de masses calculées, elles ont été approximées afin de les fusionner avec la règle de combinaison de Dempster-Shafer. Pour cela, nous nous sommes inspirés des travaux de Shou et al [102].

4.2.1. Approximation des fonctions de masses

Pour un cadre de discernement Θ avec N éléments, supposons la masse de deux éléments de preuve désignés par m_1, m_2. Le coefficient de corrélation est défini comme:

$$r_{Masse}(m_1, m_2) = \frac{c(m_1, m_2)}{\sqrt{c(m_1, m_1).c(m_2, m_2)}} \qquad (8)$$

Où c(m_1, m_2) est le degré de corrélation noté par :

$$c(m_1, m_2) = \sum_{i=1}^{2^N} \sum_{j=1}^{2^N} m_1(A_i) m_2(A_j) \left| \frac{A_i \cap A_j}{A_i \cup A_j} \right| \quad (9)$$

Où $i, j = 1, .., 2^N$; A_i, A_j est les éléments focaux de la masse ; et | · | est la cardinalité d'un sous-ensemble.

Supposons qu'on ait une fonction de masse m avec L éléments focaux $A_1, \dots, A_L$. Puisque A_i est un élément focal constitué d'un ou plusieurs ensembles uniques, le A_i peut être divisé en un ou plusieurs éléments, qui peuvent être notés E_n $(1 \leq n \leq |A_i|)$. La masse de chaque élément est m (A_i) / | A_i|. Ces éléments E_n peuvent être composés d'un ensemble, qui peut être représenté par S_i.

La mise en œuvre de notre approche consiste à suivre quatre étapes. Supposons que le nombre désiré d'éléments focaux restants est k.

- **Étape 1 :** Retirer un élément focal A_i de m. Et la masse de A_i est assignée à l'A_j restant pour générer une nouvelle masse m_i' , où $i \neq j$ et i, j ∈ {1, ..., L}. La stratégie d'attribution est que l'ensemble S_i peut-être généré par A_i. Ensuite, si $E_n \cap A_j \neq 0$, où $E_n \in S_i$ et 1 ≤ n ≤ | A_i |, E_n de S_i est uniformément assigné en A_j. Si tous les éléments focaux A_j et E_n se croisent, la masse de E_n est normalisée pour tous les éléments focaux sortants A_j.
- **Etape 2 :** Calculer la similarité entre m_i' et m en utilisant le coefficient de corrélation noté r (i). Exécutez les étapes 1 et 2 pour tout i = 1, ..., L.
- **Etape 3 :** Trier le $A_i$de m, dans l'ordre croissant, en fonction de la similarité r, où i ∈ {1, ..., L}.
- **Étape 4 :** Retirer les k éléments focaux ayant le plus grand m en fonction du résultat du tri. La masse des éléments focaux supprimés est assignée aux éléments focaux sortants en fonction de la stratégie d'assignation à l'étape 1. Enfin, une nouvelle masse m peut être obtenue.

Un exemple illustratif de la nouvelle approximation de l'assignation de probabilité de base est donné en utilisant le coefficient de corrélation (voir annexe D).

Les fonctions de masses sont ainsi approximées, nous les avons fusionnées grâce à l'utilisation de la règle de combinaison de Dempster-Shafer. Les résultats de reconnaissance sont donnés dans **la figure 9**.

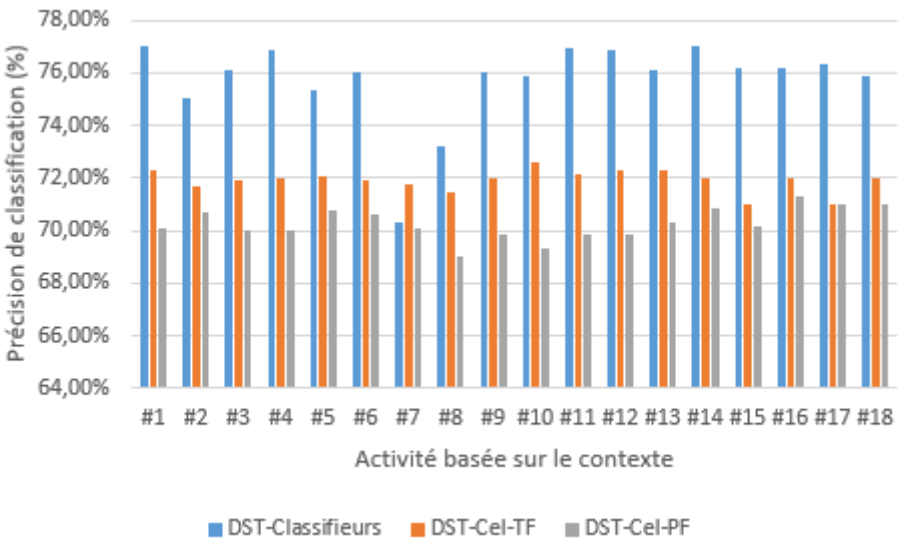


Figure 9 : Taux de prédiction des activités physiques basées sur le contexte.

Les taux de reconnaissance des activités physiques élémentaires étaient compris entre 68% et 72% selon que la fonction de masse est calculée avec les méthodes des classifieurs (DST-classifieurs), les méthodes des cellules à tailles fixes (DST-Cel-TF), et à plages fixes (DST-Cel-PF). Alors que le taux de détection des activités physiques basées sur le contexte est un peu plus haut (70%-76%).

Les précisions de reconnaissance des activités basées sur le contexte est bien meilleure que celle des activités physiques élémentaires, 68% et 76% de taux de reconnaissance pour les activités élémentaires et basées sur le contexte respectivement. Ces résultats peuvent etre expliquées par la quantité d'information utile apportée par les différentes sources de données. Par ailleurs, les activités basées sur le contexte sont enrichies par les différentes sources de données disponibles, ce qui n'était pas le cas des activités physiques élémentaires.

Le calcul de la fonction de masse, pour la reconnaissance de l'activité physique élémentaire et les activités physiques basées sur le contexte nécessite le parcours de tout l'ensemble du cadre de discernement. Ainsi, cela nécessite $2^{\#\theta}$ combinaisons d'éléments focaux (au minimum) [165]. Ceci génère un temps de calcul important avec une croissance en exponentielle mais aussi une complexité de calcul (l'ordre de $O(\frac{n^3}{3})$, dans le cas d'une optimisation par descente de gradient stochastique pour l'apprentissage du MLP vu dans le chapitre 3). Il est donc nécessaire d'optimiser les calculs en particulier lorsque ces calculs sont embarqués dans un objet connecté.

4.2.2. Etude comparative des méthodes d'approximation des fonctions de masse

L'état de l'art a montré le temps de combinaison entre les élements focaux est un facteur majeur de la complexité et lenteur des calculs [115, 149, 189]. Nous allons donc prendre en compte cette mesure afin de comparer entre une DST approximée et une DST non approximée.

Nous avons fait varier le nombre d'éléments focaux dans le cadre de discernement. Nous avons ensuite calculé les fonctions de masse par la méthode des classifieurs pour reconnaitre les activités physiques basées sur le contexte. Les masses calculées ont été fusionnées en utilisant la règle de combinaison de Dempster-Shafer. Deux cas ont été étudié : i) dans le premier cas, les masses calculées n'ont pas été approximées (cext-DST-classifieurs). Ii) dans le second cas, les masses calculées ont été approximées en utilisant notre méthode décrite plus haut. Puis, les masses ainsi approximées ont été fusionnées en utilisant la règle de combinaison de Dempster-Shafer (cext-DST-Classifieurs-approximée). Nous avons fait varier le nombre d'éléments focaux dans les deux cas. Nous avons ensuite étudié l'influence de la variation des éléments focaux sur le temps de combinaion des éléments focaux. **La figue 10** présente les résultats obtenus.

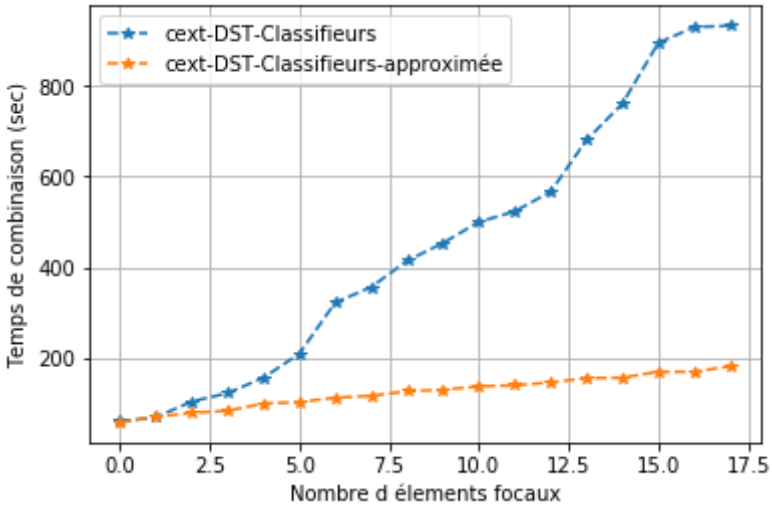


Figure 10 : Compraison entre la DST approximée appliquée aux activitésd physiques basées sur le contexte (cext-DST-Classifieurs-approximée) et la DST non approximée appliquées aux activités physiques basées sur le contexte (cext-DST-Classifieurs) en termes d'influence du nombre d'éléments focaux sur le temps de combinaison des éléments focaux.

Nous remarquons que sans approximation des fonctions de masse, le temps de combinaion des éléments focaux est directement proportionnel au nombre d'éléments focaux.

En revanche, le temps de combinaison des éléments focaux est beaucoup moins élévé dans le cas de la DST approximée. Donc l'approximation proposée est beaucoup moins gourmande en temps de calcul et combinaison des éléments focaux.

Nous allons ensuite étudier l'impact du nombre d'éléments focaux sur le temps de combinaison entre les éléments focaux. Cette étude a deux objectifs : 1) de montrer que le nombre d'éléments focaux influs sur le temps de combinaison entre les éléments focaux pour les deux types d'activités. 2) monter que les différentes méthodes de calcul des fonctions de masse définies auparavant sont impactées par l'augmentation du nombre d'éléments focaux ainsi que le temps de combinaison entre les éléments focaux.

Afin de répondre à ces deux objectifs, nous avons fait varier le nombre d'éléments focaux et calculer le temps de combinaison entre les éléments focaux pour les deux types d'activité. Nous avons aussi mis en évidence toutes les méthodes de calculs de fonctions de masse. Les résultats de cette étude sont présentés dans **la figure 11.**

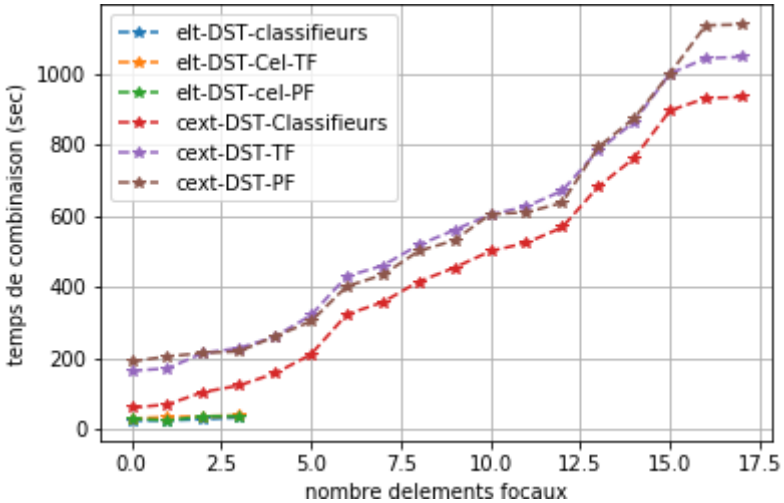


Figure 11 : Evolution du nombre d'éléments focaux dans le cadre de discernement en fonction du temps de combinaison entre les éléments focaux. Où Elt-DST-classifieurs représente l'application de la DST avec la méthode des classifieurs pour la reconnaissance des activités physiques élémentaires. Elt-DST-TF représente la reconnaissance des activités élémentaires par la méthode de DST à cellules à tailles fixes. Elt-DST-PF représente la reconnaissance des activités élémentaires par la méthode de DST à cellules à plages fixes.

Nous pouvons voir clairement que :

- Le temps de combinaison entre les éléments focaux est sensible au nombre d'éléments focaux dans les deux cas d'activités.
- Le temps de combinaison est plus court pour les activités élémentaires que pour les activités basées sur le contexte.
- Toutes les approches de calculs de fonctions de masse que nous avons proposés plus haut réagissent de la même manière (memes tendances). En effet, plus le nombre d'éléments focaux est élevé, le temps de combinaison entre les éléments focaux devient grand.

Afin de valider notre approche d'approximation des fonctions de masse, nous avons mené une étude comparative. Cette étude met en évidence quatre méthodes d'approximation de fonctions de masse déjà bien connues dans l'état de l'art, notées parApp_1, App_2, App_3 et App_4 définies dans **le tableau 3**, ainsi que notre méthode. Nous proposons de fixer les différents paramètres mis en jeux dans cette étude comparative comme suit :

Supposons que tous les capteurs des trois objets connectés décrivent tous les activités citées dans le cadre de discernement suivant Θ = $\{f_1, f_2, f_3, f_4, f_5\}$. Pour une activité cible, deux capteurs renvoient deux masses en termes d'assignation définis sur Θ, respectivement.

$$\boldsymbol{m_1(\{f_2, f_3\}) = 0.0326, m_1(\{f_3\}) = 0.0638}$$

$$\boldsymbol{m_1(\{f_1, f_2, f_3\}) = 0.3859, m_1(\{f_2, f_4\}) = 0.3130}$$

$$\boldsymbol{m_1(\{f_1, f_3, f_4\}) = 0.0597, m_1(\{f_3, f_5\}) = 0.1450}$$

$$\boldsymbol{m_2(\{f_1, \theta_3\}) = 0.1305, m_2(\{f_1, f_3, f_4\}) = 0.1123}$$

$$\boldsymbol{m_2(\{f_3, f_4\}) = 0.1438, m_2(\{f_1, f_2, f_4\}) = 1320}$$

$$\boldsymbol{m_2(\{f_1, f_2, f_3, f_4\}) = 0.1804, m_2(\{f_1, f_4\}) = 0.2010}$$

$$\boldsymbol{m_2(\{f_3, f_5\}) = 0.1000}$$

Le **tableau 3** détaille les résultats des différents calculs des différentes masses selon les quatre approches citées auparavant.

Tableau 3: Détail du calcul des masses selon les différentes techniques de l'état de l'art ainsi que notre technique proposée.

f	Assignation d'origine	App_1	App_2	App_3	App_4	Notre méthode
$\{f_1\}$	**0.0911**	0.1020	0.0872	0.1130	0.1082	**0.1046**
$\{f_2\}$	**0.0051**	0	0	0.0063	0	0
$\{f_3\}$	**0.2826**	0.2287	0.1843	0.3024	0.2808	**0.2815**
$\{f_4\}$	**0.1680**	0.1882	0.1960	0.1572	0.1505	**0.1467**
$\{f_5\}$	**0**	0	0	0	0	**0**
$\{f_1, f_2\}$	**0.0598**	0.0670	0.0573	0.0742	0.0710	**0.0717**
$\{f_1, f_3\}$	**0.1191**	0.1335	0.1622	0.0734	0.0811	**0.0824**
$\{f_1, f_4\}$	**0.0233**	0.0261	0.0209	0	0.0277	**0.0282**
$\{f_1, f_5\}$	**0**	0	0	0	0	**0**
$\{f_2, f_3\}$	**0.0069**	0	0	0.0086	0	**0**
$\{f_2, f_4\}$	**0.1148**	0.1286	0.1099	0.1424	0.1364	**0.1379**
$\{f_2, f_5\}$	**0**	0	0	0	0	**0**
$\{f_3, f_4\}$	**0.0101**	0.0113	0.0149	0	0.0120	**0.0124**
$\{f_3, f_5\}$	**0.0170**	0	0.0346	0.0211	0.0202	**0.0195**
$\{f_4, f_5\}$	**0**	0	0	0	0	**0**
$\{f_1, f_2, f_3\}$	**0.0817**	0.0916	0.0783	0.1014	0.0971	**0.0992**
$\{f_1, f_2, f_4\}$	**0**	0	0.0137	0	0	**0**
$\{f_1, f_2, f_5\}$	**0**	0	0	0	0	**0**
$\{f_1, f_3, f_4\}$	**0.0205**	0.0230	0.0220	0	0.0150	**0.0159**
$\{f_1, f_3, f_5\}$	**0**	0	0	0	0	**0**
$\{f_1, f_4, f_5\}$	**0**	0	0	0	0	**0**
$\{f_2, f_3, f_4\}$	**0**	0	0	0	0	**0**
$\{f_2, f_3, f_5\}$	**0**	0	0	0	0	**0**
$\{f_2, f_4, f_5\}$	**0**	0	0	0	0	**0**
$\{f_3, f_4, f_5\}$	**0**	0	0	0	0	**0**
$\{f_1, f_2, f_3, f_4\}$	**0**	0	0.0187	0	0	**0**
$\{f_1, f_2, f_3, f_5\}$	**0**	0	0	0	0	**0**

Où:

- Assignation d'origine ou masse d'origine: Les assignations originales ont été fusionnées en utilisant la règle de combinaison Dempster-Shafer.
- App_1: L'assignation de probabilité de base a été approximé par la méthode *k-l-x* et ont été fusionnés en utilisant la règle de combinaison de Dempster. Lorsque nous utilisons k-l-x pour supprimer un élément focal dans m_1 et m_2, respectivement, le paramètre est défini sur k = l = 5, x = 0,1 pour m_1 et k = l = 16, x = 0,1 pour m_2. Ensuite, les deux

assignations approximées sont fusionnées en utilisant la règle de combinaison de Dempster Shafer.

- App_2: Les assignations approximatives ont été obtenues par méthode du résumé et ont été fusionnés en utilisant la règle de combinaison de Dempster-Shafer.
- App_3: La méthode d'approximation fondée sur la fusion au niveau de la classe. Les assignations ont été fusionnées en utilisant la règle de combinaison de Dempster. Lorsque nous utilisons cette méthode pour supprimer un élément focal en m_1 et m_2, respectivement, le paramètre est fixé à k = l = 5, x = 0,1, α = 0,5 pour m_1 et k = l = 6, x = 0,1, α = 0,5 pour m_2.
- App_4: La méthode d'approximation fondée sur la distance de la preuve a été fusionnée à l'aide de la règle de combinaison de Dempster Shafer.
- Notre méthode: Les assignations approximatives par notre méthode ont été fusionnées en utilisant la règle de combinaison de Dempster Shafer.

Grâce au calcul ci-dessus, le résultat de la fusion des fonctions approximatives et des fonctions d'origines peut être obtenu. Nous remarquons que les différences entre les masses d'origines et les masses approximatives n'est pas très significatif. Ceci veut dire que notre approche d'approximation est sans perte d'information.

Nous avons donc gardé la DST avec classifieurs comme notre approche de référence et nous avons décidé de la comparer avec les autres approches listées dans l'état de l'art. Voici les résultats trouvés dans le graphique de la **Figure 12**.

Dans cet exemple, nous avons fixé les paramètres suivants k = l, x = 0,1 dans la méthode k-l-x et α = 0,5.

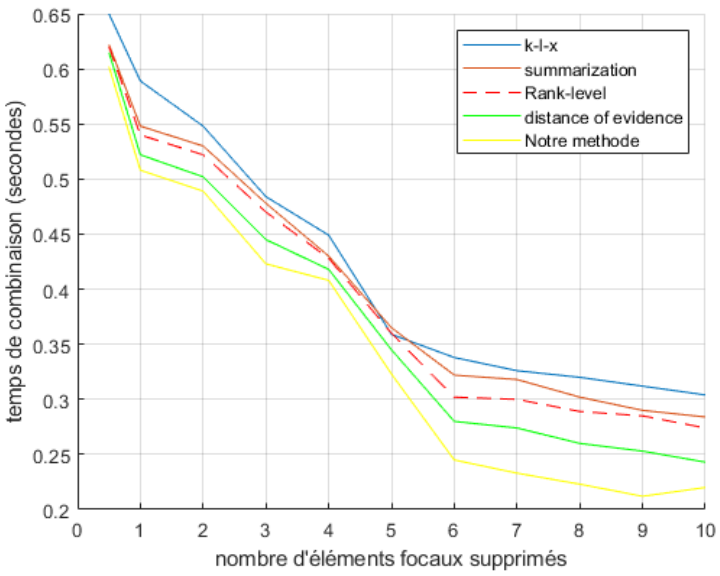


Figure 12 : Comparaison entre les différentes implémentations de la fusion de données en utilisant notre méthode (DST-classifieur).

Nous remarquons qu'en supprimant des éléments focaux du cadre de discernement, le temps de combinaison diminue. Par ailleurs, avec notre méthode, la diminution est plus marquante par rapport aux autres approches.

5. Validation et discussion des résultats

Afin de mieux visualiser l'impact de l'optimisation du calcul des fonctions de masse, nous proposons d'appliquer la DST-classifieurs à notre problème de détection de l'attention à la TV (vu dans le chapitre 3). Pour rappel, notre système de détection de l'attention à la TV était composé de trois sous-systèmes : i) un système de reconnaissance de l'activité physique, ii) un système de détection du sommeil et enfin iii) un système d'interaction sociale (voir **Figure 13**).

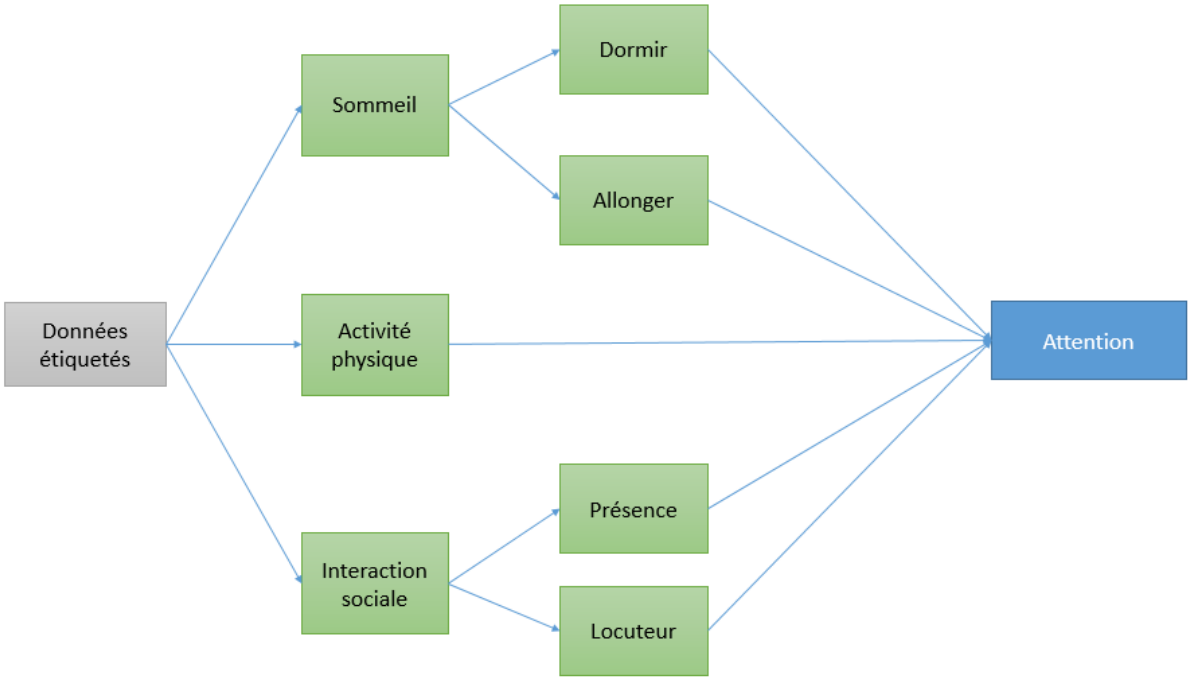


Figure 13 : Système de détection de la signature comportementale.

Dans le chapitre précédent, l'utilisation d'algorithmes d'apprentissage automatique ont permis d'avoir une précision moyenne de 70.21%. Nous proposons de voir ce que cela donne en

appliquant à la fois une DST-classifieur optimisée et une DST-classifieurs non optimisé (voir **Figure 14**). Les résultats sont résumés dans le **Tableau 4.**

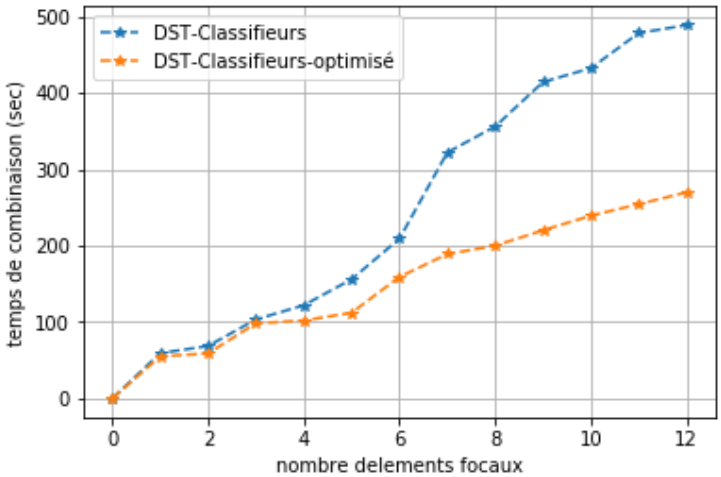


Figure 14 : Comparaison des résultats de prédiction de l'attention à la TV en utilisant la théorie de Dempster-Shafer ainsi que l'approximation de la fonction de masse par la méthode des classifieurs.

Nous constatons que le temps de combinaison de l'algorithme de DST (DST-Classifieurs-optimisé) est beaucoup moins important que celui de la DST classique, avec notre approche de classifieurs (DST-Classifieurs).

Nous proposons de voir si la précision de classification a changé, par rapport à l'utilisation des algorithmes d'apprentissage. Le **tableau 4**, compare entre les deux approches.

Tableau 4 : Comparaison entre le système de fusion avec optimisation de la fonction de masse et le système de fusion avec concaténation.

	SVM	*MLP*	*NB*	*RF*	*DST-Classifieurs-optimisé*
Système 1	*85.5%*	*70.1%*	*55%*	*70.8%*	*71.33%*
Système 2	*73.1%*	*66.6%*	*57.5%*	*66.05%*	*70.95%*
Système 3	*75.4%*	*69.47%*	*60.04%*	*70.1%*	*66.32%*
Concaténation	*70.21%*	*64.41%*	*61.01%*	*69.02%*	*68.1%*
Fusion DST	--	--	--	--	*69.05%*

Les algorithmes SVM, MLP, NB et RF ont relativement une bonne précision de reconnaissance. La concaténation des descripteurs des trois sous-systèmes donne une précision autour de 70.21%. L'application de la DST a permis de distinguer deux critères : (i) la possibilité de prendre en compte l'incertitude et la confiance sur les données provenant des trois systèmes, ce qui est en cohérence avec l'état de l'art. (ii) d'optimiser l'application de la DST à notre cas d'étude. Par ailleurs, la concaténation des fonctions de masses des sous-

systèmes a permis d'avoir pratiquement la même précision de reconnaissance que celle des algorithmes d'apprentissage automatique. La précision de reconnaissance apportée par la DST est certes un peu moins élevée que les méthodes de l'apprentissage automatique, mais cela reste de l'ordre de moins de 2% de différence, qui pourrait s'agir d'un bruit, dans une grande partie des études similaires.

Sur le plan applicatif, il serait plus intéressant de mettre en place, dans ce type d'application, une approche prenant en compte des incertitudes et avoir une précision moins bonne que d'appliquer des techniques d'apprentissage automatique. A contrario, il existe des domaines tels que la santé, par exemple, qui nécessite d'avoir la donnée la plus fiable possible.

Conclusion

Dans ce chapitre, nous avons pu proposer une méthode de reconnaissance des activités physiques basées sur l'utilisation de la théorie de Demspter-Shafer. En effet, la nature hétérogène de nos sources de données exige de faire appel à des méthodes ayant pour particularité la prise en compte des incertitudes des données provenant de ces capteurs.

De ce fait, nous avons exploré la méthode de la DST et proposé des améliorations notamment sur le calcul et l'optimisation de la fonction de masse adaptée à la reconnaissance de l'activité physique.

Nous avons proposé une approximation des fonctions de masse afin de limiter l'impact de la combinaison de plusieurs éléments focaux pour la prédiction de l'activité physique et ainsi d'aller un peu plus vite dans le calcul. Nous avons aussi démontré que cela n'a pas d'impact sur la précision de reconnaissance de l'activité physique et donc en terme de perte d'information significative.

Nous avons aussi proposé une étude comparative en implémentant les principales techniques d'approximation existantes dans l'état de l'art et avons comparé avec notre approche en termes de précision de reconnaissance et de perte d'information.

L'approche de prédiction des activités physiques par la théorie de Dempster-Shafer a montré que la prise en compte des incertitudes ne permet pas d'obtenir des algorithmes robustes. En effet, les taux de détection sont compris entre 61% et 80%. De plus, elle nécessite une sélection et calcul des descripteurs qui peut très rapidement devenir fastidieuse. Par ailleurs, certaines applications liées à la reconnaissance de l'activité physique (de bien-être et du sport, par exemple) basées sur le contexte, se focalisent sur la qualité de l'algorithme en terme de précision et robustesse.

Comment peut-on donc améliorer la précision de reconnaissance tout en éliminant la sélection de descripteurs ? Ou comment pouvons-nous bénéficier de la suppression des descripteurs tout en ayant des modèles robustes ?

Nous proposons d'aborder, dans le chapitre 5, de nouvelles techniques d'apprentissage automatique pouvant aider à répondre à ces limitations.

Chapitre 5

Optimisation de l'architecture d'entrainement des données de l'activité physique par apprentissage en profondeur

1. Introduction …… ....…153
2. Etude de la séléction de descripteurs pour l'apprentissage en profondeur...…….154
2.1. Avec séléction de descripteurs .....……154
2.2. Sans séléction de descripteurs……..…….156
3. Etude de l'optimisation de l'entrainement du modèle CNN-brutes……... …….....157
3.1. L'architecture proposée.……..…….158
3.2 Résultats et discussion …….....……161
Conclusion…..…….....……177

Chapitre 5

1. Introduction

Les méthodes d'apprentissage automatiques explorées dans les chapitres 2 et 3 ont permis d'atteindre de bonnes performances de reconnaissance des activités. Toutefois, ces algorithmes nécessitent une étape de calcul et sélection de descripteurs, très fastidieuse à mettre en place.

L'exploration des réseaux de neurones profonds, dans différents domaines tels que le traitement d'images, le traitement de la parole, etc, a permis de répondre à ces limitations. En effet, ce type de modèles permet d'apprendre des formes sans avoir à sélectionner des descripteurs. Des résultats très satisfaisants ont été atteints. La mise en place d'un tel type d'approches, pour la reconnaissance de l'activité, nécessite une adaptation des modèles existants, tel que les modèles développées dans le traitement d'images, par exemple, afin qu'ils soient exploitables pour la reconnaissance des activités physiques. En particulier, les modèles de réseaux de neurones convolutifs CNN, qui nécéssitent soit un redimentionnement des données d'entrées, soit une modification du noyau de convolution [206-209].

L'état de l'art a montré que les modèles de réseaux de neurones profonds utilisés pour la reconnaissance de l'activité physique étaient très vastes. En effet, suivant le nombre d'activités à reconnaitre et le type de données à entrainer, l'architecture du modèle devient de plus en plus profonde (plusieurs couches). Ceci peut causer plusieurs problèmes tels que la disparition et/ou d'explosion de Gradients et la lenteur du processus d'entrainement. Enfin, un tel type de modèles contenant des milliers de paramètres risquera fort de conduire au surajustement du modèle.

Nous proposons dans ce chapitre de répondre à ces limitations à travers trois points : i) nous mènerons une étude approfondie afin de voir l'impact du choix et sélection de descripteurs pour la reconnaissance des activités élémentaires et basées sur le contexte. Un modèle de réseau de neurones profonds sera proposé et entrainé sur la base des descripteurs ainsi extraits. Un deuxième modèle d'apprentissage profonds sera proposé, celui-ci prendra comme entrée toutes les données brutes des activités physiques. Une étude comparative entre les deux modèles ainsi développés sera proposée. Ii) Nous proposerons une étude approfondie permettant de dégager les meilleures paramètres du modèle d'apprentissage profond afin de prendre en compte l'explosion et/ou la disparition des gradients. Afin d'éviter le sur-apprentissage, le réseau de neurones profonds entrainé pour les activités physiques élémentaires sera réutiliser pour détecter les activités physiques basées sur le contexte.Le modèle ainsi obtenu sera optimisé pour répondre aux problèmes de disparitions/explosion des gradients ainsi que la lenteur du procéssus d'entrainement. Iii) Les résultats trouvés dans

l'étape précédente permettent d'avoir un modèle plus rapide. En revanche, celui-ci est moins précis. Nous avons proposé d'utiliser une heuristique permettant d'améliorer les précisions de reconnaissance tout en garantissant la rapidité du procéssus d'entrainement. Cette heuristique est appelée l'optimisation par essaim particulaire (PSO).

2. Etude de la séléction de descripteurs pour l'apprentissage en profondeur

L'objectif de cette séction est de développer deux modèles de réseaux de neurones profonds afin de reconnaitre les activités physiques élémentaires et basées sur le contexte. Deux critères seront étudiés : i) la séléction de descripteurs. Ii) sans séléction de descripteurs. Nous commençons par établir les architectures des modèles de reconnaissance de l'activité avec séléction de descripteurs. Puis, nous présenterons les architectures des modèles de reconnaissance des données brutes de l'activité physique. Enfin, nous présenterons les résultats obtenus.

2.1. Avec séléction de descripteurs

2.1.1. L'architecture proposée

Notre première étude a été menée avec une architecture classique de réseau de neurones profond avec une couche d'entrée correspondant aux données capteurs, une ou plusieurs couches cachées et enfin, une couche de sortie correspondant aux activités physiques à reconnaitre. L'entrainement de ce modèle sera effectué avec des descripteurs sélectionnés préalablement (Cf. Chapitre 2 et Chapitre 3).

La reconnaissance des activités élémentaires s'appuie sur les données des capteurs suivants : accéléromètres, gyroscopes et magnétomètres des trois objets connectés. Pour reconnaitre les activités physiques basées sur le contexte, nous rajoutons l'ensemble des données des capteurs, comme vu dans le chapitre 3.

Nous commençons par proposer un modèle d'apprentissage afin de reconnaitre les activités physiques élémentaires. La procédure d'implémentation de ce modèle, appelé DNN-DCT, consiste à suivre les étapes suivantes : après avoir divisé les signaux temporels en fenêtres à durées égales, nous avons extrait la DCT comme principal descripteur (voir chapitre 2) et concaténé tous les descripteurs extraits. Enfin, nous avons entrainé le modèle DNN-DCT en utilisant les descripteurs extraits.

Le réseau de neurones ainsi proposé est composé de 5 couches cachées, le nombre de neurones dans la couche d'entrée correspond à la dimension des descripteurs et le nombre de neurones dans la couche de sortie correspond aux activités de sortie : 4 pour les activités élémentaires et 18 pour les activités basées sur le contexte.

Après avoir extrait les descripteurs, le modèle doit etre entrainé. Nous avons entrainé le DNN-DCT en suivant la procédure suivante : nous avons commencé par concaténer les descripteurs de 13 cadres (ou *frame*, en anglais) en incluant le centre du cadre, les 6 cadres précédents et les 6 cadres suivants en utilisant une propriété clé des réseaux de neurones profonds qui est la capacité à traiter des descripteurs de grandes dimensions et séries temporelles. Au total nous avons utilisé 1872 (144*13) comme dimension de descripteurs d'entrée pour le modèle. Ensuite, nous avons normalisé le vecteur d'entrée.

L'étape d'entrainement nécéssite un pré-entrainement afin de mettre en place les paramètres initiaux du modèle, puis un entrainement proprement dit. Pour cela, nous avons d'abord utilisé la méthode du RBM (Restricted Boltzmann Machine, en anglais) pour pré-entrainer le modèle. Puis, l'entrainement du réseau en utilisant une rétro-propagation du gradient sur les données labélisées. Ainsi, nous avons adopté deux approches. La première, la méthode du Dropout [166] avec un taux d'apprentissage fixe. La deuxième, consiste à fixer un taux d'apprentissage, et après chaque 5 époques, nous avons commencé à diviser le taux d'apprentissage par deux à chaque époque. Les détails du réglage de l'apprentissage du modèle sont résumés dans le **tableau 1**.

Afin de reconnaitre les activités physiques basées sur le contexte, nous avons utilisé la même architecture que précédemment, notamment les couches inférieures. Nous avons, ensuite, extrait les mêmes descripteurs que ceux du chapitre 3. **La figure 2** représente les résultats de classification obtenus. Nous avons conservé la même dénomination de modèles (DNN-DCT) afin de montrer que la DCT est le principal descripteur des données des accéléromètres, gyroscopes et magnétomètres des trois objets connectés.

Les résultats de reconnaissance sont présentés dans la section suivante. L'architecture du réseau de neurones profond utilisée est présentée dans **la figure 1.**

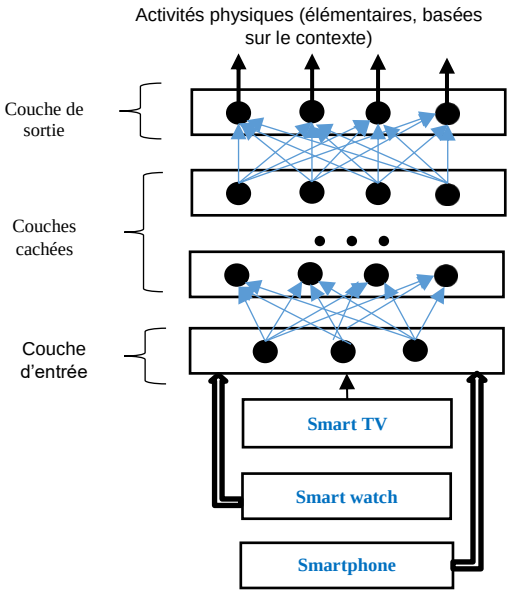


Figure 1 : Architecture de reconnaissance des activités physiques élémentaires et basées sur le contexte en utilisant le smartphone, la smartwatch et la smart TV. Les données des capteurs embarqués dans les trois objets connectés sont extraites et données en entrée du modèle ayant cette architecture.

Tableau 1 : Conditions de réglage du réseau de neurones profond (DNN-DCT).

Nombre de couches	8
Nombre de neurones dans les couches cachées	4096
Taux d'apprentissage	0.009 (sans dropout) 0.09 (avec dropout)
L2	0.0 (sans dropout) 0.000001 (avec dropout)
Taux d'abandon de la couche d'entrée	0.3
Taux d'abandon des couches cachées	0.6
Epoques	30 avec dropout 400 sans dropout

2.1.2. Résultats et discussion

La figure 2 présente les résultats de classification des activités physiques élémentaires. **La figure 3** présente les résultats de reconnaissance des activités physiques basées sur le contexte.

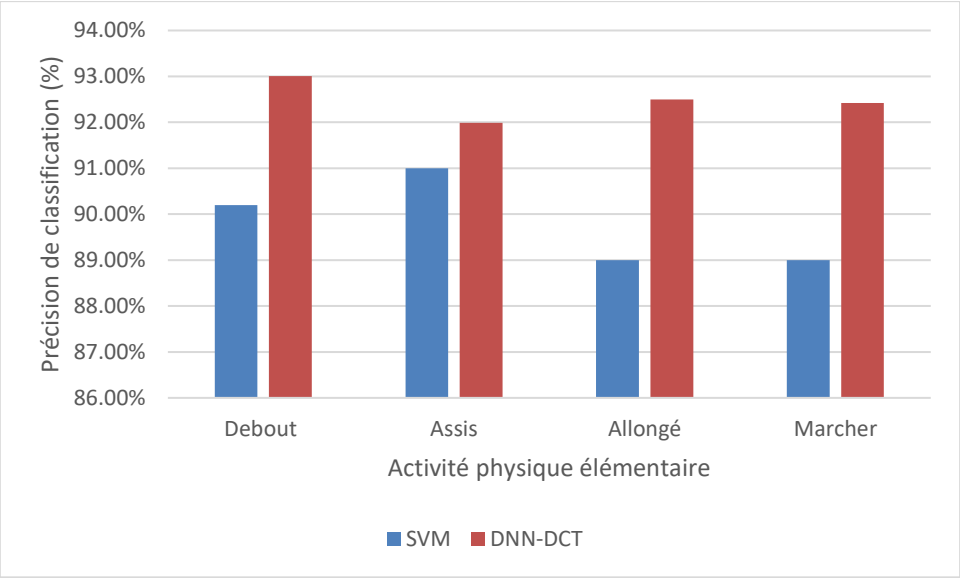


Figure 2 : Précisions de classification des activités physiques élémentaires.

Le réseau de neurones DNN-DCT permet d'atteindre une performance de plus 92.57% en moyenne. A titre de comparaison, ce modèle est meilleur que le SVM appliqué dans le chapitre 2 où une performance de 89% a été atteinte. Le caractère « profond » de l'algorithme d'apprentissage et sa capacité à apprendre des structures à grandes échelles a permis d'atteindre ce résultat. Les résultats ainsi obtenus sont comparables à ceux de l'état de l'art, notamment les travaux de luyou et al. [167] ayant obtenu une précision de 93% en détectant seulement deux activités et dans un environnement contrôlé.

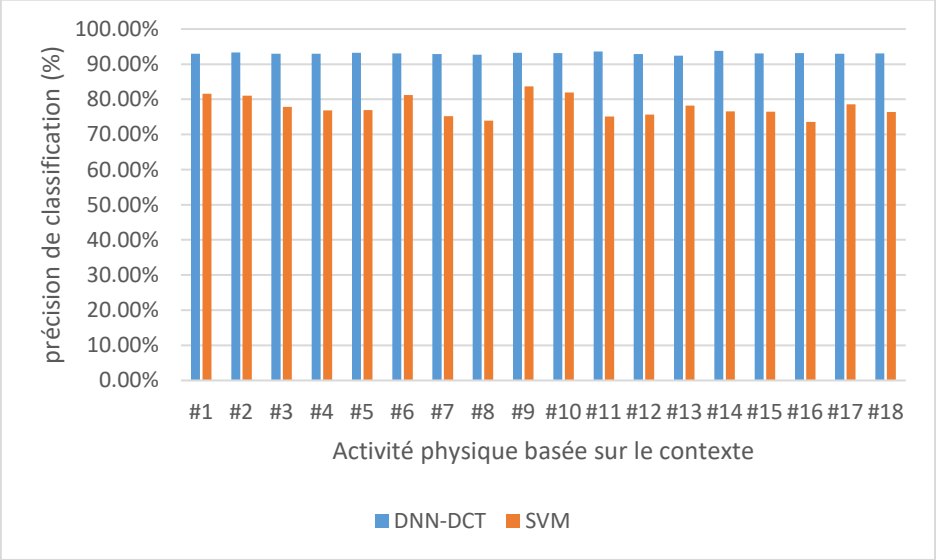


Figure 3 : Précision de classification des activités physiques basées sur le contexte (en utilisant le modèle DNN-DCT vs SVM).

La précision de reconnaissance des activités physiques basées sur le contexte est en moyenne égale à 93%, contre 92% pour les activités physiques élémentaires. Ceci s'explique

par l'incorporation d'informations supplémentaires provenant des autres capteurs. Par conséquent, le modèle DNN-DCT est meilleur que SVM à noyau Gaussien utilisé dans le chapitre 3. Nos résultats sont comparables à ceux de l'état de l'art, notamment le travail de Ordonez et Rodger [168] et Ronao et Chan [169] ayant détecté des activités basées sur le contexte dans un environnement contrôlé en utilisant des capteurs fixés à l'environement. La précision de reconnaissance atteinte était de 92% et 94%, respectivement.

Nous proposons dans la séction suivante de reconnaitre les activités physiques en entrainant un réseau de neurones profond avec les données brutes de l'ensemble des données capteurs.

2.2. Sans séléction de descripteurs

L'étude précédente a démontré le potentiel de la séléction de descripteurs pour la reconnaissance des activités physiques avec un réseau de neurones profond. Cependant, ce type d'approches est très fastidieux à mettre en place. Afin de remédier à ce problème, nous proposons dans cette deuxième étude d'explorer les données brutes renvoyées par les capteurs.

Parmi les différentes approches existantes, nous avons retenu les réseaux de neurones convolutifs (CNN) qui présentent une bonne alternative. Ce type de modèles a été appliqué avec succès dans différents domaines notamment en traitement d'image où les caractéristiques ont été automatiquement extraites des données brutes. Les résultats de reconnaissance des motifs dans des images étaient très satisfaisants, de l'ordre de 97% [166] [170].

Cette étude propose ainsi l'exploitation des réseaux de neurones profonds, notées CNN-brutes, afin de reconnaitres les activités physiques.

2.2.1. L'architecture proposée

Nous commençons par présenter l'architceture du modèle de reconnaissance des activités élémentaires, puis, nous présenterons l'architecture du modèle de reconnaissance des activités basées sur le contexte.

L'élaboration de notre modèle d'inspire des travaux de T. Plotz et al [171] [166]. **La figure 4** représente la structure du modèle. En suivant les paramétrages effectués dans [140], nous utilisons une fenêtre glissante avec une longueur de ωet avec un certain pourcentage de chevauchement pour extraire des données d'entrée pour le CNN.

L'architecture que nous proposons contient *C* couches avec trois types de couches :

a) Une couche d'entrée, avec h_i^0 neurones dont les valeurs sont fixées par les données d'entrée.

b) Une ou plusieurs couches cachées contenant h_i^l neurones dont les valeurs sont dérivées des couches précédentes.

c) Une couche de sortie avec h_i^L neurones dont les valeurs sont dérivées de la dernière couche cachée du réseau de neurones.

Comme c'est le cas des réseaux de neurones classiques, notre réseau de neurones CNN-brutes apprend en ajustant un ensemble w_{ij}^l de neurones, tel que w_{ij}^l représente le poids d'une certaine entrée de la couche h_i^l ayant comme sortie dans la couche h_j^{l+1}. Notons x_i^l l'entrée totale du neurone u_i^l ($i^{ème}$ neurone dans la couche l), et y_i^l sa sortie dans la coucheh_i^l).

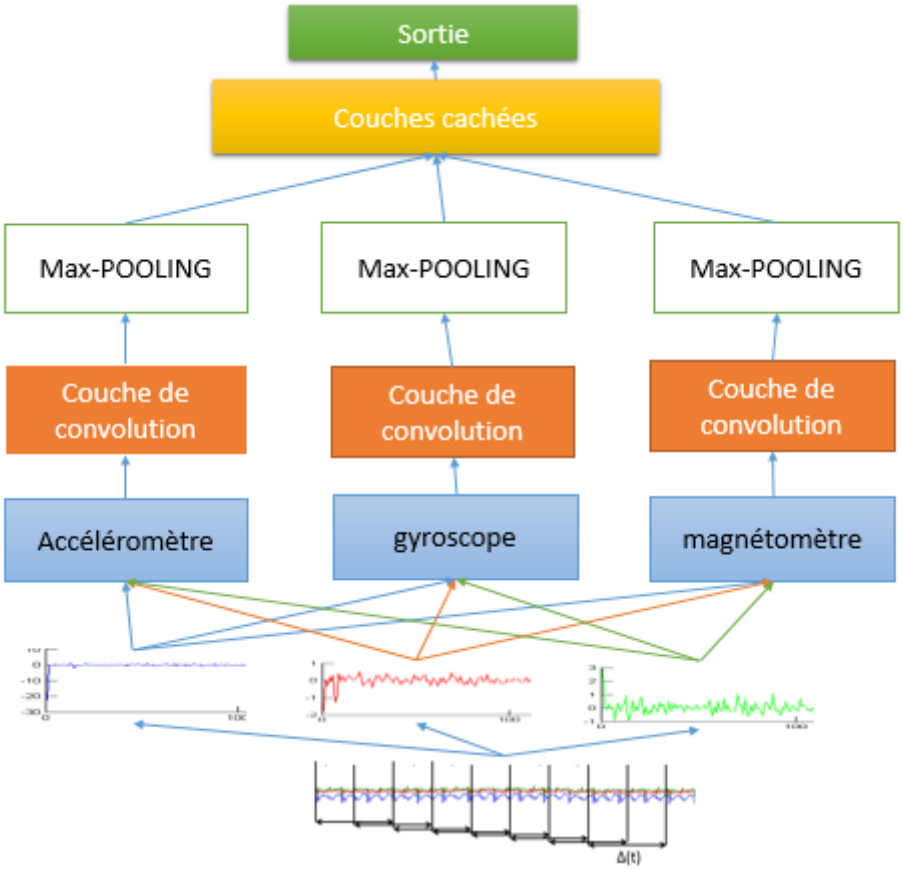


Figure 4 : Structure du modèle CNN-brutes pour la reconnaissance de l'activité humaine, élémentaire, dans ce cas précis. La dimension des données d'entrée est de 48, la sortie convolutionnelle de dimension 12, la sortie de dimension max-pooling est 4. Ce modèle contient trois couches cachées de dimension 1024, 64 et 30. La couche supérieure est un classificateur classique Softmax.

Les détails des différentes composantes de cette architecture du CNN-brutes sont donnés dans la partie suivante.

- **La couche de convolution :**

La couche de convolution sert à capturer les dépendances locales et les caractéristiques invariantes d'échelle des signaux d'activité. Afin de capturer les dépendances locales des données issues de l'accéléromètre, du gyroscope et du magnétomètre, nous appliquons une contrainte de connectivité locale entre les unités de couches adjacentes [166,182]. Cette contrainte s'inspire des structures bilogiques. En effet, nous savons qu'il existe un arrangement complexe de cellules dans le cortex visuel sensibles aux petites régions de l'entrée, appelées un champ réceptif, et sont corrélées pour générer tout le champ visuel. Ces filtres, appelés également filtres locaux, sont locaux dans l'espace d'entrée et sont donc adaptés pour exploiter la corrélation locale cachée dans les données [172-173]. Ainsi, le poids $w_{i,j}$ du bord connecté du i^{eme} neurone au j^{eme} neurone, peut être réduit par w_a, e t$w_{i,j} = w_{i,j} + m = w_a$, où m est la largeur du filtre local. L'opération de convolution est effectuée sur le sous-ensemble local. Cette contrainte topologique correspond à l'apprentissage d'une matrice de poids avec contrainte de parcimonie [189,201]. Ceci permet d'une part d'extraire des dépendances locales. D'autres part, il réduit également la complexité de calcul. Cet ensemble de filtres locaux constitue une carte de caractéristiques. À chaque position temporelle correspondant aux signaux d'entrées, différents types d'unités dans différentes mappes de caractéristiques, ou descripteurs, calculent différents types de descripteurs.

Afin de former une représentation plus riche des données, les couches convolutionnelles sont composées d'un ensemble de cartes de caractéristiques multiples,$x^{(.,j)}$, j = 1 ... J.

Supposons que nous ayons une couche de N neurones en entrée suivie d'une couche convolutionnelle. Si nous utilisons le filtre w de largeur m, la sortie convolutionnelle serait composée de (N - m + 1) neurones. Dans ce cas, la sortie de la couche convolutionnelle *l* est donnée par:

$$x_i^{l,j} = \sigma\left(b_j + \sum_{a=1}^{m} \omega_a^j x_{i+a-1}^{l-1,j}\right) \qquad (1)$$

Où $x_i^{l,j}$est la sortie de la $l^{ième}$couche de convolution de la $j^{ième}$ carte de caractéristiques du $i^{ième}$neurone, et σ est une application non linéaire qui exploite généralement la fonction tangente hyperbolique, notée tanh (·).

Dans un modèle CNN traditionnel [174], chaque filtre local est en outre répliqué sur l'ensemble de l'espace d'entrée. Cela signifie que les poids des filtres locaux sont liés et partagés, forcés d'être identiques, par toutes les positions des neurones dans l'ensemble de l'espace d'entrée. Les poids répliqués permettent de détecter les caractéristiques indépendamment de la position du neurone, ce qui est également bénéfique pour la préservation de l'invariance d'échelle.

Dans le domaine du traitement d'images [175], par exemple, le partage de poids complet est une bonne approche car le même motif d'image peut apparaître à n'importe quelle position dans une image. Cependant, dans le domaine de la reconnaissance de l'activité physique élémentaire et basée sur le contexte, le signal apparait à différentes unités et peut se comporter de manière très différente car différents motifs apparaissent dans une trame différente. Par conséquent, il peut être préférable de relâcher la contrainte de partage de poids, c'est-à-dire que les poids de la même couleur et du même type sont forcés à être identiques. Cette stratégie de partage de poids est décrite dans [132], et nous l'avons appelée partage de poids partiel dans notre étude. Cette technique permet ainsi de modifier l'activation de la fonction dans la couche de convolution comme suit :

$$w_{i,k}^{l,j} = \sigma(b_j + \sum_{a=1}^{m} \omega_{a,k}^{j} x_{i+(k-1)\times s+a-1}^{l-1,j} \qquad (2)$$

Où $x_{i,k}^{l,j}$ est l'un des $i^{ième}$ neurones de la $j^{ième}$ carte de la $k^{ième}$ section dans la $l^{ième}$ couche, et s est la plage de la séction. La différence entre l'équation (1) et l'équation (2) se situe au niveau de la plage de partage des poids. En effet, en utilisant la fenêtre $(k-1) \times s + i + a$ au lieu de$(i + a)$, ceci permet de conduire l'opération de convolution.

- **La couche de Max-pooling :**

Une fois qu'une caractéristique ait été détectée dans la couche convolutionnelle, son emplacement exact devient moins important, tant que sa position approximative par rapport aux autres caractéristiques est préservée. Par conséquent, une couche supplémentaire qui effectue une mise en pool maximale est empilée sur une couche convolutionnelle afin de réduire la sensibilité de la sortie. La conservation des descripteurs invariants à l'échelle est une autre caractéristique clé du CNN, celle ci est réalisée par la couche de mutualisation maximale (Max-pooling). Dans la couche Max-pooling, les descripteurs de la couche convolutionnelle sont divisés en plusieurs partitions. Dans chaque partition, nous appliquons l'opération max (décrite ci-dessous, l'équation 3) pour afficher les valeurs. La fonction d'activation dans la couche max-pooling dans le CNN traditionnel est donnée par:

$$x_i^{l,j} = max_{k=1}^{r} \left(x_{(i-1)\times s+k}^{l-1,j} \right) \qquad (3)$$

En raison de la structure de partage de poids partiel, décrite auparavant, seuls les filtres locaux proches les uns des autres partagent des poids et sont agrégés ensemble dans la couche de mutualisation maximale. La fonction de regroupement (équation 4) est légèrement différente de la fonction de regroupement (Pooling function, en anglais) traditionnelle, puisque l'opération max n'est effectuée que dans la même section de poids partagé:

$$x_i^{l,j} = max_{k=1}^{r}\left(x_k^{l-1,j}\right) \qquad (4)$$

Ainsi, les filtres qui fonctionnent sur la fenêtre temporelle locale fourniront un moyen efficace de représenter ces structures locales et leurs combinaisons tout au long de l'axe temporel. Ils pourront éventuellement être utilisés pour reconnaître différentes activités, notamment celles basées sur le contexte.

Le CNN peut contenir une ou plusieurs paires de couches de convolution et de regroupement maximal. Les couches supérieures utilisent des filtres plus larges pour traiter des parties plus complexes de l'ensemble de données d'entrées. Les couches supérieures de CNN-brutes sont empilées par un ou plusieurs réseaux de neurones normaux entièrement connectés. La finalité est qu'ils combinent différentes structures locales dans les couches inférieures pour la classification finale.

Comme nous venons de le décrire, nous n'utilisons qu'une couche de convolution et une couche de Max-pooling, et trois réseaux de neurones normaux entièrement connectés. Dans la phase d'apprentissage, les paramètres du CNN-brutes sont estimés par des algorithmes de propagation avant et arrière standards afin de minimiser la fonction objective. Nous proposons de détailler cela dans la partie suivante :

a) Propagation avant :

La propagation avant est réalisée dans les couches de convolution avec N neurones par l'équation (2). Dans le cas où nous utilisons m filtres ω, la sortie de la couche convolutionnelle partielle sera de taille$(N - m + 1)$. La sortie de la couche convolutionnelle alimente la couche de Max-Pooling, réalisée par l'équation (4).
Supposons que nous prenons une fenêtre de taille k et donnant une seule sortie qui représente le maximum dans cette fenêtre. Par exemple, si la couche d'entrée contient N neurones, sa sortie contiendra $\frac{N}{k}$ neurones dans la couche de Max-Pooling car chaque fenêtre k est réduite à une seule valeur via la fonction max. Puis, un réseau entièrement connecté est suivi de la couche max-pooling. La propagation avant est ainsi définie par :

$$x_i^l = \sum_j w_{j,i}^{l-1}\sigma\left(x_i^{l-1}\right) + b_i^{l-1} \qquad (5)$$

Plus haut dans la la structure du réseau, la fonction Softmax interviendra pour la classification finale des activités physiques (élémentaires et basées sur le contexte).

$$f(x) = argmax_c P(y = c|x) = \ argmax_c \frac{e^{x^T\omega_j}}{\sum_{k=1}^{K} e^{x^T\omega_k}} \qquad (6)$$

Où

- *c* est une étiquette de classe (activité basée sur le contexte ou activité élémentaire),
- *x* est une caractéristique d'échantillon,
- *y* est une variable d'étiquette, et w est un vecteur de poids,
- *K* est le nombre de classe (4 dans le cas des activités physiques élémentaires et 18 dans le cas des activités basées sur le contexte).

b) Propagation arrière

L'itération de propagation vers l'avant est réalisée, nous aurons la valeur d'erreur, avec la fonction de perte L, nous utilisons ici la norme L2. Nous sommes ainsi capables d'utiliser la descente de gradient pour mettre à jour les poids ω. Pour la couche entièrement connectée, le gradient peut être calculé avec la méthode classique de la manière suivante:

$$\frac{\partial L}{\partial w_{i,j}^{l}} = y_i^l \frac{\partial L}{\partial x_j^{l+1}} \qquad (7)$$

Telle que

- la fonction de mappage non linéaire est donnée par l'équation suivante $y_i^l = \sigma(x_i^l) + b_i^l$
- x_j^{l+1} est le neurone j dans la couche $(l+1)$, $x_i^l = \sum_j w_{i,j}^{l-1} y_j^{l-1}$.

Le gradient dans la couche de convolution est calculé comme suit :

$$\frac{\partial L}{\partial \omega_{a,b}} = \sum_{i=1}^{N-m-1} y_{(i+a)}^{l-1} \frac{\partial L}{\partial y_i^l} \sigma'(x_i') \qquad (8)$$

Pour le gradient dans la couche max-pooling, comme indiqué précédemment, le pool maximum réduit simplement la taille de la sortie convolutionnelle en introduisant l'éparpillement. Dans l'étape de propagation avant, la taille k de la fenêtre est réduite à une seule valeur. Cette valeur contribue à une erreur de propagation vers l'arrière à partir de la couche précédente, couche convolutionnelle.

Pour le cas des activités physiques basées sur le contexte, Il s'agit ici de prendre en compte l'ensemble des données des capteurs des trois objets connectés. La **figure 5**, présente l'architecture retenue pour la reconnaissance de l'activité physique basée sur le contexte.

L'apprentissage s'appuie sur le modèle présenté précédemment, avec les données de l'accéléromètre, gyroscope et magnétomètre. Afin d'entrainer ce modèle pour la reconnaissance des activités physiques basées sur le contexte, nous rajoutons les autres sources de données, redimensionnées et en l'intégrant directement dans la couche de convolution (**Figure 5).** En effet, l'état de l'art a montré que cette approche permet de mieux

optimiser le procéssus de traitement [117, 135,141]. Ainsi, en utilisant des modèles permettant de résoudre des taches similaires, en loccurence la reconnaissance d'activités élémentaires, la reconnaissance d'activités basées sur le contexte s'appuie directement sur l'archtecture existante en ajoutant les données des autres capteurs directement à la couche de convolution.

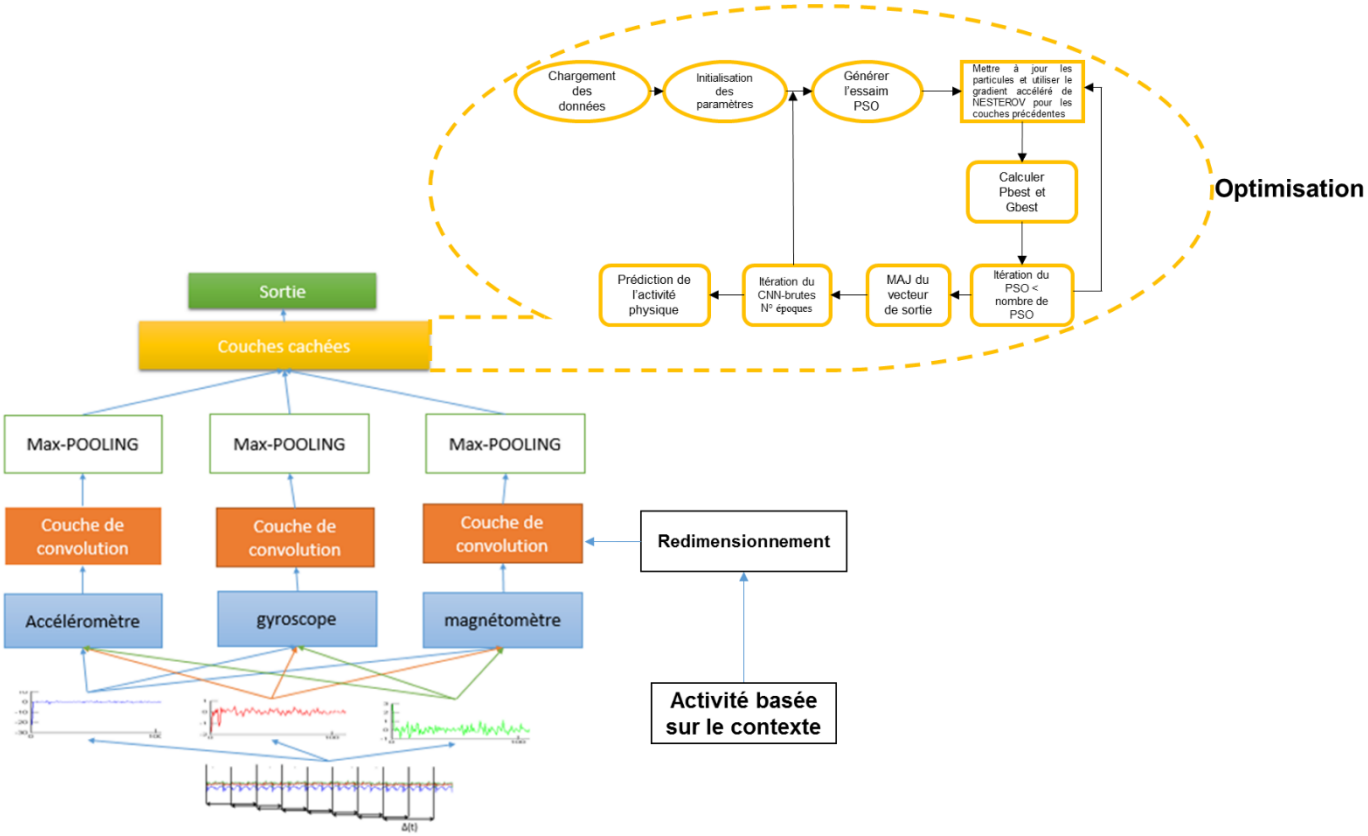


Figure 5: Schéma globale de l'architecture du modèle d'apprentissage CNN-brutes prenant en compte toutes les données des trois objets connectés. Une fois le modèle de reconnaissance des activités élémentaires entrainé, ses paramètres seront fixés. Ce modèle sera résutilisé pour l'apprentissage des données de l'activité basée sur le contexte en ajoutant les données de l'ensemble des capteurs directement dans les couches intermédiaires de convolution. Cette approche permet d'accélérer le processus d'apprentissage des réseaux de neurones profonds, que nous allons détailler dans la partie suivante. La composante « Optimisation », sur cette figure, concerne la contribution apportée dans ce chapitre en matière d'optimisation du processus d'entrainement. Ce sera traité dans la section 3.

2.2.2. Résultats et discussion

Le tableau 3 présente les résultats de classification des activités physiques élémentaires. Nous constatons que les précisions de reconnaissance ont considérablement progressés pour passer de 92.57% dans le cas du DNN-DCT à plus de 97% en utilisant ce modèle et sans sélection de descripteurs.

Tableau 2 : Résultats de classification de l'activité physique élémentaire avec un modèle de CNN.

	Debout	Assis	Allongé	marcher
CNN-brutes	98.12%	98.14%	97.1%	98.02%
DNN-DCT	91,01%	91,87%	92,33%	91,10%

Afin d'approfondir cette étude, nous proposons de comparer les deux approches en terme de vitesse de converge. La Figure 6 présente l'évolution du temps d'apprentissage des deux modèles DNN-DCT et DNN-brutes.

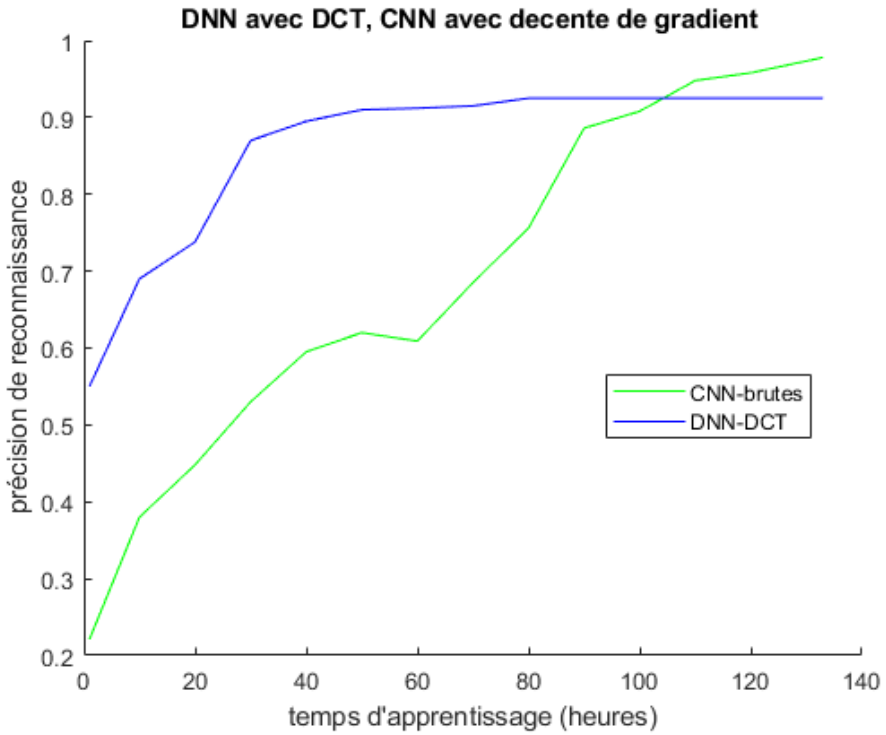


Figure 6 : Evolution du temps d'apprentissage des deux modèles.

Le CNN-brutes est plus précis. En revanche, celui-ci nécessite beaucoup plus de temps d'apprentissage par rapport au modèle DNN-DCT. A contrario, le DNN-DCT est beaucoup moins précis alors qu'il est plus rapide.

Les résultats de la reconnaissance des activités physiques basées sur le contexte, utilisant la totalité des données des capteurs embarqués dans les trois objets connectés, sont donnés dans le graphique suivant (**Figure 6**) :

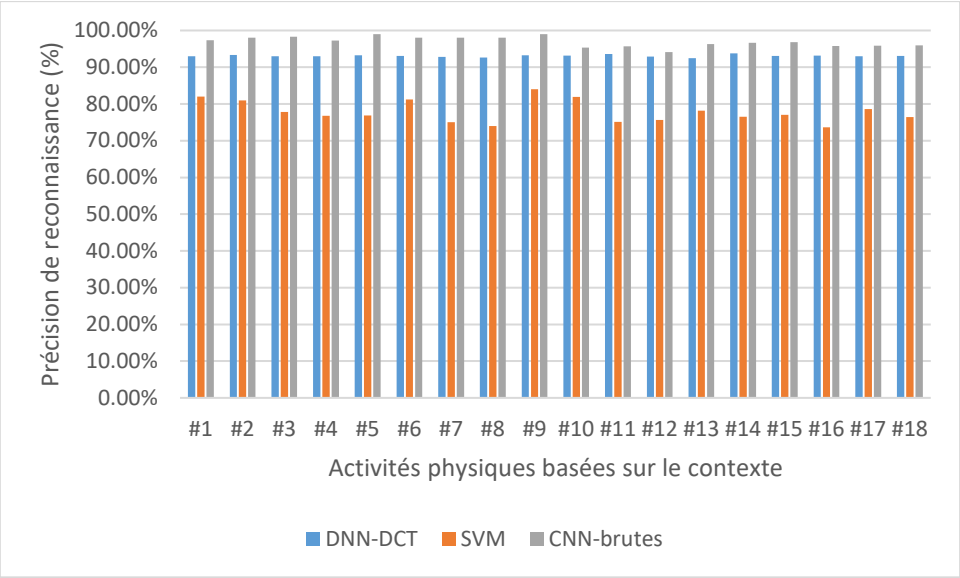


Figure 7 : Distribution de la reconnaissance des activités physiques basées sur le contexte en utilisant les modèles DNN-DCT vs SVM vs CNN-brutes.

Nous constatons que les précisions de reconnaissances sont bien meilleures en appliquant le CNN-brutes. La précision moyenne atteinte est de 96.8%, en moyenne. L'apprentissage des données brutes permet au CNN d'extraire des caractéristiques bien plus marquantes et plus informatives que dans le cas d'une extraction manuelle des descripteurs. Nous avons enregistré les échantillons du temps d'exécution par rapport à la performance de reconnaissance. **La figure 8** représente les résultats de cette étude.

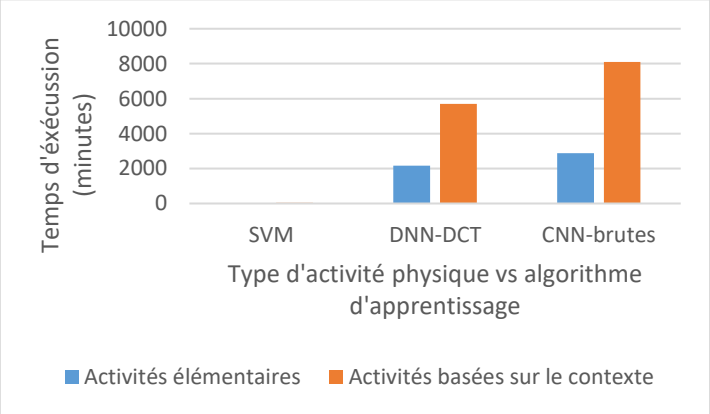


Figure 8 : Evolution du temps d'exécution en fonction du type d'activité détectée et l'algorithme d'apprentissage appliqué.

Les architectures des algorithmes DNN-DCT et DNN-brutes sont les plus performants par rapport à d'autres algorithmes d'apprentissage comme SVM. Touefois, ils restent très lents en termes de temps d'entrainement.

La sélection de descripteurs pour la reconnaissance de l'activité permet d'avoir des résultats moins précis en comparaison avec un modèle qui prend en compte des données brutes de l'activité physique. Ce dernier a l'inconvénient d'être très lent en terme de temps d'apprentissage.

Cette contrainte représente ainsi un dilemme entre la précision de reconnaissance qui devrait être assez élevée et le temps d'apprentissage qui doit rester raisonnable. Il serait donc intéressant d'optimiser l'apprentissage du CNN-Brutes afin de garder une précision élevée avec un temps d'apprentissage plus court.

En termes de précision, le modèle d'apprentissage des données brutes de l'activité humaine reste le plus intéressant, puis qu'il ne requiert pas de séléction manuelle de desctripteurs. C'est donc ce modèle que nous allons optimiser.

3. Etude de l'optimisation de l'entrainement du modèle CNN-brutes

L'état de l'art a montré que l'optimisation de réseaux de neurones profonds nécéssite de traiter le problème de disparition et/ou d'explosion de gradient ainsi que la lenteur du procéssus d'entrainement. Ceci passe par l'élaboration d'une stratégie optimale afin de i) choisir la bonne fonction d'activation du réseau de neurones et ii) choisir des optimiseurs rapides.

Afin de répondre à ces deux critères, nous avons mené une étude exploratoire afin d'implémenter les fonctions d'activations et optimiseurs les plus répandus dans l'état de l'art. Le détail de cette étude est résumée dans l'annexe E. Le tableau 3 représente les meilleurs paramètres retenus de cette étude.

Tableau 3 : combinaison optimale d'apprentissage d'un réseau de neurones profonds.

Initialisation	Initialisation de He
Fonction d'activation	SELU
Normalisation	Aucune (grâce à SELU)
Régularisation	Alpha dropout
Optimiseur	Gradient accéléré de Nesterov
Echéancier d'apprentissage	Planification par exponentielle

En utilisant cette configuration, nous arrivons à obtenir un résultat de classification de l'ordre de 96% et 95.4% des activités élémentaires et basées sur le contexte respectivement. Avec un gain en temps d'apprentissage de 6h32 et 5h40 pour les activités élémentaires et basées sur le contexte, respectivement.

Il est clair que les résultats d'optimisation du temps d'entrainement sont plus interesants que les résultats de précision de classification. Ces derniers ont été revus à la baisse suite à l'application de cette configuration (voir tableau 3), 96% contre 97.84% (voir tableau 2), pour la classification des activités élémentaires, et 95.4% contre 98% en moyenne pour les activités basées sur le contexte.

Afin d'améliorer les précisions de classifications tout en gardant les résultats d'optimisation du temps d'entrainement acquis, nous proposons d'explorer une nouvelle approche, appelée optimisation par essaim particulaire ou PSO, permettant d'améliorer les résultats de classification en agissant directement dans la couche de décision.

3.1. L'architecture proposée

L'une des particularités de notre modèle CNN-brutes est que la dernière couche responsable de la classification est complétement connectée avec les couches précédentes. L'état de l'art a montré que dans des cas pratiques, une petite zone, et donc une petite quantité de neurones, soit nécessaire à atteindre le résultat de classification souhaité, en loccurence 98%.Il serait donc intéressant de tenter d'optimiser l'algotithme du CNN-brutes à ce niveau.

Étant donné que la dernière couche est responsable du résultat de la classification, elle est très sensible aux poids de connexion avec la couche précédente pour déterminer le niveau d'appartenance de la caractéristique de l'activité courante : activité élémentaire et basée sur le contexte. Afin d'améliorer la précision de la classification, nous avons étudié l'optimisation du CNN-brutes, plus exactement, du dernier vecteur de poids en utilisant l'algorithme PSO. Nous appelons ce modèle CNN-brutes-PSO (**Figure 9**).

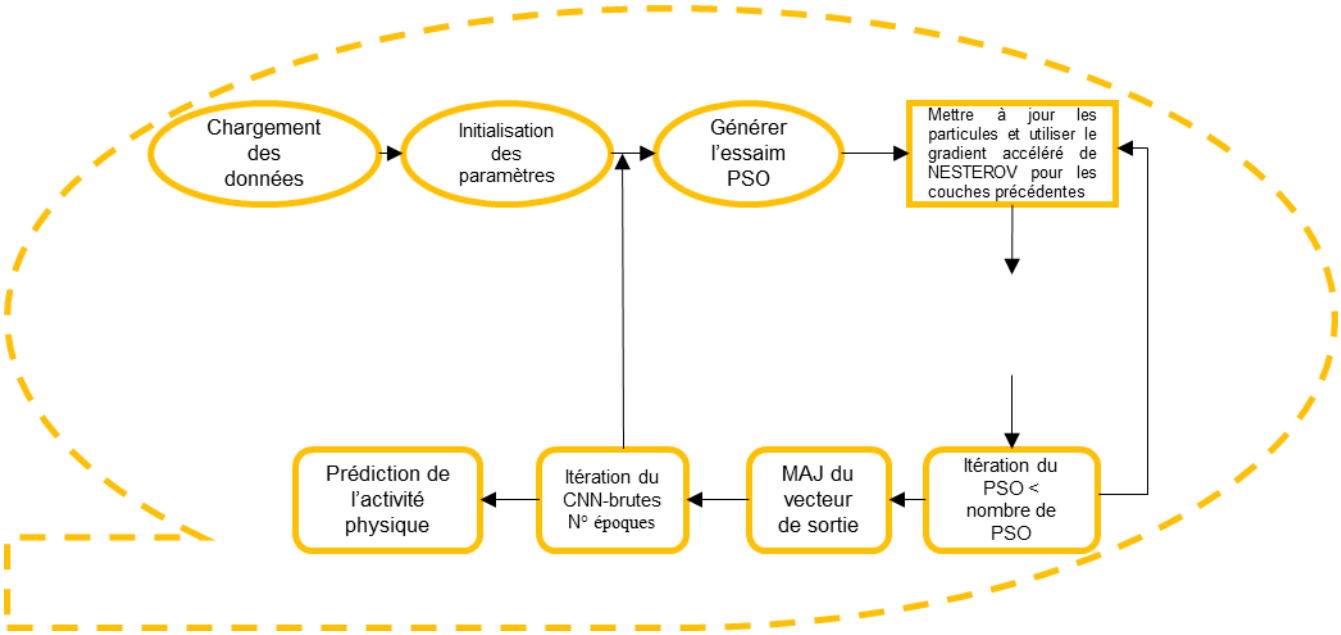


Figure 9 : Etapes de l'implémentation du PSO pour notre algorithme CNN-brutes-PSO.

Afin d'entrainé le modèle, nous proposons de suivre la procédéure suivante : Après avoir chargé l'ensemble de données, nous avons commencé par initialiser certains paramètres du modèle comme la taille du lot, le nombre d'époques, le nombre d'itérations du PSO et la taille de l'essaim du PSO. L'objectif d'entrainement du CNN-brutes-PSO est de trouver les meilleures valeurs de chaque couche et établir une bonne relation entre les couches du CNN-brutes-PSO pour assurer la bonne identification des caracteristiques.

Dans le modèle standard, l'algorithme de la décente de Gradient accéléré de Nesterov, utilisé pour entrainer le modèle CNN-brutes, est utilisé pour optimiser tous les paramètres du modèle, y compris les filtres de convolution et les deux poids des couches entièrement connectés.

La première étape de l'algorithme PSO est de générer aléatoirement les particules de l'essaim, y compris les positions et les vitesses. Après l'évaluation de chaque particule, nous déterminons les meilleures positions locales (Pbest) et la meilleure position globale dans l'essaim (Gbest). Pendant l'apprentissage, les particules sont mises à jour en utilisant les équations (1) et (2) suivantes :

$$v_{n+1} = v_n + c_1 r_1 (Pbest - x_n) + c_2 r_2 (Gbest - x_n) \quad (1)$$

$$x_{n+1} = x_n + v_{n+1} \quad (2)$$

A la fin des itérations du PSO, le vecteur de poids de sortie est mis à jour en fonction de la meilleure particule globale finale de l'essaim. Ensuite, le modèle entrainé est testé par l'ensemble de données de test pour calculer le taux de classification. L'organigramme du modèle proposé est représenté sur la **figure 9** ci-dessus. Les résultats de l'entrainement du modèle CNN-beutes-PSO sont renseignés dans **le tableau 4**.

3.2. Résultats et discussion

Dans l'étude expérimentale, nous définissons la taille du lot à 50, l'itération PSO à 10, les constantes PSO c1 = c2 = 0,2, la taille de l'essaim PSO à 10 et le nombre d'époques d'entraînement dans la plage de 10 à 100.

Lorsque *x* est égal à 100, nous constatons que notre modèle atteint de meilleurs résultats car la fonction de perte est égale à 0,02 lorsque le CNN-brutes standard atteint environ 0,44. L'utilisation de l'algorithme PSO permet une grande diminution de la fonction de perte qui résulte d'un meilleur taux de précision pendant les différentes époques.

Tableau 4 : Précision de classification des modèles CNN-brutes-PSO et CNN-brutes.

Nombre d'époques	10	20	30	40	50	60	70	80	90	100
CNN-brutes	85.01%	86.3%	87.4%	83.6%	96.7%	95.1%	96.7%	96.9%	97.3%	97.8%
CNN-brutes-PSO	91.3%	93.6%	94.1%	95.4%	96.1%	97.2%	97.8%	98%	98.2%	99.1%

Comme présenté dans le **tableau 4**, la performance d'exactitude atteinte est 99,1% en 100 époques. ce qui signifie que le modèle proposé est un modèle efficace pour la classification des activités physiques basées sur le contexte.

Notre approche combine deux algorithmes d'optimisation différents afin d'optimiser le résultat de performance de précision de l'architecture CNN-brutes initiale. Il utilise l'architecture CNN-brutes déjà représentée comme utilisant l'algorithme PSO pour la dernière couche qui est le vecteur de sortie et utilise l'algorithme de Gradient accéléré de Nesterov pour les premières couches. Notre résultat montre une amélioration du résultat de précision par rapport à l'architecture standard du CNN-brutes qui utilise uniquement l'algorithme d'optimisation du Gradient accéléré de Nesterov (après réglages) dans toutes les couches.

Afin de voir son impact sur la vitesse d'apprentissage, nous représentons dans la **figure 10**, les graphiques montrants le temps d'apprentissage des différents modèles : CNN-brutes, DNN-DCT et CNN-brutes-PSO.

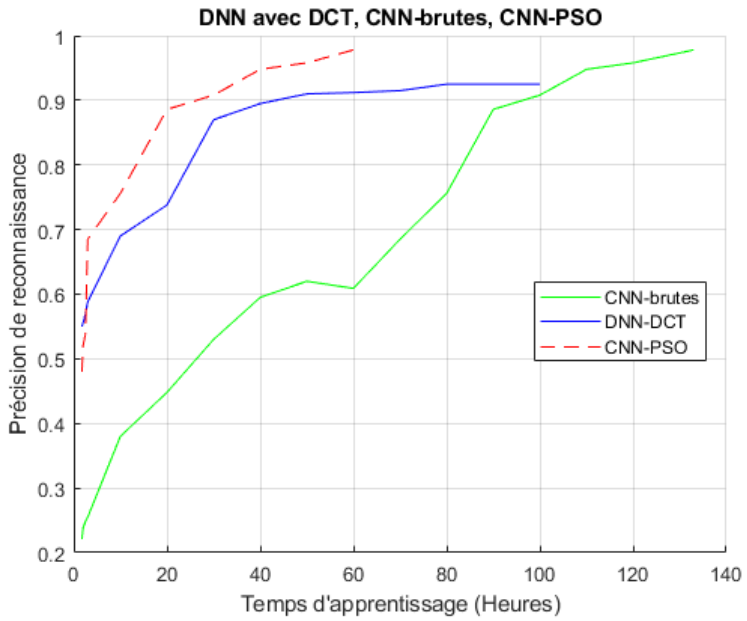


Figure 10 : Comparaison entre les trois modèles en termes de vitesse de convergence et précision de reconnaissance.

Notre modèle CNN-brutes-PSO dépasse les deux autres modèles pour la classification des activités physiques. Il est plus précis et plus rapide. Il permet de réduire le temps d'entrainement et d'améliorer la précision de classification. De plus, l'approche proposée est applicable non seulement dans le cas des activités élémentaires mais aussi pour les activités basées sur le contexte.

L'apprentissage du CNN-brutes avec l'algorithme du PSO a permis d'avoir des résultats meilleurs que ceux de l'état de l'art sur la reconnaissance de l'activité. A titre d'exemple, nous pouvons citer les travaux de Masaya et al [130] qui ont appris un modèle basé sur l'apprentissage en profondeur récurrent pour détecter 6 types d'activités de 7 personnes différentes avec une précision de 95%, dans un environnement contrôlé. L'implémentation de leur approche sur nos données d'activités basées sur le contexte a donné une précision de 94.89% en moyenne, avec un temps de calcul égal à 5.7 fois le temps de calcul de notre modèle CNN-PSO brutes.

En résumé, nous proposons de mettre les principaux résultats de cette étude dans le tableau 5 suivant :

Tableau 5 : Synthèse des résultats de l'optimisation de l'architecture du CNN-brutes.

	Methode d'entrainement/optimisation	Le gain en temps	Précision de classification
CNN-brutes	Gradient accéléré de Nesterov	5h40	95.4%
CNN-brutes	Décente de grandient	0	96.8%
CNN-brutes-PSO	Gradient accéléré de Nesterov + PSO	9h30	99.1%

Comme le montre le tableau 5, notre approche dépasse les approches classiques utilisées au début de notre étude. Ainsi, la précision atteinte dépasse le modèle initial : 99.1% de précision avec un temps d'entrainement qui s'est nettement amélioré pour passer de 5h40 à presque 10 heures.

Conclusion

Dans ce chapitre, nous avons étudié quelques facteurs qui influencent l'optimisation des réseaux de neurones profonds pour la reconnaissance de l'activité physique.

Nous avons montré que la sélection de descripteurs permet d'entrainer des modèles rapidement. En revanche, l'utilisation des données brutes pour l'entrainement du modèle d'apprentissage en profondeur permet d'avoir une meilleure précision de reconnaissance. Toutefois, le processus d'apprentissage est plus lent dans ce cas.

Notre étude comparative des différents paramètres d'entrainement d'un réseau de neurones profond a permis de choisir la meilleure fonction d'activation. Celle-ci s'avère efficace pour entrainer rapidement les réseaux de neurones profonds. Nous avons également exploré des

optimiseurs plus rapides, comme la méthode Adam, le Gradient accéléré de Nesterov et la Descente de Gradient Stochastique. Ces méthodes restent moins précises. Ceci peut s'expliquer par le fait que ce type de méthodes requiert beaucoup d'ajustements manuels (hyperparamètres).

Afin de répondre à ces limitations, une heuristique appelée optimisation par essaim particulaires (PSO) a été explorée. Celle-ci a offert des résultats encourageants en termes de précision et de temps de calcul. En effet, l'utilisation de l'algorithme du PSO a permis d'opérer directement sur la dernière couche permettant la classification. Enfin, la précision de reconnaissance ainsi que la vitesse d'apprentissage ont été nettement améliorés.

CONCLUSION GENERALE

Rappel des objectifs de recherche

Dans cette thèse, nous nous sommes intéressés à la reconnaissance de l'activité physique en utilisant des objets connectés portables dans un environnement non contrôlé.

Ceci est rendu possible grace aux nouvelles technologies de l'internet des objets qui permettent de communiquer les signaux de l'activité physique enrichis d'un certain contexte (signaux audio, données des surfaces tactiles, des données des boutons des télécommandes ou smartphones, etc.), ce qui pose le problème de l'exploration de cette multitude de type de données. Comment pourrons-nous les explorer efficacement afin de reconnaitre les activités physiques ?

La conception d'un système de reconnaissance de l'activité physique est réalisée de deux façons différentes. La première, la plus répandue, consiste à extraite des descripteurs les plus informatifs à partir de différentes sources de données collectées, les concaténer et les donner en entrées d'algorithmes d'apprentissage automatique (SVM, MLP, DNN, etc.). La deuxième approche consiste à développer un modèle de fusion prédictif permettant d'exploiter les données des descripteurs ainsi extraits et les fusionner. La sortie de ce type de modèles est une prédiction d'une activité physique. Les deux approches nécessitent une étape d'extraction de descripteurs, ce pose alors la question du choix du type de descripteurs. Quels sont les meilleurs descripteurs à extraire ?

Le calcul et la sélection de descripteurs est un travail fastidieux. En effet, il faudrait non seulement choisir le bon descripteur à utiliser dans l'étape de classification par des algorithmes d'apprentissage automatique, mais aussi y apporter les réglages nécessaires en terme de taille, dimension du descripteur, etc. Comment explorer efficacement les données des différentes sources des objets connectés sans avoir à calculer et sélectionner des descripteurs ?

Les données extraites des objets connectés sont, par définition, imparfaites. Les différents canaux de communication ainsi que les protocoles d'exploration et de traitement sont en générales sujettes à des erreurs (de transmission, d'incertitude, etc.). Comment pourrons-nous explorer efficacement cette nouvelle contrainte d'incertitude et d'imperfection afin de reconnaitre les activités physiques ?

Contributions de la thèse

Nos travaux de thèse avaient pour objectif de répondre à ces problématiques. Une étude bibliographique nous a permis de mettre en évidence la pertinence de l'utilisation des techniques d'apprentissage automatique pour la reconnaissance de l'activité physique. Celles-ci requierts de suivre les étapes suivantes : i) collecte de données, ii) extraction de descripteurs et iii) apprentissage.

Nous résumons, dans ce qui suit, les principales contributions de cette thèse.

1) Création d'un nouveau corpus de données d'activités

Nous avons commencé par proposer un corpus de données d'activités physiques de 33 participants. La durée des enregistrements était de 6 mois. Les utilisateurs portaient trois objets connectés : un smartphone, une smartwatch et une smart TV. Nous avons extrait toutes les données de l'ensemble des capteurs des trois objets connectés. L'experience s'est déroulée dans un environnement non controlé ; les participants ne recevaient aucunes instructions relatives à l'exécussion de leurs activités.

2) Utilisation de la DCT comme principal descripteur

Afin de choisir le meilleur descripteur, nous avons commencé par utiliser les trois capteurs (accéléromètres, gyroscopes et magnétomètres) des trois objets connectés, afin de détcter quatre types d'activités : debout, assis, allongé et marcher. Nous avons appelé ce type d'activité, les activités physiques élémentaires. Après avoir extrait les principaux descripteurs les plus utilisés dans l'état de l'art, une étude comparative a montré la pertinence de choisir la transformée en cosinus discrète (DCT).

3) Gestion de l'imperfection des données capteurs

Lors de mon séjour à l'université de North Dakota aux USA, j'ai pu mettre en évidence l'importance de prendre en compte l'imperfection des données capteurs.

L'étude de l'état de l'art nous a permis d'identifier un cadre théorique adapté et nouveau dans le domaine de la reconnaissance de l'activité physique avec des objets connectés : la théorie de Dempster-Shafer. Nous avons examiné les différents critères d'implémentation de cette théorie pour la reconnaissance de l'activité physique. En particulier, le calcul des fonctions de masse. A ce titre, nous avons proposé deux approches, une première approche consistant à subdiviser l'espace des descripteurs en cellules. La deuxième approche consiste à utiliser des classifieurs. Les résultats ont montré que la méthode des classifieurs était la plus précise et adapée.

Avec l'augmentation du nombre d'activités à reconnaitre, notre étude a montré que le calcul et la fusion des fonctions de masse devient très lent et complexe. Nous avons proposé une approche basée sur l'approximation des fonctions de masse et l'utilisation des fonctions de corrélation, afin de répondre à cette limitation. De plus, la fusion des fonctions de masse approximées permet d'atteindre des résultats meilleurs que ceux de l'état de l'art. Notre approche est plus précise. L'approximation des fonctions de masse conserve une majeure partie de l'information apportée par les masses d'origine. Ainsi, notre approche nous permet d'éviter une perte d'information.

4) Les réseaux de neurones profonds pour la reconnaissance de l'activité

Lors de mon séjour à l'université de Colombie Britannique (British Columbia University), j'ai pu mettre en avant l'interet de l'exploration des méthodes d'apprentissage profond et les techniques d'optimisation pour la reconnaissance de l'activité.

En effet, afin de répondre à la problématique du calcul et sélection de descripteurs, nous avons exploré une approche d'apprentissage en profondeur sans sélection de descripteurs (appelée CNN-brutes). Nous avons mené une étude comparative entre le modèle ainsi développé et un modèle d'apprentissage profond avec sélection de descripteur (appelé DNN-DCT). Les résultats obtenus avec le CNN-brutes étaient meileurs en termes de précision de classification.

L'état de l'art a montré que les modèles d'apprentissage profonds utilisés sont généralement optimisés avec une simple décente de gradient (stochastique). Ceci peut être source de plusieurs problèmes comme la disparition/explosion de gradients, la lenteur du processus d'entrainement. Enfin, en présente d'un grand nombre d'activités physiques à détecter, les réseaux de neurones deviennent très profonds (des millions de paramètres, hyperparamètres, etc). Afin de répondre à cette problématique, nous avons mené une étude exploratoire afin d'aboutir aux meilleurs paramétrages (fonctions d'activation, optimiseurs, etc.). Les résultats ont montré que notre modèle souffrait de lenteur dans le processus d'entrainement. La tentative de réduire le temps d'entrainement avec les optimiseurs classiques a été faite au détriment de la précision de classification. Afin d'améliorer ce résultat, nous avons proposé une méthode permettant d'optimiser la précision de classification tout en optimisant le temps d'entrainement. Cette méthode consiste à coupler l'optimisation par gradient accéléré de Nesterov avec l'algorithme d'optimisation de l'Essaim Particulaire (PSO). Les résultats étaient meiileurs que ceux de l'état de l'art.

Limites de cette thèse et perspectives de recherche

Des éléments de réponse on put être apportés aux questions de recherche émises au début de cette thèse. Nos contributions soulèvent de nouvelles interrogations.

L'utilisation de la théorie de Dempster-Shafer dans sa forme actuelle ne prend pas en compte les valeurs mesurées par les capteurs à chaque instant t pour produire une inférence sur l'état de la personne. Dans des conditions expérimentales favorables et stationnaires, qui ne changent pas au cours du temps pendant les expérimentations, notre modèle présente de bonnes performances, comme celles présentées dans le Chapitre 4. Cependant, la non-stationnarité des signaux recueillis sur les capteurs du système considéré peut conduire à une dégradation des conditions expérimentales. La présence de bruit de mesure, la variabilité des signaux enregistrés par les capteurs, les capteurs défaillants ou non fiables, peuvent rendre notre modèle incohérent dans sa décision. Plusieurs travaux ont tenté d'apporter des éléments de réponses. Les travaux existants ont utilisé des réseaux d'évidence dynamiques. Ces modèles peuvent prendre en compte la dynamique temporelle de l'expérience (données supplémentaires, décision en ligne, etc.). En revanche, une étape importante de séléction de descripteurs doit etre faite. Il serait donc intéréssant d'explorer des techniques prenant en compte ces différentes contraintes afin de rendre les modèles basées sur la théorie de Demspter-Shafer plus précis, plus robustes et auto adaptatifs.

L'optimisation de l'entrainement par PSO a apporté des contributions en termes de vitesse d'entrainement et précision de classification. En revanche, le modèle appris nécessite l'utilisation d'une méthode d'optimisation avant d'appliquer l'algorithme du PSO. Ceci peut être sources de plusieurs problèmes (la précision de classification atteinte n'est pas maximale, etc) notamment si le modèle à mettre en œuvre nécessite d'être plus profond. Afin d'éviter d'inclure des modèles d'optimisation basées sur le calcul de gradients, il est nécéssaire d'étudier la propagation du PSO en utilisant différents types d'architectures du CNN-brutes et éviter ainsi de l'appliquer à la dernière couche du modèle seulement. Un travail intéressant dans le domaine du traitement d'images a été proposé dans l'état de l'art permettant d'entrainer un réseau de neurones sans utilisation de Gradients.

RÉFÉRENCES BIBLIOGRAPHIQUES

[1] R. M. Palmer, A. G. Ferrige, et S. Moncada, « Nitric oxide release accounts for the biological activity of endothelium-derived relaxing factor », Nature, vol. 327, no 6122, p. 524, 1987.

[2] O. D. Lara et M. A. Labrador, « A survey on human activity recognition using wearable sensors. », IEEE Communications Surveys and Tutorials, vol. 15, no 3, p. 1192–1209, 2013.

[3] S. C. Mukhopadhyay, « Wearable sensors for human activity monitoring: A review », IEEE sensors journal, vol. 15, no 3, p. 1321–1330, 2015.

[4] D. Preuveneers et Y. Berbers, « Mobile phones assisting with health self-care: a diabetes case study », in Proceedings of the 10th international conference on Human computer interaction with mobile devices and services, 2008, p. 177–186.

[5] E. Kim, S. Helal, et D. Cook, « Human activity recognition and pattern discovery », IEEE Pervasive Computing, vol. 9, no 1, 2010.

[6] M. Tambe et al., « Intelligent agents for interactive simulation environments », AI magazine, vol. 16, no 1, p. 15, 1995.

[7] C. E. Zsambok et G. Klein, Naturalistic decision making. Psychology Press, 2014.

[8] J. K. Aggarwal et Q. Cai, « Human motion analysis: A review », Computer vision and image understanding, vol. 73, no 3, p. 428–440, 1999.

[9] G. M. Weiss, J. L. Timko, C. M. Gallagher, K. Yoneda, et A. J. Schreiber, « Smartwatch-based activity recognition: A machine learning approach », in Biomedical and Health Informatics (BHI), 2016 IEEE-EMBS International Conference on, 2016, p. 426–429.

[10] T. V. Duong, H. H. Bui, D. Q. Phung, et S. Venkatesh, « Activity recognition and abnormality detection with the switching hidden semi-markov model », in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, 2005, vol. 1, p. 838–845.

[11] L. Atzori, A. Iera, et G. Morabito, « The internet of things: A survey », Computer networks, vol. 54, no 15, p. 2787–2805, 2010.

[12] G. Kortuem, F. Kawsar, V. Sundramoorthy, et D. Fitton, « Smart objects as building blocks for the internet of things », IEEE Internet Computing, vol. 14, no 1, p. 44–51, 2010.

[13] J. Gubbi, R. Buyya, S. Marusic, et M. Palaniswami, « Internet of Things (IoT): A vision, architectural elements, and future directions », Future generation computer systems, vol. 29, no 7, p. 1645–1660, 2013.

[14] N. Koshizuka et K. Sakamura, « Ubiquitous ID: standards for ubiquitous computing and the Internet of Things », IEEE Pervasive Computing, no 4, p. 98–101, 2010.

[15] L. Tan et N. Wang, « Future internet: The internet of things », in Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on, 2010, vol. 5, p. V5–376.

[16] C. Perera, A. Zaslavsky, P. Christen, et D. Georgakopoulos, « Context aware computing for the internet of things: A survey », IEEE communications surveys & tutorials, vol. 16, no 1, p. 414–454, 2014.

[17] M. Aazam, I. Khan, A. A. Alsaffar, et E.-N. Huh, « Cloud of Things: Integrating Internet of Things and cloud computing and the issues involved », in Applied Sciences and Technology (IBCAST), 2014 11th International Bhurban Conference on, 2014, p. 414–419.

[18] T. O'Reilly et J. Battelle, Web squared: Web 2.0 five years on. O'Reilly Media, Inc., 2009.
[19] A. Serbanati et al., « Internet of Things Architecture », Concept and Solutions for Privacy and Security in the Resolution Infrastructure, EU Project IoT-A, Project Report D, vol. 4, 2013.
[20] B. David et al., « SMART-CITY: Problematics, techniques and case studies », in Computing Technology and Information Management (ICCM), 2012 8th International Conference on, 2012, vol. 1, p. 168–174.
[21] I. Chiuchisan, H.-N. Costin, et O. Geman, « Adopting the internet of things technologies in health care systems », in Electrical and Power Engineering (EPE), 2014 International Conference and Exposition on, 2014, p. 532–535.
[22] M. Chen, Y. Miao, Y. Hao, et K. Hwang, « Narrow band internet of things », IEEE Access, vol. 5, p. 20557–20577, 2017.
[23] N. Gershenfeld, R. Krikorian, et D. Cohen, « The internet of things », Scientific American, vol. 291, no 4, p. 76–81, 2004.
[24] D. Miorandi, S. Sicari, F. De Pellegrini, et I. Chlamtac, « Internet of things: Vision, applications and research challenges », Ad hoc networks, vol. 10, no 7, p. 1497–1516, 2012.
[25] E. Belove et al., « Distributed, intermittently connected, object-oriented database and management system », févr-1996.
[26] R. Khan, S. U. Khan, R. Zaheer, et S. Khan, « Future internet: the internet of things architecture, possible applications and key challenges », in Frontiers of Information Technology (FIT), 2012 10th International Conference on, 2012, p. 257–260.
[27] L. Da Xu, W. He, et S. Li, « Internet of things in industries: A survey », IEEE Transactions on industrial informatics, vol. 10, no 4, p. 2233–2243, 2014.
[28] K. L. Levy et G. B. Rhoads, « Connected audio and other media objects », janv-2003.
[29] V. Baumgarte, G. Ehlers, F. May, A. Nückel, M. Vorbach, et M. Weinhardt, « PACT XPP—A self-reconfigurable data processing architecture », the Journal of Supercomputing, vol. 26, no 2, p. 167–184, 2003.
[30] M. Shapiro, « Structure and encapsulation in distributed systems: the proxy principle », in Int. Conf. on Distr. Comp. Sys.(ICDCS), 1986, p. 198–204.
[31] A. Amirat et M. Oussalah, « First-class connectors to support systematic construction of hierarchical software architecture », The Journal of Object Technology, vol. 8, no 7, p. 107–130, 2009.
[32] R. Seller et A. Szullo, « Wide area multilateration demonstration system », in Applied Electromagnetics and Communications (ICECom), 2013 21st International Conference on, 2013, p. 1–5.
[33] J. Takacs, C. L. Pollock, J. R. Guenther, M. Bahar, C. Napier, et M. A. Hunt, « Validation of the Fitbit One activity monitor device during treadmill walking », Journal of Science and Medicine in Sport, vol. 17, no 5, p. 496–500, 2014.
[34] A. Pintus, D. Carboni, et A. Piras, « Paraimpu: a platform for a social web of things », in Proceedings of the 21st International Conference on World Wide Web, 2012, p. 401–404.
[35] A. Stisen et al., « Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition », in Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems, 2015, p. 127–140.
[36] A. Bulling, U. Blanke, et B. Schiele, « A tutorial on human activity recognition using body-worn inertial sensors », ACM Computing Surveys (CSUR), vol. 46, no 3, p. 33, 2014.
[37] J. Meyer, P. Lukowicz, et G. Troster, « Textile pressure sensor for muscle activity and

motion detection », in Wearable Computers, 2006 10th IEEE International Symposium on, 2006, p. 69–72.

[38] P. Kahn, A. Kinsolving, M. A. Christensen, B. Y. Lee, et D. Vogel, « Eyewear having human activity monitoring device », juill-2011.

[39] M. Bobin, H. Amroun, S. Coquillart, F. Bimbard, et M. Ammi, « SVM based approach for the assessment of elbow flexion with smart textile sensor », in Systems, Man, and Cybernetics (SMC), 2017 IEEE International Conference on, 2017, p. 2129–2134.

[40] C. Schuldt, I. Laptev, et B. Caputo, « Recognizing human actions: a local SVM approach », in Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, 2004, vol. 3, p. 32–36.

[41] H. Wang, M. M. Ullah, A. Klaser, I. Laptev, et C. Schmid, « Evaluation of local spatio-temporal features for action recognition », in BMVC 2009-British Machine Vision Conference, 2009, p. 124–1.

[42] A. F. Bobick et J. W. Davis, « The recognition of human movement using temporal templates », IEEE Transactions on pattern analysis and machine intelligence, vol. 23, no 3, p. 257–267, 2001.

[43] I. Laptev, M. Marszalek, C. Schmid, et B. Rozenfeld, « Learning realistic human actions from movies », in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, p. 1–8.

[44] J. K. Aggarwal et M. S. Ryoo, « Human activity analysis: A review », ACM Computing Surveys (CSUR), vol. 43, no 3, p. 16, 2011.

[45] R. Poppe, « A survey on vision-based human action recognition », Image and vision computing, vol. 28, no 6, p. 976–990, 2010.

[46] P. Turaga, R. Chellappa, V. S. Subrahmanian, et O. Udrea, « Machine recognition of human activities: A survey », IEEE Transactions on Circuits and Systems for Video technology, vol. 18, no 11, p. 1473–1488, 2008.

[47] P. Dollár, V. Rabaud, G. Cottrell, et S. Belongie, « Behavior recognition via sparse spatio-temporal features », in Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 2nd Joint IEEE International Workshop on, 2005, p. 65–72.

[48] A. Shahroudy, J. Liu, T.-T. Ng, et G. Wang, « NTU RGB+ D: A large scale dataset for 3D human activity analysis », in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, p. 1010–1019.

[49] J. Yamato, J. Ohya, et K. Ishii, « Recognizing human action in time-sequential images using hidden markov model », in Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on, 1992, p. 379–385.

[50] J.-D. Haynes et G. Rees, « Neuroimaging: decoding mental states from brain activity in humans », Nature Reviews Neuroscience, vol. 7, no 7, p. 523, 2006.

[51] F. ERCIS, « COMPARISON OF HISTOGRAMS OF ORIENTED OPTICAL FLOW BASED ACTION RECOGNITON METHODS », PhD Thesis, MS Thesis, Middle East Technical University, Turkey, 2012.

[52] R. Chavarriaga et al., « The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition », Pattern Recognition Letters, vol. 34, no 15, p. 2033–2042, 2013.

[53] A. M. Khan, Y.-K. Lee, S. Y. Lee, et T.-S. Kim, « Human activity recognition via an accelerometer-enabled-smartphone using kernel discriminant analysis », in Future Information Technology (FutureTech), 2010 5th International Conference on, 2010, p. 1–6.

[54] O. D. Lara et M. A. Labrador, « A survey on human activity recognition using wearable sensors. », IEEE Communications Surveys and Tutorials, vol. 15, no 3, p. 1192–1209, 2013.
[55] Z. Zeng, M. Pantic, G. I. Roisman, et T. S. Huang, « A survey of affect recognition methods: Audio, visual, and spontaneous expressions », IEEE transactions on pattern analysis and machine intelligence, vol. 31, no 1, p. 39–58, 2009.
[56] D. Anguita, A. Ghio, L. Oneto, X. Parra, et J. L. Reyes-Ortiz, « A public domain dataset for human activity recognition using smartphones. », in ESANN, 2013.
[57] O. D. Lara, A. J. Pérez, M. A. Labrador, et J. D. Posada, « Centinela: A human activity recognition system based on acceleration and vital sign data », Pervasive and mobile computing, vol. 8, no 5, p. 717–729, 2012.
[58] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, et P. Havinga, « Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey », in Architecture of computing systems (ARCS), 2010 23rd international conference on, 2010, p. 1–10.
[59] D. Anguita, A. Ghio, L. Oneto, X. Parra, et J. L. Reyes-Ortiz, « Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine », in International workshop on ambient assisted living, 2012, p. 216–223.
[60] K. Altun et B. Barshan, « Human activity recognition using inertial/magnetic sensor units », in International workshop on human behavior understanding, 2010, p. 38–51.
[61] Y. E. Ustev, O. Durmaz Incel, et C. Ersoy, « User, device and orientation independent human activity recognition on mobile phones: Challenges and a proposal », in Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication, 2013, p. 1427–1436.
[62] A. Anjum et M. U. Ilyas, « Activity recognition using smartphone sensors », in Consumer Communications and Networking Conference (CCNC), 2013 IEEE, 2013, p. 914–919.
[63] D. Riboni et C. Bettini, « Context-aware activity recognition through a combination of ontological and statistical reasoning », in International Conference on Ubiquitous Intelligence and Computing, 2009, p. 39–53.
[64] G. Masuyer, S. L. Schwager, E. D. Sturrock, R. E. Isaac, et K. R. Acharya, « Molecular recognition and regulation of human angiotensin-I converting enzyme (ACE) activity by natural inhibitory peptides », Scientific reports, vol. 2, p. 717, 2012.
[65] A. Bayat, M. Pomplun, et D. A. Tran, « A study on human activity recognition using accelerometer data from smartphones », Procedia Computer Science, vol. 34, p. 450–457, 2014.
[66] J.-L. Reyes-Ortiz, L. Oneto, A. Samà, X. Parra, et D. Anguita, « Transition-aware human activity recognition using smartphones », Neurocomputing, vol. 171, p. 754–767, 2016.
[67] H. Wang, A. Kläser, C. Schmid, et C.-L. Liu, « Action recognition by dense trajectories », in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, 2011, p. 3169–3176.
[68] H. Davies et al., « Mutations of the BRAF gene in human cancer », Nature, vol. 417, no 6892, p. 949, 2002.
[69] W. J. Storkus, H. d Zeh, M. J. Maeurer, R. D. Salter, et M. T. Lotze, « Identification of human melanoma peptides recognized by class I restricted tumor infiltrating T lymphocytes. », The Journal of Immunology, vol. 151, no 7, p. 3719–3727, 1993.
[70] N. A. Capela, E. D. Lemaire, et N. Baddour, « Feature selection for wearable smartphone-based human activity recognition with able bodied, elderly, and stroke patients », PloS one, vol. 10, no 4, p. e0124414, 2015.

[71] S. Dernbach, B. Das, N. C. Krishnan, B. L. Thomas, et D. J. Cook, « Simple and complex activity recognition through smart phones », in Intelligent Environments (IE), 2012 8th International Conference on, 2012, p. 214–221.
[72] F. Palumbo, C. Gallicchio, R. Pucci, et A. Micheli, « Human activity recognition using multisensor data fusion based on reservoir computing », Journal of Ambient Intelligence and Smart Environments, vol. 8, no 2, p. 87–107, 2016.
[73] E. Waltz et J. Llinas, Multisensor data fusion, vol. 685. Artech house Boston, 1990.
[74] D. L. Hall et J. Llinas, « An introduction to multisensor data fusion », Proceedings of the IEEE, vol. 85, no 1, p. 6–23, 1997.
[75] H. Medjahed, D. Istrate, J. Boudy, J.-L. Baldinger, et B. Dorizzi, « A pervasive multi-sensor data fusion for smart home healthcare monitoring », in Fuzzy Systems (FUZZ), 2011 IEEE International Conference on, 2011, p. 1466–1473.
[76] K. Hung, Y.-T. Zhang, et B. Tai, « Wearable medical devices for tele-home healthcare », in Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE, 2004, vol. 2, p. 5384–5387.
[77] D. L. Hall et S. A. McMullen, Mathematical techniques in multisensor data fusion. Artech House, 2004.
[78] E. Rahm et H. H. Do, « Data cleaning: Problems and current approaches », IEEE Data Eng. Bull., vol. 23, no 4, p. 3–13, 2000.
[79] T. Bass, « Intrusion detection systems and multisensor data fusion », Communications of the ACM, vol. 43, no 4, p. 99–105, 2000.
[80] K. Liu, N. Z. Gebraeel, et J. Shi, « A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis », IEEE Transactions on Automation Science and Engineering, vol. 10, no 3, p. 652–664, 2013.
[81] Z. Li et R. M. Narayanan, « Data level fusion of multilook Inverse Synthetic Aperture Radar (ISAR) images », in Applied Imagery and Pattern Recognition Workshop, 2006. AIPR 2006. 35th IEEE, 2006, p. 2–2.
[82] F. R. Al-Osaimi, M. Bennamoun, et A. Mian, « Spatially optimized data-level fusion of texture and shape for face recognition », IEEE Transactions on Image Processing, vol. 21, no 2, p. 859–872, 2012.
[83] G. Niu, B.-S. Yang, et M. Pecht, « Development of an optimized condition-based maintenance system by data fusion and reliability-centered maintenance », Reliability Engineering & System Safety, vol. 95, no 7, p. 786–796, 2010.
[84] Y. Lu et J. E. Michaels, « Feature extraction and sensor fusion for ultrasonic structural health monitoring under changing environmental conditions », IEEE Sensors Journal, vol. 9, no 11, p. 1462–1471, 2009.
[85] C. Zhu et W. Sheng, « Human daily activity recognition in robot-assisted living using multi-sensor fusion », in Robotics and Automation, 2009. ICRA'09. IEEE International Conference on, 2009, p. 2154–2159.
[86] C. Zhu et W. Sheng, « Multi-sensor fusion for human daily activity recognition in robot-assisted living », in Proceedings of the 4th ACM/IEEE international conference on Human robot interaction, 2009, p. 303–304.
[87] P. K. Varshney, « Multisensor data fusion », Electronics & Communication Engineering Journal, vol. 9, no 6, p. 245–253, 1997.
[88] N. Oliver, E. Horvitz, et A. Garg, « Layered representations for human activity recognition », in Proceedings of the 4th IEEE International Conference on Multimodal Interfaces, 2002, p. 3.

[89] S. Liu, R. X. Gao, D. John, J. W. Staudenmayer, et P. S. Freedson, « Multisensor data fusion for physical activity assessment », IEEE Transactions on Biomedical Engineering, vol. 59, no 3, p. 687–696, 2012.

[90] C. Chen, R. Jafari, et N. Kehtarnavaz, « Action recognition from depth sequences using depth motion maps-based local binary patterns », in Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on, 2015, p. 1092–1099.

[91] R. Gravina, P. Alinia, H. Ghasemzadeh, et G. Fortino, « Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges », Information Fusion, vol. 35, p. 68–80, 2017.

[92] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, et P. J. Havinga, « Fusion of smartphone motion sensors for physical activity recognition », Sensors, vol. 14, no 6, p. 10146–10176, 2014.

[93] M. Shoaib, H. Scholten, et P. J. Havinga, « Towards physical activity recognition using smartphone sensors », in Ubiquitous intelligence and computing, 2013 ieee 10th international conference on and 10th international conference on autonomic and trusted computing (uic/atc), 2013, p. 80–87.

[94] X. Yang et Y. L. Tian, « Eigenjoints-based action recognition using naive-bayes-nearest-neighbor », in Computer vision and pattern recognition workshops (CVPRW), 2012 IEEE computer society conference on, 2012, p. 14–19.

[95] A. A. Efros, A. C. Berg, G. Mori, et J. Malik, « Recognizing action at a distance », in null, 2003, p. 726.

[96] J. Mantyjarvi, J. Himberg, et T. Seppanen, « Recognizing human motion with multiple acceleration sensors », in Systems, Man, and Cybernetics, 2001 IEEE International Conference on, 2001, vol. 2, p. 747–752.

[97] S. S. Keerthi et C.-J. Lin, « Asymptotic behaviors of support vector machines with Gaussian kernel », Neural computation, vol. 15, no 7, p. 1667–1689, 2003.

[98] M. W. Gardner et S. R. Dorling, « Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences », Atmospheric environment, vol. 32, no 14-15, p. 2627–2636, 1998.

[99] A. Y. Ng et M. I. Jordan, « On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes », in Advances in neural information processing systems, 2002, p. 841–848.

[100] C. Zhang et Y. Ma, Ensemble machine learning: methods and applications. Springer, 2012.

[101] P.-E. Aguera, K. Jerbi, A. Caclin, et O. Bertrand, « ELAN: a software package for analysis and visualization of MEG, EEG, and LFP signals », Computational intelligence and neuroscience, vol. 2011, p. 5, 2011.

[102] S. Khalifa, G. Lan, M. Hassan, A. Seneviratne, et S. K. Das, « Harke: Human activity recognition from kinetic energy harvesting data in wearable devices », IEEE Transactions on Mobile Computing, vol. 17, no 6, p. 1353–1368, 2018.

[103] W. Wang, A. X. Liu, M. Shahzad, K. Ling, et S. Lu, « Understanding and modeling of wifi signal based human activity recognition », in Proceedings of the 21st annual international conference on mobile computing and networking, 2015, p. 65–76.

[104] Y. Wang, K. Huang, et T. Tan, « Human activity recognition based on r transform », in Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, 2007, p. 1–8.

[105] A. M. Khan, Y.-K. Lee, S. Y. Lee, et T.-S. Kim, « A triaxial accelerometer-based physical-

activity recognition via augmented-signal features and a hierarchical recognizer », IEEE transactions on information technology in biomedicine, vol. 14, no 5, p. 1166–1172, 2010.

[106] F. S. Samaria et A. C. Harter, « Parameterisation of a stochastic model for human face identification », in Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on, 1994, p. 138–142.

[107] Z. He et L. Jin, « Activity recognition from acceleration data based on discrete consine transform and SVM », in Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on, 2009, p. 5041–5044.

[108] M. J. Snadden, J. M. McGuirk, P. Bouyer, K. G. Haritos, et M. A. Kasevich, « Measurement of the Earth's gravity gradient with an atom interferometer-based gravity gradiometer », Physical review letters, vol. 81, no 5, p. 971, 1998.

[109] E. M. Tapia et al., « Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor », in Wearable Computers, 2007 11th IEEE International Symposium on, 2007, p. 37–40.

[110] A. M. Khan, Y.-K. Lee, S. Y. Lee, et T.-S. Kim, « A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer », IEEE transactions on information technology in biomedicine, vol. 14, no 5, p. 1166–1172, 2010.

[111] Y. Kawasaki, Y. Maeda, et N. Fukushima, « Parallelized and Vectorized Implementation of DCT denoising with FMA instructions », in Proc. International Workshop on Advanced Image Technology (IWAIT), 2018.

[112] F. M. de S Matos et L. V. Batista, « Face recognition using DCT coefficients selection », in Proceedings of the 2008 ACM symposium on Applied computing, 2008, p. 1753–1757.

[113] M. Mahmoodi, L. A. James, et T. Johansen, « Automated advanced image processing for micromodel flow experiments; an application using labVIEW », Journal of Petroleum Science and Engineering, 2018.

[114] N. Ravi, N. Dandekar, P. Mysore, et M. L. Littman, « Activity recognition from accelerometer data », in Aaai, 2005, vol. 5, p. 1541–1546.

[115] P. C. Ribeiro, J. Santos-Victor, et P. Lisboa, « Human activity recognition from video: modeling, feature selection and classification architecture », in Proceedings of International Workshop on Human Activity Recognition and Modelling, 2005, p. 61–78.

[116] Q. Huynh-Thu, M. Barkowsky, et P. Le Callet, « The importance of visual attention in improving the 3D-TV viewing experience: Overview and new perspectives », IEEE Transactions on broadcasting, vol. 57, no 2, p. 421–431, 2011.

[117] S.-R. Ke, H. L. U. Thuc, Y.-J. Lee, J.-N. Hwang, J.-H. Yoo, et K.-H. Choi, « A review on video-based human activity recognition », Computers, vol. 2, no 2, p. 88–131, 2013.

[118] A. T. Duchowski, « Eye tracking methodology », Theory and practice, vol. 328, 2007.

[119] M. S. Ryoo et J. K. Aggarwal, « Recognition of composite human activities through context-free grammar based representation », in Computer vision and pattern recognition, 2006 ieee computer society conference on, 2006, vol. 2, p. 1709–1718.

[120] R. Cucchiara, C. Grana, M. Piccardi, et A. Prati, « Detecting moving objects, ghosts, and shadows in video streams », IEEE transactions on pattern analysis and machine intelligence, 2003.

[121] J. Wu, A. Osuntogun, T. Choudhury, M. Philipose, et J. M. Rehg, « A scalable approach to activity recognition based on object use », in Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, 2007, p. 1–8.

[122] M. Philipose et al., « Inferring activities from interactions with objects », IEEE pervasive computing, vol. 3, no 4, p. 50–57, 2004.

[123] H. S. Koppula, R. Gupta, et A. Saxena, « Learning human activities and object affordances from rgb-d videos », The International Journal of Robotics Research, vol. 32, no 8, p. 951–970, 2013.

[124] J. Sung, C. Ponce, B. Selman, et A. Saxena, « Unstructured human activity detection from rgbd images », in Robotics and Automation (ICRA), 2012 IEEE International Conference on, 2012, p. 842–849.

[125] D. J. Moore, I. A. Essa, et M. H. Hayes, « Exploiting human actions and object context for recognition tasks », in Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, 1999, vol. 1, p. 80–86.

[126] K. Kuutti, « Activity theory as a potential framework for human-computer interaction research », Context and consciousness: Activity theory and human-computer interaction, vol. 17, 1996.

[127] L. Xia, C.-C. Chen, et J. K. Aggarwal, « View invariant human action recognition using histograms of 3d joints », in Computer vision and pattern recognition workshops (CVPRW), 2012 IEEE computer society conference on, 2012, p. 20–27.

[128] A. Al-Rahayfeh et M. Faezipour, « Eye tracking and head movement detection: A state-of-art survey », IEEE journal of translational engineering in health and medicine, vol. 1, 2013.

[129] C. Chamaret et O. Le Meur, « Attention-based video reframing: validation using eye-tracking », in Pattern Recognition, 2008. ICPR 2008. 19th International Conference on, 2008, p. 1–4.

[130] M. Inoue, S. Inoue, et T. Nishida, « Deep recurrent neural network for mobile human activity recognition with high throughput », Artificial Life and Robotics, vol. 23, no 2, p. 173–185, 2018.

[131] Q. Huynh-Thu, M. Barkowsky, et P. Le Callet, « The importance of visual attention in improving the 3D-TV viewing experience: Overview and new perspectives », IEEE Transactions on broadcasting, vol. 57, no 2, p. 421–431, 2011.

[132] A. P. Kumar, R. Roy, S. Rawat, R. Chaturvedi, A. Sharma, et C. S. Kumar, « Speech Recognition with Combined MFCC, MODGDF and ZCPA Features Extraction Techniques Using NTN and MNTN Conventional Classifiers for Telugu Language », in Soft Computing: Theories and Applications, Springer, 2018, p. 697–706.

[133] J. P. Lesso et J. L. Melanson, « Speaker recognition », févr-2018.

[134] V. Vestman, D. Gowda, M. Sahidullah, P. Alku, et T. Kinnunen, « Speaker recognition from whispered speech: A tutorial survey and an application of time-varying linear prediction », Speech Communication, vol. 99, p. 62–79, 2018.

[135] A. V. Haridas, R. Marimuthu, et V. G. Sivakumar, « A critical review and analysis on techniques of speech recognition: The road ahead », International Journal of Knowledge-based and Intelligent Engineering Systems, vol. 22, no 1, p. 39–57, 2018.

[136] B. Barai, D. Das, N. Das, S. Basu, et M. Nasipuri, « Closed-Set Text-Independent Automatic Speaker Recognition System Using VQ/GMM », in Intelligent Engineering Informatics, Springer, 2018, p. 337–346.

[137] J. F. Thayer, S. S. Yamamoto, et J. F. Brosschot, « The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors », International journal of cardiology, vol. 141, no 2, p. 122–131, 2010.

[138] A. H. Kemp, D. S. Quintana, M. A. Gray, K. L. Felmingham, K. Brown, et J. M. Gatt, « Impact of depression and antidepressant treatment on heart rate variability: a review and meta-analysis », Biological psychiatry, vol. 67, no 11, p. 1067–1074, 2010.

[139] D. Phan, L. Y. Siong, P. N. Pathirana, et A. Seneviratne, « Smartwatch: Performance evaluation for long-term heart rate monitoring », in Bioelectronics and Bioinformatics (ISBB), 2015 International Symposium on, 2015, p. 144–147.

[140] S. C. Matta, Z. Sankari, et S. Rihana, « Heart rate variability analysis using neural network models for automatic detection of lifestyle activities », Biomedical Signal Processing and Control, vol. 42, p. 145–157, 2018.

[141] V. P. da Silva et al., « Heart Rate Variability Indexes in Dementia: A Systematic Review with a Quantitative Analysis », Current Alzheimer Research, vol. 15, no 1, p. 80–88, 2018.

[142] A. Tolstikov, X. Hong, J. Biswas, C. Nugent, L. Chen, et G. Parente, « Comparison of fusion methods based on dst and dbn in human activity recognition », Journal of Control Theory and Applications, vol. 9, no 1, p. 18–27, 2011.

[143] A. H. Milaghardan, M. Delavar, et A. Chehreghan, « Uncertainty in landslide occurrence prediction using Dempster–Shafer theory », Modeling Earth Systems and Environment, vol. 2, no 4, p. 188, 2016.

[144] J. Liao, Y. Bi, et C. Nugent, « Evidence fusion for activity recognition using the Dempster-Shafer theory of evidence », in Information Technology and Applications in Biomedicine, 2009. ITAB 2009. 9th International Conference on, 2009, p. 1–4.

[145] J. Liao, Y. Bi, et C. Nugent, « Using the Dempster–Shafer theory of evidence with a revised lattice structure for activity recognition », IEEE Transactions on Information Technology in Biomedicine, vol. 15, no 1, p. 74–82, 2011.

[146] F. Sebbak, F. Benhammadi, A. Chibani, Y. Amirat, et A. Mokhtari, « Dempster–Shafer theory-based human activity recognition in smart home environments », annals of telecommunications-annales des télécommunications, vol. 69, no 3-4, p. 171–184, 2014.

[147] S. Mckeever, J. Ye, L. Coyle, C. Bleakley, et S. Dobson, « Activity recognition using temporal evidence theory », Journal of Ambient Intelligence and Smart Environments, vol. 2, no 3, p. 253–269, 2010.

[148] A. Perez, H. Tabia, D. Declercq, et A. Zanotti, « Using the conflict in Dempster–Shafer evidence theory as a rejection criterion in classifier output combination for 3D human action recognition », Image and Vision Computing, vol. 55, p. 149–157, 2016.

[149] E. Javadi, B. Moshiri, et H. S. Yazdi, « Activity Recognition In Smart Home Using Weighted Dempster-Shafer Theory », Adm, vol. 1, no 1, p. 4, 2013.

[150] R. Kennes et P. Smets, « Computational aspects of the Mobius transform », arXiv preprint arXiv:1304.1122, 2013.

[151] J. A. Barnett, « Computational Methods for a Mathematical Theory of Evidence. », in IJCAI, 1981, vol. 81, p. 868–875.

[152] G. Shafer et R. Logan, « Implementing Dempster's rule for hierarchical evidence », Artificial Intelligence, vol. 33, no 3, p. 271–298, 1987.

[153] W. He, N. Williard, M. Osterman, et M. Pecht, « Prognostics of lithium-ion batteries based on Dempster–Shafer theory and the Bayesian Monte Carlo method », Journal of Power Sources, vol. 196, no 23, p. 10314–10321, 2011.

[154] M. Bauer, « Approximation algorithms and decision making in the Dempster-Shafer theory of evidence—An empirical study », International Journal of Approximate Reasoning, vol. 17, no 2-3, p. 217–237, 1997.

[155] T. Denoeux, « A k-nearest neighbor classification rule based on Dempster-Shafer theory », IEEE transactions on systems, man, and cybernetics, vol. 25, no 5, p. 804–813, 1995.

[156] M. Bauer, « Approximations for decision making in the Dempster-Shafer theory of

evidence », in Proceedings of the Twelfth international conference on Uncertainty in artificial intelligence, 1996, p. 73–80.

[157] J. D. Lowrance, T. D. Garvey, et T. M. Strat, « A framework for evidential-reasoning systems », in Classic Works of the Dempster-Shafer Theory of Belief Functions, Springer, 2008, p. 419–434.

[158] T. L. Wickramarathne, K. Premaratne, et M. N. Murthi, « Monte-Carlo approximations for Dempster-Shafer belief theoretic algorithms », in Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on, 2011, p. 1–8.

[159] Y. Yang, D. Han, C. Han, et F. Cao, « A novel approximation of basic probability assignment based on rank-level fusion », Chinese Journal of Aeronautics, vol. 26, no 4, p. 993–999, 2013.

[160] L. A. Klein et L. A. Klein, Sensor and data fusion: a tool for information assessment and decision making, vol. 324. SPIE press Bellingham^ eWA WA, 2004.

[161] D. Han, J. Dezert, et Y. Yang, « Two novel methods for BBA approximation based on focal element redundancy », in Information Fusion (Fusion), 2015 18th International Conference on, 2015, p. 428–434.

[162] Y. Yang et Y. Liu, « Iterative approximation of basic belief assignment based on distance of evidence », PloS one, vol. 11, no 2, p. e0147799, 2016.

[163] D. Yong, S. WenKang, Z. ZhenFu, et L. Qi, « Combining belief functions based on distance of evidence », Decision support systems, vol. 38, no 3, p. 489–493, 2004.

[164] W. Jiang, « A correlation coefficient of belief functions », arXiv preprint arXiv:1612.05497, 2016.

[165] M. F. Møller, « A scaled conjugate gradient algorithm for fast supervised learning », Neural networks, vol. 6, no 4, p. 525–533, 1993.

[166] M. Zeng et al., « Convolutional neural networks for human activity recognition using mobile sensors », in Mobile Computing, Applications and Services (MobiCASE), 2014 6th International Conference on, 2014, p. 197–205.

[167] Q. V. Le, W. Y. Zou, S. Y. Yeung, et A. Y. Ng, « Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis », in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, 2011, p. 3361–3368.

[168] F. J. Ordóñez et D. Roggen, « Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition », Sensors, vol. 16, no 1, p. 115, 2016.

[169] C. A. Ronao et S.-B. Cho, « Human activity recognition with smartphone sensors using deep learning neural networks », Expert Systems with Applications, vol. 59, p. 235–244, 2016.

[170] R. Girshick, J. Donahue, T. Darrell, et J. Malik, « Rich feature hierarchies for accurate object detection and semantic segmentation », in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, p. 580–587.

[171] T. Plötz, N. Y. Hammerla, et P. Olivier, « Feature learning for activity recognition in ubiquitous computing », in IJCAI Proceedings-International Joint Conference on Artificial Intelligence, 2011, vol. 22, p. 1729.

[172] B. G. Cragg, « The topography of the afferent projections in the circumstriate visual cortex of the monkey studied by the Nauta method », Vision research, vol. 9, no 7, p. 733–747, 1969.

[173] M. Koperski, « Human Action Recognition in Videos with Local Representation », PhD Thesis, Universite Cote d'Azur, 2017.

[174] M. Lin, Q. Chen, et S. Yan, « Network in network », arXiv preprint arXiv:1312.4400, 2013.
[175] H. Tang, R. Ni, Y. Zhao, et X. Li, « Median filtering detection of small-size image based on CNN », Journal of Visual Communication and Image Representation, vol. 51, p. 162–168, 2018.
[176] Y. LeCun, Y. Bengio, et G. Hinton, « Deep learning », nature, vol. 521, no 7553, p. 436, 2015.
[177] J. Martens, « Deep learning via Hessian-free optimization. », in ICML, 2010, vol. 27, p. 735–742.
[178] R. Pascanu, T. Mikolov, et Y. Bengio, « On the difficulty of training recurrent neural networks », in International Conference on Machine Learning, 2013, p. 1310–1318.
[179] X. Glorot et Y. Bengio, « Understanding the difficulty of training deep feedforward neural networks », in Proceedings of the thirteenth international conference on artificial intelligence and statistics, 2010, p. 249–256.
[180] D.-A. Clevert, T. Unterthiner, et S. Hochreiter, « Fast and accurate deep network learning by exponential linear units (elus) », arXiv preprint arXiv:1511.07289, 2015.
[181] B. Xu, N. Wang, T. Chen, et M. Li, « Empirical evaluation of rectified activations in convolutional network », arXiv preprint arXiv:1505.00853, 2015.
[182] G. Klambauer, T. Unterthiner, A. Mayr, et S. Hochreiter, « Self-normalizing neural networks », in Advances in Neural Information Processing Systems, 2017, p. 972–981.
[183] S. Ioffe et C. Szegedy, « Batch normalization: Accelerating deep network training by reducing internal covariate shift », arXiv preprint arXiv:1502.03167, 2015.
[184] Y. Nesterov, « A method for unconstrained convex minimization problem with the rate of convergence O (1/k^ 2) », in Doklady AN USSR, 1983, vol. 269, p. 543–547.
[185] J. Duchi, E. Hazan, et Y. Singer, « Adaptive subgradient methods for online learning and stochastic optimization », Journal of Machine Learning Research, vol. 12, no Jul, p. 2121–2159, 2011.
[186] J. Dean et al., « Large scale distributed deep networks », in Advances in neural information processing systems, 2012, p. 1223–1231.
[187] D. P. Kingma et J. Ba, « Adam: A method for stochastic optimization », arXiv preprint arXiv:1412.6980, 2014.
[188] A. C. Wilson, R. Roelofs, M. Stern, N. Srebro, et B. Recht, « The marginal value of adaptive gradient methods in machine learning », in Advances in Neural Information Processing Systems, 2017, p. 4151–4161.
[189] N. Y. Hammerla, S. Halloran, et T. Ploetz, « Deep, convolutional, and recurrent models for human activity recognition using wearables », arXiv preprint arXiv:1604.08880, 2016.
[190] A.-L. Jousselme, D. Grenier, et É. Bossé, « A new distance between two bodies of evidence », Information fusion, vol. 2, no 2, p. 91–101, 2001.
[191] A. J. Perez, M. A. Labrador, and S. J. Barbeau. G-sense: A scalable architecture for global sensing and monitoring. IEEE Network, 24(4):57– 64, 2010.
[192] Y. Jia. Diatetic and exercise therapy against diabetes mellitus. In Second International Conference on Intelligent Networks and Intelligent Systems, pages 693–696, 2009.
[193] J. Yin, Q. Yang, and J. J. Pan. Sensor-based abnormal human-activity detection. IEEE Transactions on Knowledge and Data Engineering, 20(8):1082–1090, 2008.
[194] M. Bobin, H. Amroun, S. Coquillart, F. Bimbard, et M. Ammi, « SVM based approach for the assessment of elbow flexion with smart textile sensor », in Systems, Man, and Cybernetics (SMC), 2017 IEEE International Conference on, 2017, p. 2129–2134.

[195] M. Ahad, J.K. Tan, H.S. Kim, and S. Ishikawa. Human activity recognition: Various paradigms. In International Conference on Control, Automation and Systems, pages 1896–1901, 2008.

[196] A. Ali, R.C. King, and G-Z. Yang. Semi-supervised segmentation for activity recognition with multiple eigenspaces. In International Summer School and Symposium on Medical Devices and Biosensors, pages 314–317, 2008.

[197] K. Altun and B. Barshan. Human activity recognition using inertial/- magnetic sensor units. In Human Behavior Understanding, Lecture Notes in Computer Science, pages 38–51. Springer Berlin/Heidelberg, 2010.

[198] Datasets for Human Activity Recognition from ETH's Wearable Computing Lab. http://www.wearable.ethz.ch/resources/Dataset.

[199] Datasets for Human Activity Recognition from MIT Media Lab. http://architecture.mit.edu/house_n/data/PlaceLab/PlaceLab.htm.

[200] Datasets for Human Activity Recognition from Tanzeem Choudhury's website. http://www.cs.dartmouth.edu/~tanzeem/teaching/ CS188-Fall08/dataset.html.

[201] M. E. Liechti, M. R. Saur, A. Gamma, D. Hell, et F. X. Vollenweider, « Psychological and physiological effects of MDMA ("Ecstasy") after pretreatment with the 5-HT2 antagonist ketanserin in healthy humans », Neuropsychopharmacology, vol. 23, no 4, p. 396–404, 2000.

[202] Datasets for Human Activity Recognition from Tim Van Kasteren's website. https://sites.google.com/site/tim0306/datasets.

[203] Google Places API. http://code.google.com/apis/maps/ documentation/places/.

[204] P. Antal. Construction of a classifier with prior domain knowledge formalized as bayesian network. In Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society, volume 4, pages 2527–2531, 1998.

[205] D. Arvidsson, F. Slinde, and L. Hulth´en. Free-living energy expenditure in children using multi-sensor activity monitors. Clinical nutrition (Edinburgh, Scotland), 28:3, s. 305, 2009.

[206] L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. In Pervasive, pages 1–17, 2004.

[207] M. Berchtold, M. Budde, D. Gordon, H.R. Schmidtke, and M. Beigl.Actiserv: Activity recognition service for mobile phones. In International Symposium on Wearable Computers, pages 1–8, 2010.

[208] M. Berchtold, M. Budde, D. Gordon, H.R. Schmidtke, and M. Beigl.Actiserv: Activity recognition service for mobile phones. pages 1–8,2010.

[209] M. Berchtold, M. Budde, H. Schmidtke, and M. Beigl. An extensible modular recognition concept that makes activity recognition practical. In Advances in Artificial Intelligence, Lecture Notes in Computer Science, pages 400–409. Springer Berlin/Heidelberg, 2010.

[210] U. Blanke and B. Schiele. Remember and transfer what you have learned- recognizing composite activities based on activity spotting. In International Symposium on Wearable Computers, pages 1–8, 2010.

[211] A. Blum and T. Mitchell. Combining labeled and unlabeled data with co-training. In Proceedings of the eleventh annual conference on Computational learning theory, pages 92–100. ACM, 1998.

[212] T. Brezmes, J. L. Gorricho, and J. Cotrina. Activity recognition from accelerometer data on a mobile phone. In Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living, volume 5518, pages 796–799. Springer Berlin/Heidelberg, 2009.

[213] T. Brezmes, J. L. Gorricho, and J. Cotrina. Activity recognition from accelerometer data on a mobile phone. In Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living, volume 5518 of Lecture Notes in Computer Science, pages 796–799. Springer Berlin/Heidelberg, 2009.

[214] J. Candamo, M. Shreve, D.B. Goldgof, D.B. Sapper, and R. Kasturi. Understanding transit scenes: A survey on human behavior-recognition algorithms. IEEE Transactions on Intelligent Transportation Systems, 11(1):206–224, 2010.

[215] P. Cavalin, R. Sabourin, and C. Suen. Dynamic selection approaches for multiple classifier systems. Neural Computing and Applications, pages 1–16, 2011.

[216] Y-P. Chen, J-Y. Yang, S-N. Liou, G-Y. Lee, and J-S. Wang. Online classifier construction algorithm for human activity detection using a tri-axial accelerometer. Applied Mathematics and Computation, 205(2):849–860, 2008.

[217] J. Cheng, O. Amft, and P. Lukowicz. Active capacitive sensing: Exploring a new wearable sensing modality for activity recognition. In Pervasive Computing, volume 6030 of Lecture Notes in Computer Science, pages 319–336. Springer Berlin/Heidelberg, 2010.

[218] D. Choujaa and N. Dulay. Tracme: Temporal activity recognition using mobile phone data. In IEEE/IFIP International Conference on Embedded and Ubiquitous Computing, volume 1, pages 119–126, 2008.

[219] D. Christin, A. Reinhardt, S. S. Kanhere, and M. Hollick. A survey on privacy in mobile participatory sensing applications. Journal of Systems and Software, 84(11):1928–1946, 2011.

[200] D. Coppersmith and S. Winograd. Matrix multiplication via arithmetic progressions. In Proceedings of the nineteenth annual ACM symposium on Theory of computing, pages 1–6, 1987.

[211] C. Cortes and V. Vapnik. Support-vector networks. Machine Learning, 20:273–297, 1995.

[222] G. Dahlquist and A. Bjorck. Numerical methods in scientific computing. vol 1. SIAM, 2008.

[223] W. H. E. Day. Consensus methods as tools for data analysis. In Conference of the International Federation of Classification Societies, pages 317–324, 1988.

[224] T. G. Dietterich. Approximate statistical tests for comparing supervised classification learning algorithms. Neural Computation, 10:1895–1923, 1998.

[225] R. P. W. Duin. The combining classifier: to train or not to train? In Proceedings of the 16th International Conference on Pattern Recognition, volume 2, pages 765–770, 2002.

[226] C. Elkan. The foundations of cost-sensitive learning. In Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, pages 973–978, 2001.

[227] M. Ermes, J. Parkka, and L. Cluitmans. Advancing from offline to online activity recognition with wearable sensors. In Engineering in Medicine and Biology Society. 30th Annual International Conference of the IEEE, pages 4451–4454, 2008.

[228] JAMA: A Java Matrix Package. http://mste.illinois.edu/users/ exner/java.f/leastsquares/.

[229] Source code for the Least Square algorithm. http://mste.illinois. edu/users/exner/java.f/leastsquares/.

[230] The Java Data Mining Platform. http://www.jdmp.org/.

[231] UCI datasets. http://archive.ics.uci.edu/ml/datasets.html.

[232] Use Weka in your Java code. http://weka.wikispaces.com/Use+ WEKA+in+your+Java+code.

[233] Vigilante: A Mobile Platform for Real-time Human Activity Recognition. http://www.youtube.com/watch?v=4Jen6O1lfms.

[234] Zephyr Bioharness BT Website. http://www.zephyr-technology. com/bioharness-

bt.html.
[235] G. Giacinto and F. Roli. Dynamic classifier selection based on multiple classifier behaviour. Pattern Recognition, 34:1879–1881, 2001.
[236] K. Goebel and W. Yan. Using correlation-based measures to select classifiers for decision fusion, 2005.
[237] D. Guan, W. Yuan, Y-K. Lee, A. Gavrilov, and S. Lee. Activity recognition based on semi-supervised learning. In 13th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications,pages 469–475, 2007.
[238] M. A. Hall. Correlation-based Feature Selection for Machine Learning.1999.
[239] Y. Hanai, J. Nishimura, and T. Kuroda. Haar-like filtering for human activity recognition using 3d accelerometer. In IEEE 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop, pages 675–678, 2009.
[240] R. Helaoui, M. Niepert, and H. Stuckenschmidt. Recognizing interleaved and concurrent activities: A statistical-relational approach. In IEEE International Conference on Pervasive Computing and Communications (PerCom), pages 1–9, 2011.
[241] T. K. Ho. The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(8):832–844, Aug 1998.
[242] K. L. Huang, S. S. Kanhere, and W. Hu. Preserving privacy in participatory sensing systems. Computer Communications, 33(11):1266–1280, 2010.
[243] T. Huynh and B. Schiele. Towards less supervision in activity recognition from wearable sensors. In 10th IEEE International Symposium on Wearable Computers, pages 3–10, 2006.
[244] J. Iglesias, J. Cano, A. M. Bernardos, and J. R. Casar. A ubiquitous activity-monitor to prevent sedentariness. In IEEE PerCom, pages 668–670, 2011.
[245] L. C. Jatoba, U. Grossmann, C. Kunze, J. Ottenbacher, and W. Stork. Context-aware mobile health monitoring: Evaluation of different pattern recognition methods for classification of physical activity. In 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pages 5250–5253, 2008.
[246] C. N. Joseph, S. Kokulakumaran, K. Srijeyanthan, A. Thusyanthan, C. Gunasekara, and C.D. Gamage. A framework for whole-body gesture recognition from video feeds. In International Conference on Industrial and Information Systems (ICIIS), pages 430–435, 2010.
[247] T-P. Kao, C-W. Lin, and J-S. Wang. Development of a portable activity detector for daily activity recognition. In IEEE International Symposium on Industrial Electronics, pages 115–120, 2009.
[248] A. M. Khan, Y. K. Lee, and S. Y. Lee. Accelerometer's position free human activity recognition using a hierarchical recognition model. In 12th IEEE International Conference on e-Health Networking Applications and Services (Healthcom), pages 296–301, 2010.
[249] U. Maurer, A. Smailagic, D. Siewiorek, and M. Deisher. Activity recognition and monitoring using multiple sensors on different body positions. In Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks, Washington, DC, USA, 2006. IEEE Computer Society.
[250] U. Maurer, A. Smailagic, D. Siewiorek, and M. Deisher. Activity recognition and monitoring using multiple sensors on different body positions. In Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks, pages 113–116, 2006.
[251] D. McGlynn and M. G. Madden. An ensemble dynamic time warping classifier with

application to activity recognition. In Research and Development in Intelligent Systems XXVII, pages 339–352. Springer London, 2011.

[252] K. Oh, H-S. Park, and S-B. Cho. A mobile context sharing system using activity and emotion recognition with bayesian networks. In International Conference on Ubiquitous Intelligence Computing, pages 244–249, 2010.

[253] R. T. Olszewski, C. Faloutsos, and D. Banks Dot. Generalized Feature Extraction for Structural Pattern Recognition in Time-Series Data. 2001.

[254] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen. Activity classification using realistic data from wearable sensors. IEEE Transactions on Information Technology in Biomedicine, 10(1):119–128, 2006.

[255] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and minredundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(8):1226–1238, 2005.

[256] J. Penttil, J. Peltola, and T. Seppnen. A speech/music discriminator based audio browser with a degree of certanity measure. In Proceedings of the International Workshop Information Retrieval, page 125131, 2001.

[257] N. Pham and T. Abdelzaher. Robust dynamic human activity recognition based on relative energy allocation. In Distributed Computing in Sensor Systems, volume 5067 of Lecture Notes in Computer Science, pages 525–530. Springer Berlin/Heidelberg, 2008.

[258] J.R. Quinlan. C4.5: programs for machine learning. Morgan Kaufmann series in machine learning. Morgan Kaufmann Publishers, 1993.

[259] C. Randell and H.Muller. Context awareness by analysing accelerometer data. In The Fourth International Symposium on Wearable Computers, pages 175–176, 2000.

[260] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava. Using mobile phones to determine transportation modes. ACM Transactions on Sensor Networks, 6(2):1–27, 2010.

[261] D. Riboni and C. Bettini. Cosar: hybrid reasoning for context-aware activity recognition. Personal and Ubiquitous Computing, 15:271–289, 2011.

[262] A. Santana, R. Soares, A. Canuto, and M. de Souto. A dynamic classifier selection method to build ensembles using accuracy and diversity. In SBRN. Ninth Brazilian Symposium on Neural Networks, pages 36–41, Oct. 2006.

[263] J. Sarkar, L. Vinh, Y-K. Lee, and S. Lee. Gpars: a general-purpose activity recognition system. Applied Intelligence, pages 1–18, 2010.

[264] O. Schlmilch, B. Witzschel, M. Cantor, E. Kahl, R. Mehmke, and C. Runge. Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. Computers in Human Behavior, 15(5):571–583, 1999.

[265] F. Foerster, M. Smeja, and J. Fahrenberg. Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. Computers in Human Behavior, 15(5):571–583, 1999.

[266] O. D. Lara and M. A. Labrador. A survey on human activity recognition using wearable sensors. IEEE Communications Surveys and Tutorials, (to appear), 2012.

[267] O. D. Lara and M. A. Labrador. A mobile platform for real-time human activity recognition. In IEEE Consumer Communications and Networking Conference (CCNC), pages 667 –671, January 2012.

[268] E. Munguia Tapia, S. S. Intille, W. Haskell, K. Larson, J. Wright, A. King, and R. Friedman. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart monitor. In Proceedings of the International Symposium on

Wearable Computers, 2007.
[269] Google's Android becomes the world's leading smart phone platform. http://www.canalys.com/newsroom/google%E2%80%99s-android-becomes-world%E2%80%99s-leading-smart-phone-platform.
[270] A. Teller. A platform for wearable physiological computing. Interacting with Computers, pages 917–937, 2004.
[271] A.M. Khan, Y-K. Lee, S.Y. Lee, and T-S. Kim. A triaxial accelerometerbased physical-activity recognition via augmented-signal features and a hierarchical recognizer. IEEE Transactions on Information Technology in Biomedicine, 14(5):1166–1172, 2010.
[272] C. Zhu and W. Sheng. Human daily activity recognition in robotassisted living using multi-sensor fusion. In IEEE International Conference on Robotics and Automation, pages 2154–2159, 2009.
[273] O. D. Lara, A. J. Perez, M. A. Labrador, and J. D. Posada. Centinela: A human activity recognition system based on acceleration and vital sign data. Pervasive and Mobile Computing, 8:717–729, 2012.
[274] Z-Y. He and L-W. Jin. Activity recognition from acceleration data using ar model representation and svm. In International Conference on Machine Learning and Cybernetics, volume 4, pages 2245–2250, 2008.
[275] Z. He and L. Jin. Activity recognition from acceleration data based on discrete consine transform and svm. In IEEE International Conference on Systems, Man and Cybernetics, pages 5041–5044, 2009.
[276] La Vinh, Sungyoung Lee, Hung Le, Hung Ngo, Hyoung Kim, Manhyung Han, and Young-Koo Lee. Semi-markov conditional random fields for accelerometer-based activity recognition. Applied Intelligence, 35:226– 241, 2011.
[278] Z. He, Z. Liu, L. Jin, L-X. Zhen, and J-C. Huang. Weightlessness feature; a novel feature for single tri-axial accelerometer based activity recognition. In 19th International Conference on Pattern Recognition, pages 1–4, 2008.
[279] D. Minnen, T. Westeyn, D.l Ashbrook, P. Presti, and T. Starner. Recognizing soldier activities in the field. In 4th International Workshop on Wearable and Implantable Body Sensor Networks, volume 13, pages 236–241. Springer Berlin/Heidelberg, 2007.
[280] M. Stikic, D. Larlus, S. Ebert, and B. Schiele. Weakly supervised recognition of daily life activities with wearable sensors. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(12):2521–2537, 2011.
[281] M. Stikic, D. Larlus, and B. Schiele. Multi-graph based semi-supervised learning for activity recognition. In International Symposium on Wearable Computers, pages 85–92, 2009.
[282] H. Banaee , M.U. Ahmed , A. Loutfi, Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges, Sensors 13 (12) (2013) 17472–17500 .
[283] A. Milenkovi ´c , C. Otto , E. Jovanov , Wireless sensor networks for personal health monitoring: issues and an implementation, Comput.Commun. 29 (13) (2006) 2521–2533 .
[284] K.M. Reichard , M. Van Dyke , K. Maynard , Application of sensor fusion and sig- nal classification techniques in a distributed machinery condition monitoring system, in: AeroSense 20 0 0, International Society for Optics and Photonics, 20 0 0, pp. 329–336 .
[285] G. Fortino , V. Giampá, PPG-based methods for non invasive and continuous blood pressure measurement: an overview and development issues in body sensor networks, in: IEEE International Workshop on Medical Measurements and Applications (MeMeA),

2010, pp. 10–13 .
[286] C. Otto , A. Milenkovic , C. Sanders , E. Jovanov , System architecture of a wire- less body area sensor network for ubiquitous health monitoring, J.Mobile Multimedia 1 (4) (2006) 307–326 .
[287] B.P. Lo , S. Thiemjarus , R. King , G.-Z. Yang , Body sensor network - a wire- less sensor platform for pervasive healthcare monitoring, in: 3rd International Conference on Pervasive Computing (PerCom), 2005 .
[288] W.-Y. Chung , S. Bhardwaj , A. Purwar , D.-S. Lee , R. Myllylae , A fusion health monitoring using ecg and accelerometer sensors for elderly persons at home, in: Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th An- nual International Conference of the IEEE, IEEE, 2007, pp. 3818–3821 .
[289] W. Elmenreich , An Introduction to Sensor Fusion - Research Report, Vienna University of Technology, Austria, 2002 .
[290] R. Murphy , Biological and cognitive foundations of intelligent sensor fusion, IEEE Trans. Syst. Man Cybern. 26 (1) (1996) 42–51 .
[291] D. Cook , K.D. Feuz , N.C. Krishnan , Transfer learning for activity recognition: asurvey, Knowl. Inf. Syst. 36 (3) (2012) 537–556 .
[292] A. Augimeri , G. Fortino , S. Galzarano , R. Gravina , Collaborative body sensor networks, in: IEEE International Conference on Systems, Man, and Cybernetics (SMC 2011), 2011, pp. 3427–3432 .
[293] G. Fortino , S. Galzarano , R. Gravina , W. Li , A framework for collaborative com-puting and multi-sensor data fusion in body sensor networks, Inf. Fusion 22 (2015) 50–70 .
[294] S. Galzarano , R. Giannantonio , A. Liotta , G. Fortino , A task-oriented framework for networked wearable computing, IEEE Trans. Autom. Sci. Eng. 13 (2) (2016) 621–638 .
[295] G. Fortino , A. Guerrieri , R. Giannantonio , F. Bellifemine , Platform-indepen- dent development of collaborative WBSN applications: SPINE2, in: IEEE In-ternational Conference on Systems, Man, and Cybernetics (SMC 20 09), 20 09a, pp. 3144–3150 .
[296] G. Fortino , A. Guerrieri , R. Giannantonio , F. Bellifemine , SPINE2: developing BSN applications on heterogeneous sensor nodes, in: IEEE Symposium on In- dustrial Embedded Systems (SIES'09), 2009b, pp. 128–131 .
[297] A. Augimeri , G. Fortino , M. Rege , V. Handziski , A. Wolisz , A cooperative ap- proach for handshake detection based on body sensor networks, in: IEEE In- ternational Conference on Systems, Man and Cybernetics (SMC 2010), 2010, pp. 281–288 .
[298] G. Fortino , G.D. Fatta , M. Pathan , A. Vasilakos , Cloud-assisted body area net- works: State-of-the-art and future research challenges, Wireless Netw. 20 (7) (2014a) 1925–1938 .
[299] G. Fortino , D. Parisi , V. Pirrone , G.D. Fatta , Bodycloud: A saas approach for community body sensor networks, Future Gener. Comput. Syst. 35 (2014b) 62–79 .
[300] M.L. II , D. Hall , J. Llinas , Handbook of Multisensor Data Fusion: Theory and Practice, CRC Press, 2008 .
[301] G.-Z. Yang , Body Sensor Networks, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006 .
[302] B.V. Dasarathy , Sensor fusion potential exploitation-innovative architectures and illustrative applications, Proc. IEEE 85 (1) (1997) 24–38 .
[303] D. Schuldhaus , H. Leutheuser , B.M. Eskofier ,Towards big data for activity recognition: a novel database fusion strategy, in: 9th International Conference on Body Area Networks, 2014, pp. 97–103 .

[304] C. Chen , R. Jafari , N. Kehtarnavaz , A survey of depth and inertial sensor fusion for human action recognition, Multimedia Tools Appl. (2015) 1–21 http://link.springer.com/article/10.1007/s11042-015-3177-1 .

[305] W.-J. Yi , O. Sarkar , S. Mathavan , J. Saniie , Wearable sensor data fusion for re- mote health assessment and fall detection, in: Electro/Information Technology (EIT), 2014 IEEE International Conference on, IEEE, 2014, pp. 303–307 .

[306] F. Felisberto , F. Fdez-Riverola , A. Pereira , A ubiquitous and low-cost solution for movement monitoring and accident detection based on sensor fusion, Sensors 14 (5) (2014) 8961–8983 .

[307] F. Sanfilippo , K. Pettersen , A sensor fusion wearable health-monitoring sys- tem with haptic feedback, in: 11th IEEE International Conference on Inno- vations in Information Technology (IIT), Dubai, United Arab Emirates, 2015, pp. 262–266 .

[308] P. Zappi , T. Stiefmeier , E. Farella , D. Roggen , L. Benini , G. Troster , Activity recognition from on-body sensors by classifier fusion: sensor scalability and robustness, in: Intelligent Sensors, Sensor Networks and Information, 2007. ISSNIP 2007. 3rd International Conference on, 2007, pp. 281–286.

[309] A. Bulling , U. Blanke , B. Schiele , A tutorial on human activity recognition us- ing body-worn inertial sensors, ACM Comput. Surv. 46 (3) (2014) 1–33 .

[310] I.M. Pires , N.M. Garcia , N. Pombo , F. Flórez-Revuelta , From data acquisition to data fusion: a comprehensive review and a roadmap for the identification of activities of daily living using mobile devices, Sensors 16 (2) (2016) 184 .

[311] H. Ghasemzadeh , N. Amini , R. Saeedi , M. Sarrafzadeh , Power-aware comput- ing in wearable sensor networks: an optimal feature selection, IEEE Trans. Mobile Comput. 14 (4) (2015) 800–812 .

[312] R. Saeedi , B. Schimert , H. Ghasemzadeh , Cost-sensitive feature selection for on-body sensor localization, in: 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, ACM, 2014, pp. 833–842 .

[313] R. Saeedi , R. Fallahzadeh , P. Alinia , H. Ghasemzadeh , An energy-efficient computational model for uncertainty management in dynamically changing net-worked wearables, in: International Symposium on Low Power Electronics and Design (ISLPED 2016), 2016, pp. 46–51.

[314] C. Zhu , W. Sheng , Human daily activity recognition in robot-assisted living using multi-sensor fusion, in: Robotics and Automation, 2009. ICRA'09. IEEE International Conference on, 2009, pp. 2154–2159 .

[315] N.A. Capela , E.D. Lemaire , N. Baddour , Feature selection for wearable smartphone-based human activity recognition with able bodied, elderly, and stroke patients, PLoS ONE 10 (4) (2015) 1–18 .

[316] C. Chen , R. Jafari , N. Kehtarnavaz , Improving human action recognition using fusion of depth camera and inertial sensors, IEEE Trans. Hum.-Mach. Syst. 45 (1) (2015) 51–61 .

[317] A .M. Khan , A . Tufail , A .M. Khattak , T.H. Laine , Activity recognition on smartphones via sensor-fusion and kda-based svms, Int. J. Distrib. Sensor Netw. 10 (5) (2014) 1–14 .

[318] U. Maurer , A. Smailagic , D.P. Siewiorek , M. Deisher , Activity recognition and monitoring using multiple sensors on different body positions, in: International Workshop on Wearable and Implantable Body Sensor Networks (BSN20 06), 20 06, pp. 113–116 .

[319] S. Liu , R.X. Gao , D. John , J.W. Staudenmayer , P.S. Freedson , Multisensor data fusion for physical activity assessment, IEEE Trans. Biomed. Eng. 59 (3) (2012) 687–696 .

[320] M. Zeng , X. Wang , L.T. Nguyen , P. Wu , O.J. Mengshoel , J. Zhang , Adaptive activity recognition with dynamic heterogeneous sensor fusion, in: 6th International Conference on Mobile Computing, Applications and Services (MobiCASE 2014), 2014, pp. 189–196.

[321] A.R. Maria , P. Sever , V. Carlos , Biomedical sensors data fusion algorithm for enhancing the efficiency of fault-tolerant systems in case of wearable electronics device, in: Conference on Grid, Cloud & High Performance Computing in Science (ROLCG 2015), IEEE, 2015, pp. 1–4 .

[322] N. Ravi , N. Dandekar , P. Mysore , M.L. Littman , Activity recognition from accelerometer data, in: 17th Conference on Innovative Applications of Artificial Intelligence - Volume 3, in: IAAI'05, 2005, pp. 1541–1546 .

[323] L. Mo , S. Liu , R.X. Gao , P.S. Freedson , Energy-efficient and data synchronized body sensor network for physical activity measurement, in: Instrumentation and Measurement Technology Conference (I2MTC), 2013 IEEE International, IEEE, 2013, pp. 1120–1124

[324] L. Mo , S. Liu , R.X. Gao , P.S. Freedson , Multi-sensor ensemble classifier for activity recognition, J. Softw. Eng.Appl. 5 (12) (2012) 113–116 .

[325] A.Y. Yang , R. Jafari , S.S. Sastry , R. Bajcsy , Distributed recognition of human actions using wearable motion sensor networks, J. Ambient Intell. Smart Environ. 1 (2) (2009) 103–115 .

[326] N. Raveendranathan , S. Galzarano , V. Loseu , R. Gravina , R. Giannantonio , M. Sgroi , R. Jafari , G. Fortino , From modeling to implementation of virtual sensors in body sensor networks, IEEE Sensors J. 12 (3) (2012) 583–593 .

[327] M. Bahrepour , N. Meratnia , Z. Taghikhaki , P.J. Havinga , Sensor fusion-based activity recognition for parkinson patients, in: Sensor Fusion - Foundation and Applications, InTech, 2011, pp. 171–190 .

[328] L. Gao , A.K. Bourke , J. Nelson , Activity recognition using dynamic multiple sensor fusion in body sensor networks, in: International conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2012), 2012, pp. 1077–1080 .

[329] F. Aiello , F. Bellifemine , S. Galzarano , R. Gravina , G. Fortino , An agent-based signal processing in-node environment for real-time human activity monitoring based on wireless body sensor networks, J. Eng. Appl. Artif. Intell. 24 (7)(2011) 1147–1161 .

[330] W. Li , J. Bao , X. Fu , G. Fortino , G. S , Human postures recognition based on d-s evidence theory and multi-sensor data fusion, in: DPMSS Workshop at the 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, 2012, pp. 912–917 .

[331] G. Fortino , R. Gravina , Fall-mobileguard: a smart real-time fall detection system, in: 10th EAI International Conference on Body Area Networks (Bodynets 2015), IEEE, 2015, pp. 44–50 .

[332] H. Ghasemzadeh , P. Panuccio , S. Trovato , G. Fortino , R. Jafari , Power-aware activity monitoring using distributed wearable sensors, IEEE Trans. Hum.-Mach. Syst. 44 (4) (2014) 537–544 .

[333] K. Van Laerhoven , H.-W. Gellersen , Spine versus porcupine: astudy in distributed wearable activity recognition, in: 8th International Symposium on Wearable Computers (ISWC '04), Washington, DC, USA, 2004, pp. 142–149 .

[334] P. Zappi , C. Lombriser , T. Stiefmeier , E. Farella , D. Roggen , L. Benini , G. Tröster , Activity recognition from on-body sensors: accuracy-power trade-offby dynamic sensor selection, in: Wireless Sensor Networks, 2008, pp. 17–33 .

[335] A.M. Nia , M. Mozaffari-Kermani , S. Sur-Kolay , A. Raghunathan , N.K. Jha , En- ergy

efficient long-term continuous personal health monitoring, Multi-Scale Comput. Syst. IEEE Trans. 1 (2) (2015) 85–98.

[336] T. Choudhury, S. Consolvo, B. Harrison, J. Hightower, A. LaMarca, L. LeGrand, A. Rahimi, A. Rea, G. Bordello, B. Hemingway, P. Klasnja, K. Koscher, J.A. Landay, J. Lester, D. Wyatt, and D. Haehnel. The mobile sensing platform: An embedded activity recognition system. Pervasive Computing, IEEE, 7(2):32–41, 2008.

[337] Waikato environment for knowledge analysis (weka). http://www.cs. waikato.ac.nz/ml/weka.

[338] R. Cilla, M.A. Patricio, A. Berlanga, and J.M. Molina. Creating human activity recognition systems using pareto-based multiobjective optimization. In Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, 2009.

[339] S.I. Gallant. Perceptron-based learning algorithms. IEEE Transactions on Neural Networks, 1(2):179–191, 1990.

[340] The R Project. http://www.r-project.org/.

[341] O. D. Lara and M. A. Labrador. A mobile human activity recognition system. In IEEE Consumer Communications and Networking Conference (CCNC), pages 38 –39, January 2012.

[342] O. D. Lara and M. A. Labrador. A mobile platform for real-time human activity recognition. In IEEE Consumer Communications and Networking Conference (CCNC), pages 667 –671, January 2012.

[343] 2011 Activity Recognition Challenge. www.oportunity-project.eu/ challenge.

[344] N.D. Lane, Ye Xu, Hong Lu, A.T. Campbell, T. Choudhury, and S.B. Eisenman. Exploiting social networks for large-scale human behavior modeling. Pervasive Computing, IEEE, 10(4):45–53, 2011.

[344] C. A. Ronao et S.-B. Cho, « Human activity recognition with smartphone sensors using deep learning neural networks », Expert Systems with Applications, vol. 59, p. 235–244, 2016.

[345] F. J. Ordóñez et D. Roggen, « Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition », Sensors, vol. 16, no 1, p. 115, 2016.

[346] Y. Chen et Y. Xue, « A deep learning approach to human activity recognition based on single accelerometer », in Systems, man, and cybernetics (smc), 2015 ieee international conference on, 2015, p. 1488–1492.

[347] M. Ma, H. Fan, et K. M. Kitani, « Going deeper into first-person activity recognition », arXiv preprint arXiv:1605.03688, 2016.

[348] Y. Kim et T. Moon, « Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks », IEEE geoscience and remote sensing letters, vol. 13, no 1, p. 8–12, 2016.

[349] D. Castro et al., « Predicting daily activities from egocentric images using deep learning », in proceedings of the 2015 ACM International symposium on Wearable Computers, 2015, p. 75–82.

[350] P. Molchanov, S. Gupta, K. Kim, et J. Kautz, « Hand gesture recognition with 3D convolutional neural networks », in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2015, p. 1–7.

[351] L. Mo, F. Li, Y. Zhu, et A. Huang, « Human physical activity recognition based on computer vision with deep learning model », in Instrumentation and Measurement Technology Conference Proceedings (I2MTC), 2016 IEEE International, 2016, p. 1–6.

[352] F. J. O. Morales et D. Roggen, « Deep convolutional feature transfer across mobile

activity recognition domains, sensor modalities and locations », in Proceedings of the 2016 ACM International Symposium on Wearable Computers, 2016, p. 92–99.

[353] R. Bodor, A. Drenner, M. Janssen, P. Schrater, et N. Papanikolopoulos, « Mobile camera positioning to optimize the observability of human activity recognition tasks », in Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on, 2005, p. 1564–1569.

[354] Z. Yu, Y. Rennong, C. Guillaume, et G. Maoguo, « Deep Residual Bidir-LSTM for Human Activity Recognition Using Wearable Sensors », arXiv preprint arXiv:1708.08989, 2017.

[355] G. Taylor, R. Burmeister, Z. Xu, B. Singh, A. Patel, et T. Goldstein, « Training neural networks without gradients: A scalable admm approach », in International Conference on Machine Learning, 2016, p. 2722–2731.

[356] K. E. Powell, P. D. Thompson, C. J. Caspersen, et J. S. Kendrick, « Physical activity and the incidence of coronary heart disease », *Annual review of public health*, vol. 8, n° 1, p. 253–287, 1987.

[357] M. Yu, A. Rhuma, S. M. Naqvi, L. Wang, et J. Chambers, « A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment », *IEEE transactions on information technology in biomedicine*, vol. 16, n° 6, p. 1274–1286, 2012.

[358] N. Noury *et al.*, « Monitoring behavior in home using a smart fall sensor and position sensors », in *Microtechnologies in Medicine and Biology, 1st Annual International, Conference On. 2000*, 2000, p. 607–610.

Annexe A

Contribution de l’apprentissage par renforcement et de la Théorie de Demspter Shafer pour la reconnaissance de l’activité physique

Annexes

L'objectif de cette section est d'utiliser des objets connectés du quotidien pour reconnaitre les activités physiques. L'approche adoptée est l'utilisation de la Théorie de Dempster Shafer pour la prise en compte de l'imperfection des données. Un modèle d'apprentissage par renforcement a été utilisé afin d'évaluer la dépense énergétique induite par les différents activités détectées. L'approche proposée est composée de deux phases distinctes consécutives : la reconnaissance d'activité physique, et la modélisation de la consommation énergétique (s'appuyant sur la phase de reconnaissance d'activité physique). Il a donc été question dans un premier temps de s'intéresser à la reconnaissance d'activité humaine par l'utilisation d'un smartphone.

1. Modélisation de la consommation énergétique

En 1993, Ainsworth *et al.* ont publié un compendium (corpus de connaissances) des activités physiques et leurs équivalents métaboliques associés. L'équivalent métabolique (MET) d'une activité physique se définie comme étant son intensité sur la demande du métabolisme de base (assis au repos, par exemple). 1 MET équivaut à 3.5 ml d'oxygène expiré pendant l'activité par kilogramme et par minute. Le compendium prend la forme d'un tableau d'équivalence entre activités physiques et MET. Ce compendium a été étoffé au fil des années, notamment en 2000 et 2011. L'idée derrière ce compendium est de regrouper un grand nombre d'études sur la consommation énergétique afin de faciliter les comparaisons entre études.

2. Analyse et plan d'action

De nombreux travaux traitent de la détection d'activité physique et de modélisation de la consommation énergétique. Cependant, il n'existe pas à ce jour de travaux modélisant la consommation énergétique en temps-réel avec un smartphone. L'objectif final étant d'aboutir à la modélisation de la consommation énergétique en temps-réel en se basant sur la reconnaissance de l'activité.

Dans un premier temps, l'objectif a été de se concentrer sur la détection d'activité physique élémentaire (marcher, courir, être debout, …). Bien que moins performants en matière de détection d'activité physique, les autres capteurs que possède le smartphone (gyroscope, magnétomètre) pourraient être utilisés eux aussi pour améliorer la détection de l'activité physique. Une approche utilisant de la fusion de données multi-capteurs a été retenues pour traiter cette partie.

Comment modéliser la consommation énergétique une fois les activités physiques élémentaires (marcher, courir, être assis, …) détectées en temps-réel ? Le compendium présenté par Ainsworth *et al.* est très pratique pour répondre à cette question. À chaque activité correspond un équivalent métabolique représentant la consommation énergétique. Cependant, la connaissance de l'activité élémentaire ne suffit pas à estimer l'équivalent métabolique avec le tableau proposé. En effet, pour pouvoir trouver une équivalence dans le tableau, il faut généralement savoir ce qu'est réellement en train de faire la personne. Voici un exemple : est-ce qu' « être assis, en regardant la télé » consomme autant d'énergie que l'activité « être assis, jouant de la batterie » ? La réponse est non, et le compendium nous le dit bien : la première activité a une valeur de 1,3 MET, tandis que la seconde a une valeur de 3,8 MET. Avant de pouvoir modéliser la consommation énergétique, il faut d'abord être capable de comprendre la nature des activités physiques exercées par l'utilisateur. L'approche qui a été retenue est la suivante : en utilisant l'activité physique élémentaire détectée par le smartphone en temps-réel ainsi que d'autres informations telles que l'heure ou la vitesse de l'utilisateur, faire apprendre à une intelligence la nature des activités à partir des habitudes de l'utilisateur.

Enfin, une fois la nature des activités détectée en temps-réel, nous pouvons faire une traduction directe en consommation énergétique grâce au tableau d'équivalence entre activités et équivalents métaboliques proposé par Ainsworth *et al.* dans leur compendium des activités physiques.

3. Reconnaissance des activités physiques élémentaires

Le but de cette partie est de reconnaître à partir d'un smartphone placé dans la poche de l'utilisateur, ses activités physiques élémentaires.

Définissons tout d'abord les activités physiques élémentaires. Elles sont au nombre de huit : courir, être allongé, être assis, être debout, descendre des escaliers, marcher, monter des escaliers, portable abandonné. L'activité élémentaire « portable abandonné » mérite quelques explications. Il s'agit de l'activité pour laquelle le portable n'est pas dans la poche de l'utilisateur mais posé quelque part. Cet état du smartphone important à reconnaître pour la partie suivant qui est la compréhension de la nature des activités.

Pour cela, nous utiliserons un smartphone. Trois capteurs du smartphone sont utilisés : l'accéléromètre, le gyroscope, et le magnétomètre. Le fait d'utiliser plusieurs capteurs nous amène à envisager une approche un peu différente de celles utilisées aujourd'hui en reconnaissance d'activité physique : la fusion de données multi-capteurs. Nous nous sommes penchés notamment sur la théorie de Dempster-Shafer, souvent utilisée en fusion de données.

Tout d'abord, il a été question de créer une application Android de collecte de données. Puis, avant d'appliquer les techniques de fusion de données sur les données collectées, il a fallu les prétraiter. Une fois les données prétraitées, nous avons pu implémenter, tester et améliorer la classification par fusion de données.

4. Collecte de données

Pour chaque activité physique élémentaire, nous cherchons à récolter des données représentant l'évolution dans le temps des valeurs de l'accéléromètre, du gyroscope et du magnétomètre. Pour cela, nous avons développé une application Android sous Android Studio.

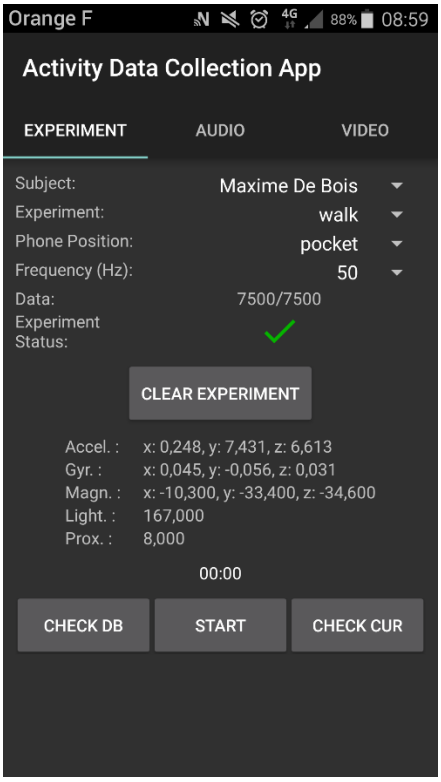


Figure 1 : Application de collecte des activités physiques élémentaires.

L'application est simple :

- L'utilisateur doit tout d'abord créer un profil d'utilisateur.
- Puis il choisit une activité pour laquelle il a envie d'enregistrer des données.
- Il peut ensuite choisir la fréquence d'échantillonnage, la position du téléphone.
- Une fois prêt, il peut lancer la collecte des données qui s'opérera sur 2 minutes 30 ou jusqu'à ce qu'il appuie sur le bouton Stop (le bouton Start se transformant en Stop après le début de la collecte).
- Une fois la collecte arrêtée, les données sont stockées dans une base de données SQLite.

Voici un schéma de la base de données :

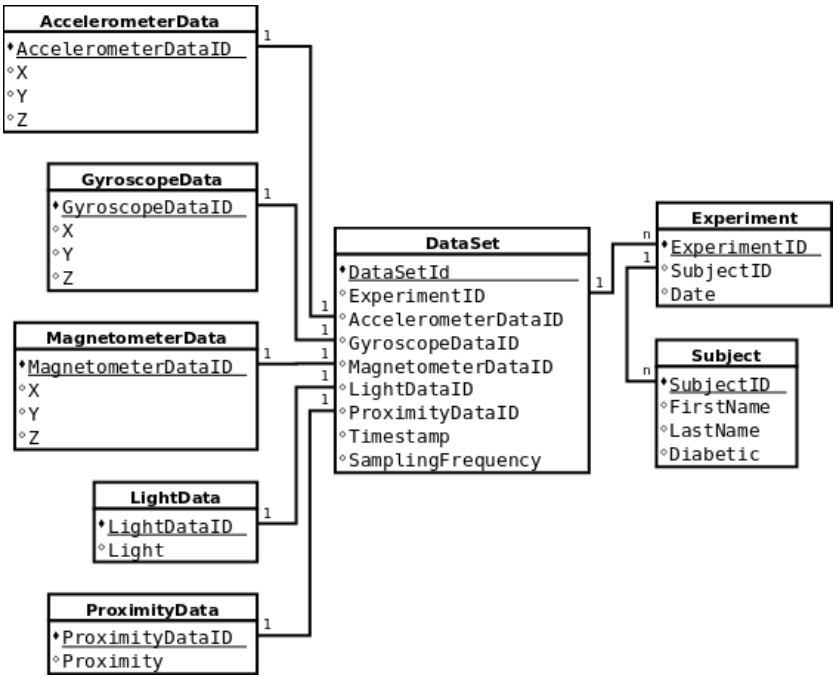


Figure 2 : Schéma de la base de données de collecte.

Pour la collecte en elle-même, elle s'est opérée sur une personne. 5 minutes de données ont été enregistrées par activité physique élémentaire, avec une fréquence d'échantillonnage de 50 Hz et avec le téléphone toujours dans la poche (sauf pour l'activité « portable abandonné »).

Une fois les données collectées, il a fallu les prétraiter afin de les rendre exploitables par la théorie de Dempster-Shafer.

La première phase de prétraitement consiste en le filtrage des données. Le signal doit être filtré entre 0.3 Hz et 20 Hz afin d'une part d'exclure la force gravitationnelle du signal et de ne garder que les mouvements du corps (en dessous de 15 Hz), en appliquant un filtre de Butterworth sur le signal.

Même filtrées, les données ne se sont toujours pas exploitables par la majeure partie des algorithmes de classification. Ainsi pour remédier à ce problème, des techniques de classification plus classiques sont utilisées, mais nécessitent une étape supplémentaire dans le prétraitement des données : l'extraction des descripteurs.

Avant d'extraire les descripteurs, il a fallu tout d'abord regrouper les données de chaque capteur (et chaque axe) en fenêtres glissantes (50%) de 2.56s (pratique pour le passage du signal dans le domaine fréquentiel). Puis nous avons extrait les descripteurs temporels et fréquentiels de chaque fenêtre. Nous en comptons sept dans le domaine temporel (moyenne,

écart type, corrélation, énergie, etc.). Et nous en comptons 4 dans le domaine temporel (entropie, asymétrie fréquentielle, etc.).

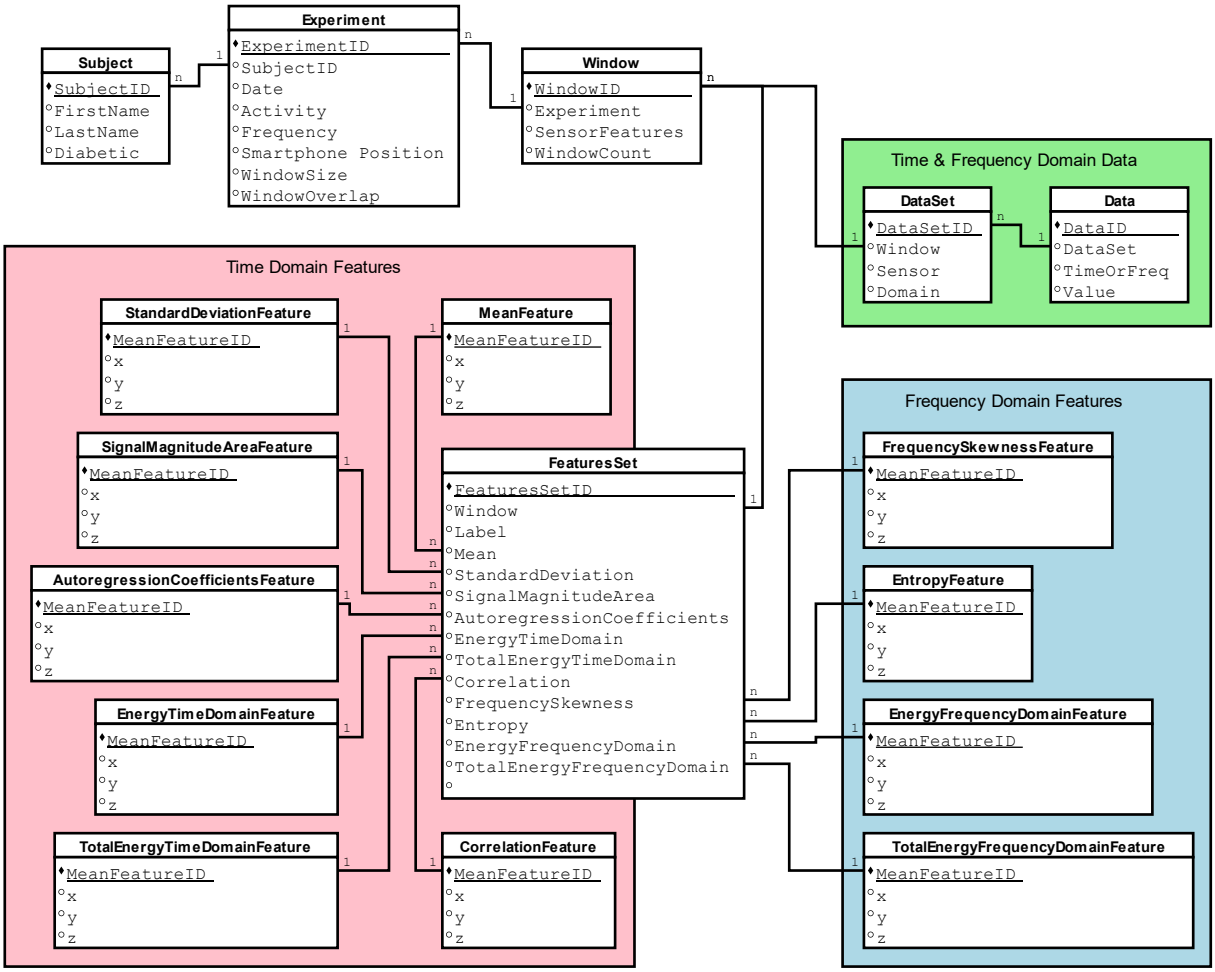


Figure 3 : Schéma base de données des descripteurs.

Voici un exemple de valeurs récoltées pour un descripteur particulier : l'énergie totale (domaine temporel) d'une fenêtre :

Figure 4 : Descripteur "énergie totale" pour les trois capteurs.

Nous pouvons remarquer que l'énergie calculée pour les activités dynamiques (courir, marcher, monter/descendre des escaliers) se distinguent nettement des activités sédentaires (les autres).

Une fois les descripteurs extraits, il a fallu les utiliser pour reconnaître les activités physiques élémentaires.

Ici, nous allons, pour chaque descripteur, faire une prédiction d'activité physique élémentaire. Pour ce faire nous pouvons utiliser un algorithme de classification de Machine Learning. Nous avons choisi les arbres de décisions car cette méthode avait les meilleurs résultats. D'autres ont été testés, comme le k plus proches voisins, les forêts aléatoires, etc.

En fonction du taux de succès de la prédiction pour chaque activité physique élémentaire, il est possible d'évaluer l'incertitude liée à la prédiction. Nous appelons la prédiction mitigée d'incertitude d'un descripteur une *masse*.

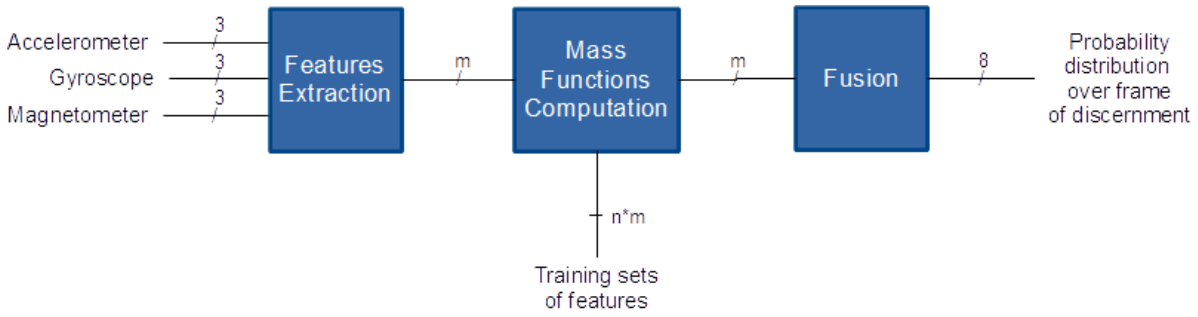


Figure 5 : Représentation graphique de la reconnaissance d'activité élémentaire.

Les masses de chaque descripteur peuvent être combinées grâce à la règle de combinaison de deux masses de la théorie de Dempster-Shafer, que l'on applique pour chaque activité physique élémentaire.

$$m_{1,2}(A) = m_1 \oplus \mathrm{m}_2(A) = \begin{cases} k^{-1} * \sum_{B \cap C = A} m_1(B) * m_2(C), & si\ A \neq \emptyset \\ 0, & si\ A = \emptyset \end{cases}$$

Équation 1 – Règle de combinaison de Dempser-Shafer

Avec k dénotant le nombre de conflits entre croyances de base :

$$k = \sum_{B \cap C \neq \emptyset} m_1(B) * m_2(C)$$

Équation 2 – Nombre de conflits entre croyances de base

Le résultat final (qui prend la forme d'une distribution de probabilité sur l'espace des activités physiques élémentaires) est obtenu en appliquant la règle de combinaison autant de fois que l'on a de descripteurs, moins une.

La bibliothèque Sklearn a été utilisée pour la partie calcul des masses avec les arbres de décisions et la théorie de Dempster-Shafer a été implémentée directement. Voici les résultats obtenus pour la reconnaissance d'activité physique élémentaire.

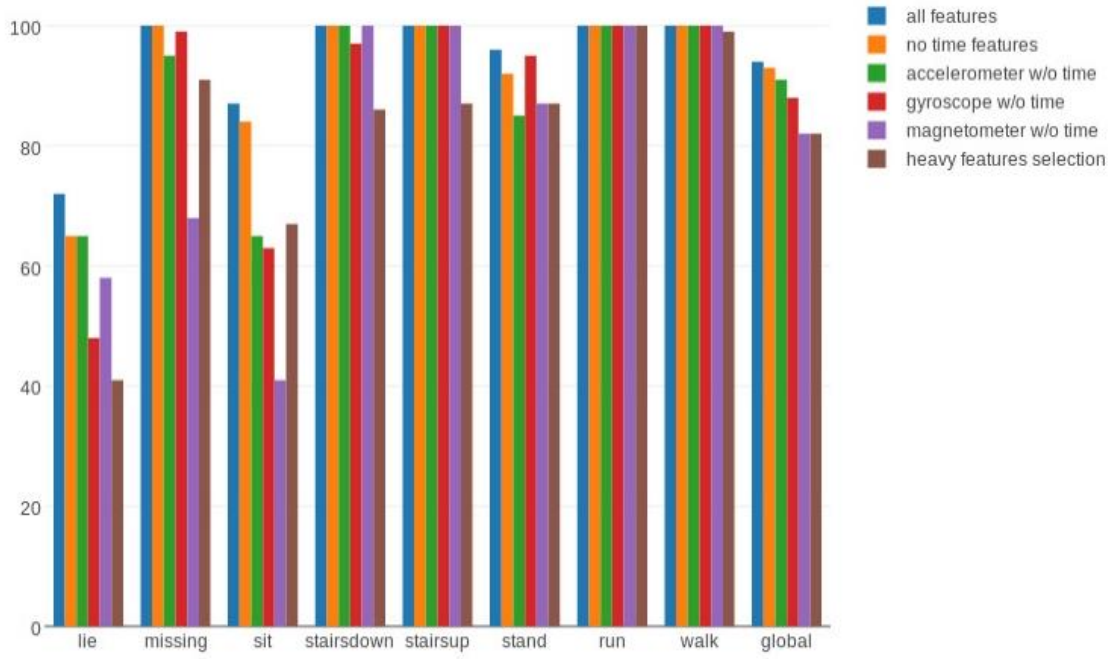


Figure 6 : Résultats de la reconnaissance d'activité physiques élémentaires.

La figure 6 représente le taux de réussite de la prédiction par activité physique élémentaire. Afin d'augmenter le taux de réussite, certains descripteurs ont été sélectionnés (les plus performants). Plusieurs choses marquent notre attention dans les résultats présentés ci-dessus.Tout d'abord, les résultats sont globalement très bons, avec 95% de succès globalement lorsque tous les descripteurs sont pris en compte. Puis, la sélection des descripteurs n'a pas amélioré les prédictions. Ce n'était pas le cas lors de tests non présentés ici (d'où la pertinence du test). Finalement, ceci est cohérent avec la théorie de Dempster-Shafer qui veut que plus on accumule de données, de pièces de conviction, plus la certitude d'une prédiction est élevée. Ainsi plus on a de descripteurs sur lesquels baser notre prédiction, plus le taux de réussite doit être élevé.

Enfin, les activités physiques élémentaires dynamiques (marcher, courir, monter/descendre des escaliers) ont un taux de réussite bien supérieur aux activités sédentaires (être allongé, assis, debout). Cela s'explique par la quasi absence de mouvements dans les activités sédentaires impliquant une plus grande difficulté dans la reconnaissance.

5. Reconnaissance de la nature des activités physiques

Comme expliqué précédemment, dans l'optique d'utiliser le tableau d'équivalence entre activités et équivalents métaboliques présenté dans le compendium des activités physiques, il faut être capable de comprendre la nature des activités élémentaires détectées dans la partie précédente. Par nature d'activité on entend par exemple « assis, en mangeant & parlant »

plutôt que juste « assis ».L'objectif est le suivant : en détectant l'activité physique élémentaire de l'utilisateur tout au long de la journée, et en ajoutant à cela des informations comme le temps et la vitesse de l'utilisateur, reconnaître la nature des activités de l'utilisateur.

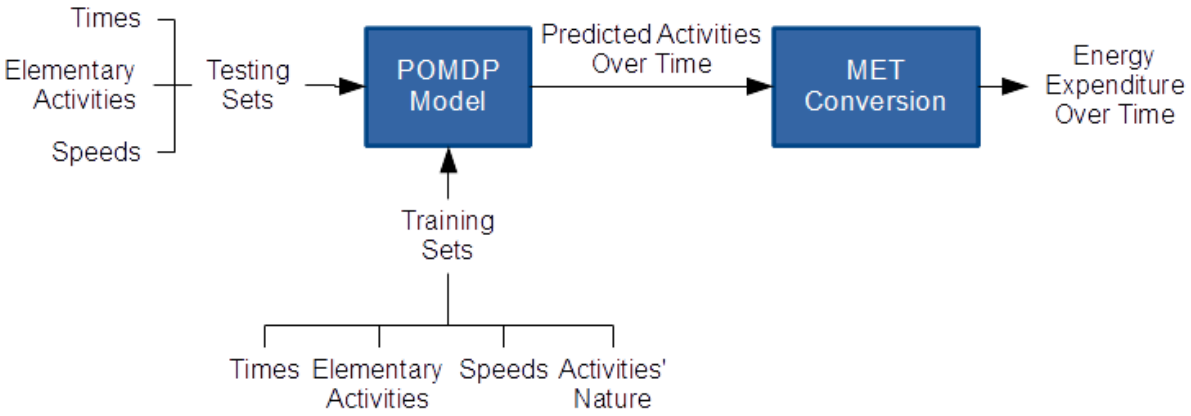


Figure 7 : Diagramme de la modélisation de la consommation énergétique.

L'approche identifiée est celle de l'apprentissage par renforcement, en particulier celle du nom de *Partially Observable Markov Decision Process*. L'agent (l'intelligence) navigue dans un environnement dont il ne perçoit que des informations partielles (activité physique élémentaire, temps, vitesse de l'utilisateur) et a pour but de deviner à instant t la nature d'activité correspondante. À chaque action, à chaque prédiction, en fonction de la réussite, l'agent reçoit une récompense. Cette récompense lui permet d'apprendre de l'expérience, et de mieux prédire par la suite.

Afin d'avoir une prédiction de qualité, il faut avoir beaucoup de données d'entraînement à donner à l'intelligence, les données ont été simulées plutôt que collectées.

Comme nous pouvons le voir sur la **Figure 7**, nous avons besoin en entrée de notre algorithme des données telles que les activités physiques élémentaires, le temps, la vitesse de l'utilisateur et la nature de l'activité qu'il est en train de réaliser.

Pour cela, nous avons observé le déroulement des journées d'une personne (jours de semaine). À partir de cela, nous avons construit des scenarios, des enchaînements d'activités, qu'elle peut suivre sur une journée de travail. Pour chaque activité, on peut associer une nature d'activité que l'on peut trouver dans le compendium des activités physiques. Voici l'un des scenarios :

Tableau 1 : Activités utilisées dans la simulation de données.

Activité	Durée (min)	Nature d'activité	Équivalent Métabolique (MET)
Lever à 8h	/	/	/
Petit déjeuner	10	13030	1.5
Se laver	30	13040	2.0
Se préparer	10	9070	1.8
Marcher jusqu'au bus	9	17190	3.5
Prendre le bus	8	16016	1.3
Marcher jusqu'au LIMSI	2	17190	3.5
Monter au 2e étage	1	17133	4.0
Travailler	1	11580	1.5
Aller aux toilettes	5	17151	2.0
Travailler	150	11580	1.5
Descendre au RDC	1	17070	3.5
Marcher jusqu'au restaurant	5	17190	3.5
Manger	30	13030	1.5
Marcher jusqu'au LIMSI	5	17190	3.5
Prendre un café	1	7040	1.3
Monter au 2e étage	1	17133	4.0
Travailler	270	11580	1.5
Descendre au RDC	1	17070	3.5
Marcher jusqu'au bus	2	17190	3.5
Prendre le bus	8	16016	1.3
Marcher jusqu'à la maison	9	17190	3.5
Goûter	5	13030	1.5
Jouer de la guitare	90	10074	2.0
Préparer à manger	10	5035	3.3
Manger	10	13030	1.5
Jouer aux jeux-vidéos	120	9045	1
Regarder un film	120	7025	1.5
Se préparer pour aller dormir	10	9070	1.8
Dormir	480	7030	1.0

Pour générer des données avec un tel scénario, il suffit de choisir l'activité physique élémentaire pour chaque activité (avec une probabilité d'avoir une « mauvaise » activité

élémentaire en fonction du taux de détection des activités élémentaires), un temps (en minutes), et une vitesse (en fonction de l'allure de la marche, course, ou de la vitesse du bus).

En agrémentant cela d'aléatoire, et en créant plusieurs autres scénarios comparables à celui-ci avec quelques variations dans les activités, nous pouvons générer suffisamment de données pour faire apprendre à notre intelligence à détecter la nature des activités.

Au total, dix scénarios (dix journées type) ont été simulés. Neuf scénarios servent à entraîner le modèle, et le dernier à le tester.

6. Apprentissage par renforcement – le modèle pomdp

Comme mentionné précédemment, le modèle d'apprentissage par renforcement du nom de *Partially Observable Markov Decision Process* (POMDP) a été choisi comme algorithme de reconnaissance de la nature des activités de l'utilisateur en temps-réel. Ce modèle répond à des problématiques particulières. Contrairement à un modèle d'apprentissage par renforcement classique, l'environnement dans lequel navigue l'agent n'est pas connu de celui-ci. L'agent n'a qu'une représentation partielle de l'environnement et doit avec cette information inférer le réel état de l'environnement et choisir une action.

Nous appelons la croyance de l'état de l'environnement l'*état de croyance* (*belief state* en anglais). L'état de croyance est une distribution de probabilités sur l'ensemble des états possibles de l'environnement. Dans notre situation, les états de l'environnement sont les activités (entendre « natures d'activité ») de l'utilisateur. Au rythme de la journée de l'utilisateur, l'environnement de l'agent va changer, les états vont se succéder les uns après les autres.

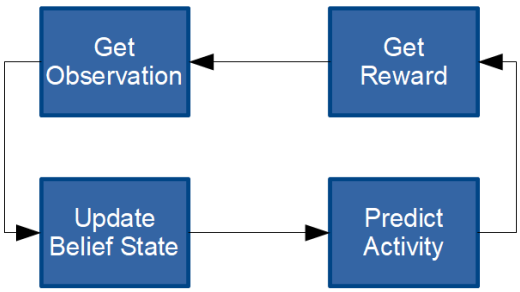


Figure 8 : Cycle de vie de l'agent dans un modèle type POMDP.

Au fur et à mesure des changements de l'environnement, l'agent doit prédire la nature de l'activité en cours. Pour cela, il se base sur les observations qu'il fait de l'environnement. Les observations sont composées du temps, de l'activité physique élémentaire et de la vitesse de l'utilisateur.

$$observation = F(temps, activité\ physique\ élémentaire, vitesse)$$

Équation 3 – Définition d'une observation

Le changement de ses observations (qui changent au même rythme que les états de l'environnement que l'agent ne connait pas) fait évoluer son état de croyance. En fonction de l'état de croyance, soit de sa croyance de l'état actuel de l'environnement et des récompenses associées à chacune de ses prédictions, il prédit la nature de l'activité en cours. En fonction de sa prédiction il reçoit donc une récompense positive ou négative.

Voici la formule permettant de mettre à jour l'état de croyance après une transition (l'agent exécutant une action et recevant une observation). Celle-ci s'applique pour tous les états possibles de l'environnement.

$$b_o^a(s') = \Pr(\mathrm{s'|b,a,o}) = \frac{\mathrm{O(o|s')} \sum_{s \in S} P(s'|s,a) b(s)}{\sum_{s \in S} \sum_{s'' \in S} O(o|s'') P(s''|s,a) b(s)}$$

Équation 4 - Mise à jour de l'état de croyance

Dans l'équation ci-dessus, $P(s'|s,a)$ représente la probabilité de transition de l'environnement d'un état s avec une action a vers l'état s'. Quant à lui, $O(o|s')$ représente la probabilité d'avoir l'observation o sachant que l'état de l'environnement est s'. Afin de calculer le nouvel état de croyance, il faut tout d'abord calculer ces deux probabilités. La manière de calculer ces probabilités influera sur la réussite de l'agent à apercevoir les changements d'états de l'environnement et ainsi prédire la bonne activité.

Une fois l'état de croyance mis à jour, l'agent va calculer la *récompense attendue* (*expected reward* en anglais) pour chaque action. L'équation ci-dessous définie cette récompense attendue, sous la forme d'une *fonction d'utilité*, V.

$$V = \max_a \left[\sum_{s \in S} r(s,a) b(s) \right]$$

Équation 5 – Fonction d'utilité

Dans l'expression de la fonction d'utilité, $r(s,a)$ représente la récompense que l'agent récupère s'il effectue l'action a à l'état s. Comme on peut le voir, la fonction d'utilité ne dépend que des récompenses (connues par l'agent) et de son état de croyance à l'instant t. À chaque étape, l'action que choisit l'agent est celle qui maximise la fonction d'utilité (dans ce cas précis,

ce n'est pas forcément vrai). Nous appelons l'association entre état de croyance et action la *politique*, $\pi(b)$. Ainsi nous avons l'équation suivante :

$$\pi(b) = \mathrm{argmax}_a \left[\sum_{s \in S} r(s,a) b(s) \right]$$

Comme mentionné précédemment, le plus grand enjeu du modèle est le calcul des probabilités de transitions d'états et d'observation. Ces probabilités peuvent être calculées directement en comptant les occurrences d'apparition des différents couples. Cependant cela s'est révélé assez inefficace. Cela s'explique par le fait que les données liées aux observations (dans le calcul de $O(o|s')$) peuvent être très clairsemées, notamment pour les activités qui n'arrivent que peu de fois pendant la journée (exemple : aller aux toilettes). Cependant, il paraitrait logique, de par la proximité temporelle d'une observation avec une autre, que la probabilité ne soit pas nulle. Par exemple, si la probabilité d'aller aux toilettes à $t = 50$ est de 50%, mesuré par l'expérience, il ne serait pas juste que la probabilité à $t = 49$ ou $t = 51$ soit de zéro. C'est à partir de ces observations que nous avons dégagé une méthodologie pour palier au caractère clairsemé des données. L'idée est de lisser les probabilités obtenues de la manière suivante : pour toute occurrence d'une observation pour une activité donnée, on simule une occurrence d'importance moindre sur les voisins temporels de l'observation enregistrée. Cette occurrence moindre est prise sur une fonction Gaussienne centrée en l'occurrence enregistrée. Il est possible de faire varier l'impact temporel des occurrences enregistrées en variant la variance de la fonction gaussienne.

7. Résultats

Dans cette partie, nous présentons les divers résultats dans la reconnaissance des natures d'activités obtenus après apprentissage de notre intelligence. Sauf si spécifié, les résultats montrés ont été obtenus par validation croisée (*leave-one-out cross-validation*) sur les dix scénarios de données simulés.

Tout d'abord, intéressons-nous au lissage de probabilités opérées sur les observations. Les deux figures ci-dessous représentent respectivement le comptage des occurrences des observations sur l'axe temporel et les probabilités résultantes du comptage, les deux pour l'activité « aller aux toilettes ». Nous avons en rouge les points, valeurs initiales. Et nous avons en bleu les points, valeurs après lissage.

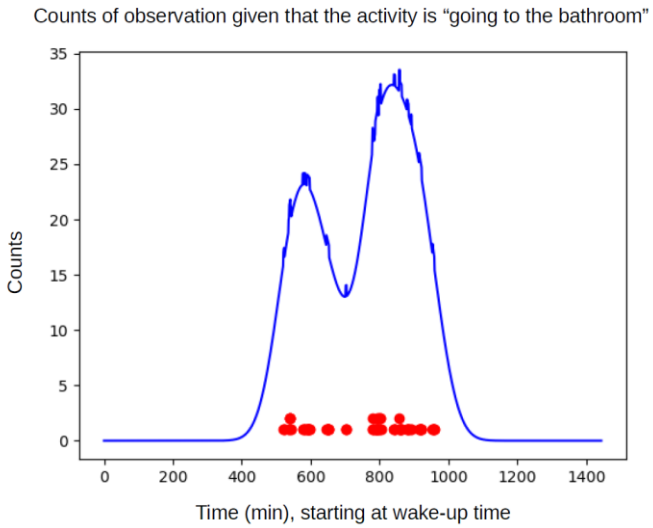


Figure 9 : Comptage des occurrences des observations pour l'activité "aller aux toilettes".

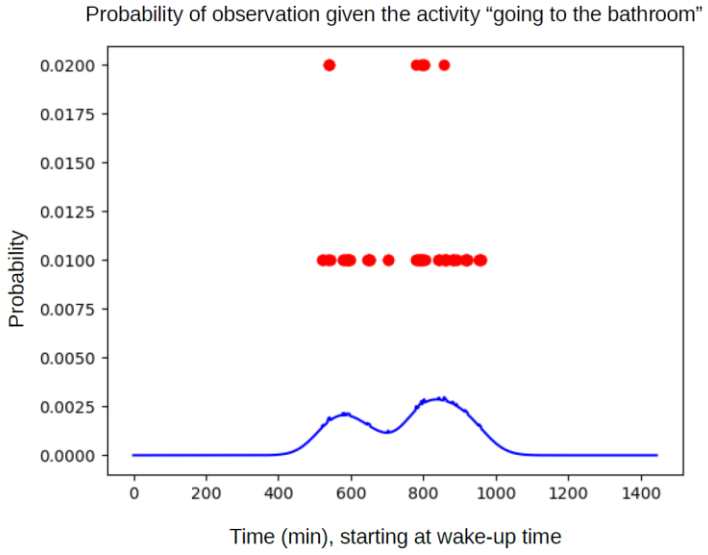


Figure 10 : Probabilités résultantes des observations pour l'activité "aller aux toilettes".

Nous pouvons voir que, en effet, nous avons bien réussi à lisser les probabilités. Reste à voir l’impact de la méthode sur la prédiction de la nature des activités.

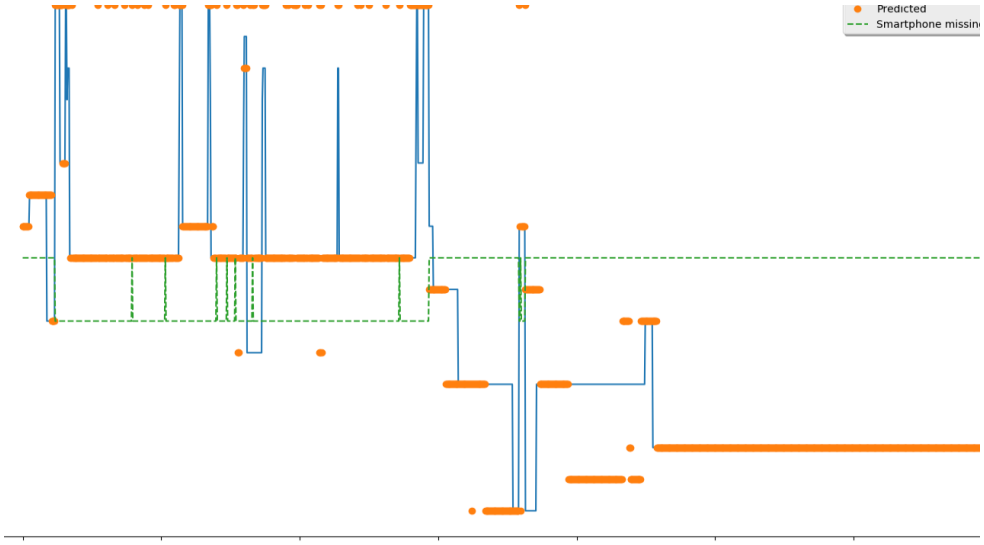


Figure 11 : Prédiction VS réalité en fonction du temps (sans lissage).

Le graphique ci-dessus représente les résultats de la prédiction du scénario 1 quand les autres scénarios sont pris en apprentissage. En bleu, nous avons la réelle évolution des natures d'activités, tandis qu'en orange nous avons la prédiction de l'agent. En vert est représenté l'état du smartphone, abandonné ou non.

Nous pouvons donc voir sur le graphique que la prédiction n'est pas mauvaise (75% en moyenne temporelle, 45% en moyenne sur les activités). Mais il y a tout de même un grand nombre de mauvaises prédictions. Voyons si le lissage a pu régler le problème.

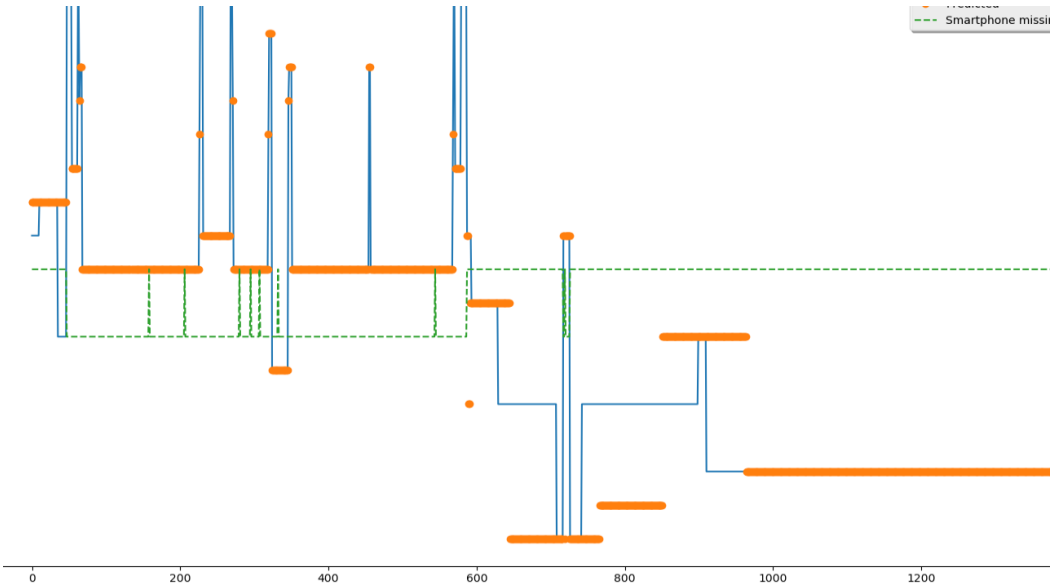


Figure 12 : Prédiction VS réalité en fonction du temps (avec lissage).

Le graphique ci-dessus représente donc la même chose que le graphique précédent, mais cette fois ci, après avoir lissé les probabilités liées aux observations. Avec 75% temporellement, et 75% en moyenne sur les activités, nous pouvons dire qu'il s'agit là d'un

succès. Les « petites activités », les activités avec peu d'occurrences sont bien mieux détectées, ce qui était le but du lissage.

Les résultats ont pu être d'avantage amélioré en modifiant les récompenses données à l'agent. Dans un souci de clarté, ces résultats ne sont pas présentés ici.

8. Modélisation de la consommation énergétique

Une fois les natures des activités détectées en temps-réel, il est maintenant question de traduire cette information en consommation énergétique.

Pour rappel, nous avions choisi d'utiliser le compendium d'activités physiques publié en 1993 (et mis à jour par la suite) par Ainsworth *et al.*. Ce compendium liste un grand nombre d'activités et leur attribue une valeur, l'équivalant métabolique, exprimant la consommation énergétique de l'activité en question.

Pour cela, comme brièvement expliqué précédemment, rien de plus simple : on utilise le tableau d'équivalence entre activités en équivalents métaboliques disponible dans le compendium des activités physiques.

Un MET s'exprimant en $kcal.kg^{-1}.h^{-1}$, il est possible de directement traduire les prédictions d'activités en prédictions de consommation énergétiques (s'exprimant en MET donc).

Comme pour la reconnaissance des natures des activités physiques, le graphique ci-dessous est tiré du premier scénario en ayant pris les neuf autres scénarios en données d'apprentissage.

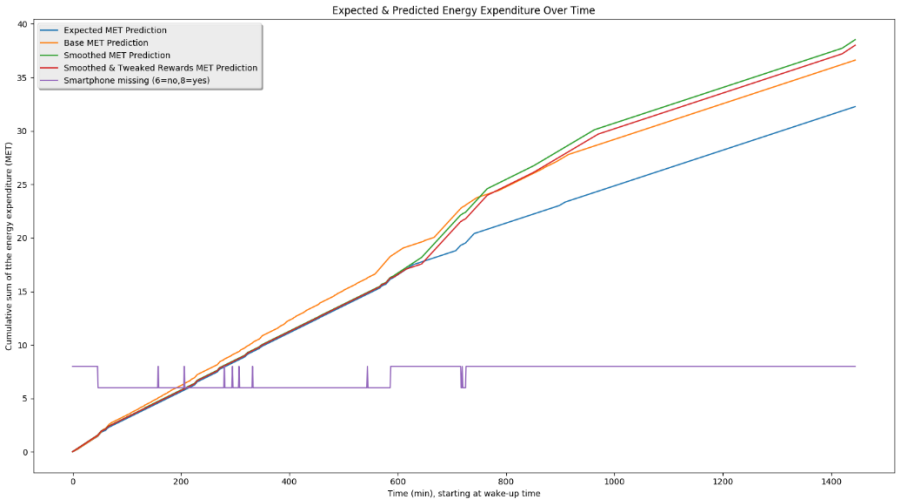


Figure 13 : Prédictions et réalité de la consommation énergétique.

Ce graphique représente les différentes prédictions et de la réalité de la consommation en fonction du temps. Par réalité, il faut comprendre : ce qui était attendu en termes de prédiction.

Il exprime la somme cumulative des équivalents métaboliques des activités faites tout au long de la journée. Nous pouvons remarquer plusieurs choses. Tout d'abord, à la fin de la journée, la meilleure prédiction d'équivalents métaboliques revient au modèle sans lissage. Cependant, nous pouvons remarquer que le lissage des données améliore nettement la prédiction entre $t = 50$ et $t = 600$. Il se trouve que pendant cette période le smartphone n'est presque jamais détecté comme étant « abandonné » ! Ce que nous pouvons voir ici reflète finalement l'incapacité de l'agent à prédire la bonne activité si on lui enlève une information capitale : l'activité physique élémentaire. Ci-dessous le compendium des activités physiques.

CODE	METS	MAJOR HEADING	SPECIFIC ACTIVITIES
09005	*2.5*	miscellaneous	casino gambling, standing
09010	1.5	miscellaneous	card playing, sitting
09013	1.5	miscellaneous	chess game, sitting
09015	1.5	miscellaneous	copying documents, standing
09020	1.8	miscellaneous	drawing, writing, painting, standing
09025	1.0	miscellaneous	laughing, sitting
09030	1.3	miscellaneous	sitting, reading, book, newspaper, etc.
09040	1.3	miscellaneous	sitting, writing, desk work, typing
09045	1.0	miscellaneous	sitting, playing traditional video game, computer game
09050	*1.8*	miscellaneous	standing, talking in person, on the phone, computer, or text messaging, light effort
09055	*1.5*	miscellaneous	sitting, talking in person, on the phone, computer, or text messaging, light effort
09060	1.3	miscellaneous	sitting, studying, general, including reading and/or writing, light effort
09065	1.8	miscellaneous	sitting, in class, general, including note-taking or class discussion
09070	*1.8*	miscellaneous	standing, reading
09071	2.5	miscellaneous	standing, miscellaneous
09075	1.8	miscellaneous	sitting, arts and crafts, carving wood, weaving, spinning wool, light effort
09080	*3.0*	miscellaneous	sitting, arts and crafts, carving wood, weaving, spinning wool, moderate effort
09085	2.5	miscellaneous	standing, arts and crafts, sand painting, carving, weaving, light effort
09090	*3.3*	miscellaneous	standing, arts and crafts, sand painting, carving, weaving, moderate effort
09095	*3.5*	miscellaneous	standing, arts and crafts, sand painting, carving, weaving, vigorous effort
09100	1.8	miscellaneous	retreat/family reunion activities involving sitting, relaxing, talking, eating
09101	3.0	miscellaneous	retreat/family reunion activities involving playing games with children
09105	*2.0*	miscellaneous	touring/traveling/vacation involving riding in a vehicle
09106	*3.5*	miscellaneous	touring/traveling/vacation involving walking
09110	*2.5*	miscellaneous	camping involving standing, walking, sitting, light-to-moderate effort
09115	*1.5*	miscellaneous	sitting at a sporting event, spectator
10010	1.8	music playing	accordion, sitting
10020	2.3	music playing	cello, sitting
10030	2.3	music playing	conducting orchestra, standing
10035	2.5	music playing	double bass, standing
10040	3.8	music playing	drums, sitting
10045	3.0	music playing	drumming (e.g., bongo, conga, benbe), moderate, sitting
10050	2.0	music playing	flute, sitting
10060	1.8	music playing	horn, standing
10070	2.3	music playing	piano, sitting
10074	2.0	music playing	playing musical instruments, general
10077	2.0	music playing	organ, sitting
10080	3.5	music playing	trombone, standing
10090	1.8	music playing	trumpet, standing
10100	2.5	music playing	violin, sitting
10110	1.8	music playing	woodwind, sitting
10120	*2.0*	music playing	guitar, classical, folk, sitting
10125	*3.0*	music playing	guitar, rock and roll band, standing
10130	*4.0*	music playing	marching band, baton twirling, walking, moderate pace, general
10131	5.5	music playing	marching band, playing an instrument, walking, brisk pace, general
10135	*3.5*	music playing	marching band, drum major, walking
11003	2.3	occupation	active workstation, treadmill desk, walking
11006	3.0	occupation	airline flight attendant
11010	*4.0*	occupation	bakery, general, moderate effort
11015	2.0	occupation	bakery, light effort
11020	2.3	occupation	bookbinding

Activity Code	Description	Metabolic Equivalent (kcal.h-1.kg-1)

5035	kitchen activity, general, (e.g., cooking, washing dishes, cleaning up), moderate effort	3.3
7025	sitting, listening to music (not talking or reading) or watching a movie in a theater	1.5
7030	sleeping	1.0
7040	standing quietly, standing in a line	1.3
9045	sitting, playing traditional video game, computer game	1.0
9055	sitting, talking in person, on the phone, computer, or text messaging, light effort	1.5
9070	standing, reading	1.8
10074	playing musical instruments, general	2.0
11580	sitting tasks, light effort (e.g., office work, chemistry lab work, computer work, light assembly repair reading, desk work)	1.5
13030	eating, sitting	1.5
13040	grooming, washing hands, shaving, brushing teeth, putting on make-up, sitting or standing	2.0
16016	riding in a bus or train	1.3
17070	descending stairs	3.5
17133	stair climbing, slow pace	4.0
17151	walking, less than 2.0 mph, level, strolling, very slow	2.0
17152	walking, 2.0 mph, level, slow pace, firm surface	2.0
17190	walking, 2.8 to 3.2 mph, level, moderate pace, firm surface	3.5

9. Activités du compendium utilisées pour la simulation de données

Activity Code	Description	Metabolic Equivalent (kcal.h-1.kg-1)
5035	kitchen activity, general, (e.g., cooking, washing dishes, cleaning up), moderate effort	3.3
7025	sitting, listening to music (not talking or reading) or watching a movie in a theater	1.5
7030	sleeping	1.0
7040	standing quietly, standing in a line	1.3
9045	sitting, playing traditional video game, computer game	1.0
9055	sitting, talking in person, on the phone, computer, or text messaging, light effort	1.5
9070	standing, reading	1.8
10074	playing musical instruments, general	2.0
11580	sitting tasks, light effort (e.g., office work, chemistry lab work, computer work, light assembly repair reading, desk work)	1.5
13030	eating, sitting	1.5
13040	grooming, washing hands, shaving, brushing teeth, putting on make-up, sitting or standing	2.0
16016	riding in a bus or train	1.3
17070	descending stairs	3.5
17133	stair climbing, slow pace	4.0
17151	walking, less than 2.0 mph, level, strolling, very slow	2.0
17152	walking, 2.0 mph, level, slow pace, firm surface	2.0
17190	walking, 2.8 to 3.2 mph, level, moderate pace, firm surface	3.5

Annexe B

Mise en œuvre d'un perceptron multicouches pour la reconnaissance de l'activité

Annexes

Dans cette séction, nous décrivons la mise en œuvre d'une méthode de classification non linéaire de type perceptron multicouche. Le but étant de de montrer l'impact de ce type d'approches sur la reconnaissance de l'activité physique.

1. Mise en œuvre d'une classification non linéaire

Nous avons montré dans le chapitre 2, l'interet d'utiliser des méthodes linéaires. Maintenant que nous avons utilisé un classifieur linéaire nous allons mettre en œuvre une méthode de classification non linéaire à l'aide d'un réseau de neurones à savoir un perceptron multi couches. Nous avons à notre disposition des descripteurs de dimension 144*3, ce qui est énorme pour estimer les paramètres d'un MLP. Pour remédier à ce problème, nous proposons de réduire la dimension des données en utilisons une ACP.

2. Réduction de la dimension des données

A l'aide de la technique d'analyse en composantes principales 'ACP' nous allons réduire la dimension des données et passer de 144 à 20 puis à 5 puis à 2.

Cette manière de faire, c'est-à-dire, réduire de 144 à 20 puis à 5 puis à 2 a été choisi aléatoirement dans le but de ne pas compliquer les calculs et multiplier les différents états.

3. Détermination de la taille de la couche cachée

Nous allons mettre en œuvre une méthode de détermination du nombre de neurones sur la couche cachée. Nous avons extrait un ensemble d'apprentissage et de test (50:50), Et grâce à cet ensemble extrait nous avons déroulé l'algorithme d'apprentissage du MLP pour un nombre de neurones de la couche cachée différent (m $\in$\{1, 3, 5, 7,9,…, 49\}). Nous avons calculé les performances en généralisation et on a retenu le tableau qui suit:

Tableau 1 : Nombre de neurones sur la couche cachée en fonction de la dimension des descripteurs.

Dimension / Nb neurones couche cachée	2	5	20
1	0.7374	0.91	0.7305
3	0.6279	0.912	0.6363
5	0.5763	0.9221	0.5795
7	0.5500	**0.931**	0.5511
9	0.5542	0.4842	0.5021
11	0.5521	0.4937	0.5053
49	0.5353	0.4000	0.3568

Nous remarquons qu'avec 7 neurones sur la couche cachée, nous avons une performance de 93.1% correspondant à une dimension des descripteurs de 5, qui sera retenue.

4. Résultats réduction de la dimension

L'utilisation de l'ACP pour réduire la dimension des descripteurs, a donné des performances de l'ordre de 93.1% contrairement à la SVM qui est de l'ordre de 98,3% d'où l'intérêt d'utiliser la SVM plutôt qu'un MLP.

Annexe C

Détails techniques sur le système de reconnaissance de l’attention envers la TV

Annexes

Dans cette partie, nous présentons les détails de l'implémentation des

1. Reconnaissance du locuteur pour la détection de l'attention envers la TV

1.1. Coefficients cepstraux de fréquence de Mel (MFCC)

MFCC est la méthode de base dans ce domaine. Elle est basée sur un axe de fréquence à échelle perceptuelle. L'échelle Mel offre une bonne résolution de fréquence aux basses fréquences et opère inversement aux fréquences plus élevées. Cette échelle est basée exactement sur le même principe que l'oreille humaine.

1.2. Méthode prédictive linéaire spectrale-perceptuelle relative (RASTA-PLP)

Un filtre passe-bande spécial est ajouté à chaque sous-bande de fréquence dans la technique PLP traditionnelle, afin de réduire les variations à court terme du bruit et d'éliminer tout décalage constant dans le canal vocal (ou la source vocale). L'amplitude spectrale est modifiée à l'aide d'une transformation de compression non linéaire. Ensuite, la trajectoire temporelle de chaque composante spectrale transformée est filtrée par un filtre passe-bande afin de transformer le discours filtré. L'étape suivante consiste à simuler la loi de puissance de l'audition et enfin nous calculons un modèle multipolaire du spectre tel que dans la méthode conventionnelle PLP.

1.3. Combinaison des méthodes MCFF et RASTA-PLP

Afin de concevoir une nouvelle extraction de caractéristiques robuste, nous avons utilisé un algorithme hybride basé sur une combinaison des méthodes d'extraction de caractéristiques décrites précédemment MFCC et RASTA-PLP. Chacune de ces méthodes est d'abord utilisée séparément, puis concaténée afin d'extraire de nouveaux vecteurs de caractéristiques.

1.4. Prétraitement

Le prétraitement des signaux de la parole est une étape très importante dans la conception de tout système de reconnaissance de la parole / du locuteur, car il a une influence directe sur sa robustesse et son efficacité, l'avantage du prétraitement de préaccentuation qui rend l'affichage uniforme de l'énergie sur toutes les fréquences du signal vocal. Cette nouvelle distribution de fréquence conduira donc à une amélioration de la caractérisation du signal de parole et à une augmentation du taux d'identification du système de reconnaissance de locuteur. Cela justifie l'utilisation de cette technique.

La préaccentuation du signal de la parole a l'intention de donner plus d'énergie aux hautes fréquences. Ceci est obtenu en utilisant un filtre passe-haut ayant la fonction de transfert représentée en (1):

$$H(z) = 1 - 0.95.z^{-1} \qquad (1)$$

Nous avons construit une base de données contenant un corpus correspondant à une population de 33 locuteurs (21 hommes et 12 femmes). Pour chaque locuteur, nous avons acquis deux enregistrements: l'un d'environ 20 secondes pour l'entrainement de la base de données de référence, l'autre d'environ 10 secondes qui servira plus tard pour la phase de test. Tous les enregistrements ont été acquis au format « .wav » avec une fréquence d'échantillonnage de quantification monophonique de 16 kHz et 16 bits en utilisant le logiciel Wavesurfer. Les extracteurs de caractéristiques qui ont été considérés dans cette étude sont le PLP, le MFCC, le RASTA-PLP, MFCC & RASTA-PLP ainsi que la combinaison des 3 descripteurs.

2. Conditions experiemntales de l'étude et résultats de classification

Ensemble de descripteurs	MFCC, PLP, RASTA-PLP, MFCC & RASTA-PLP
Nombre de coefficents dans un vecteur de descripteurs	30 (10 statique+10 delta+10 delta-delta) MFCC et 10 RASTA-PLP 40 (30+10) pour MFCC et RASTA-PLP
Taille de la fenetre	20 ms
Fréquence echantillonnage	16 kHz
Ensemble d'apprentissage	21 messages
Ensemble de test	12 messages

Tableau 3 : résultats de reconnaissance des locuteurs

	HMM	**SVM**	**NN**
La moyanne	**75.33%**	**74.32%**	**70.1%**

Les locuteurs ont été détéctés avec une précision moyenne de 75.33% en utilisant un modèle HMM.

Annexe D

Calcul et approximation des fonctions de masse

Annexes

Dans cette séction, nous présentons quelques résultats de calcul des fonctions de masses en utilisant les différentes approches abordées dans le chapitre 4. Puis, nous proposons d'illustrer l'execussion de l'algorithme d'approximation proposé. Enfin, nous aborderons le dilemme « portabilité vs utilisabilité ».

1. Résultats de calcul des fonctions de masses

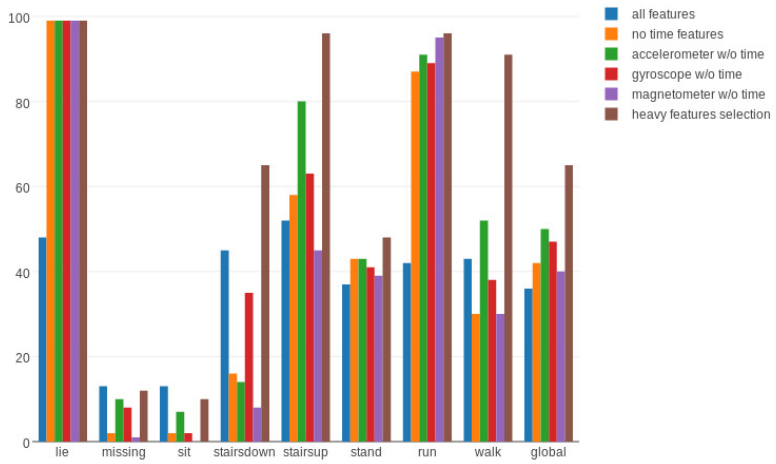


Figure 1 : Prediction Success Rate by Activity and Features Selection for Fixed-range Cells

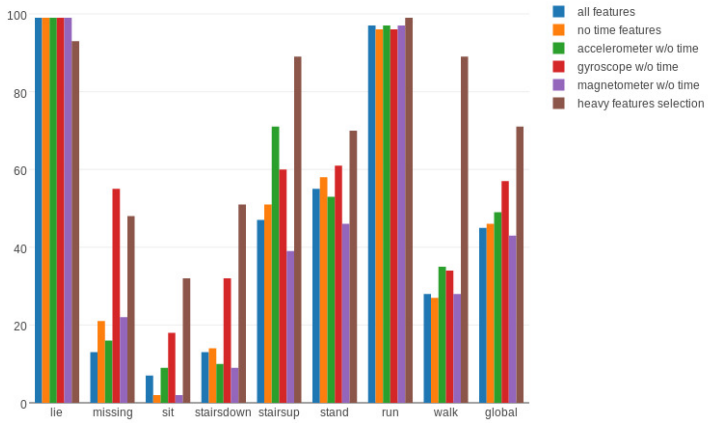


Figure 2 : Prediction Success Rate by Activity and Features Selection for Fixed-size Cells

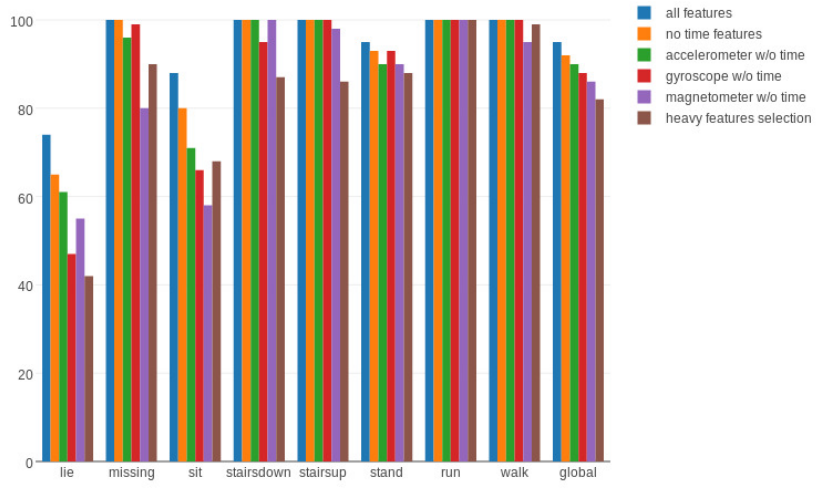


Figure 3 : *Prediction Success Rate by Activity and Features Selection for the kNN Classifier*

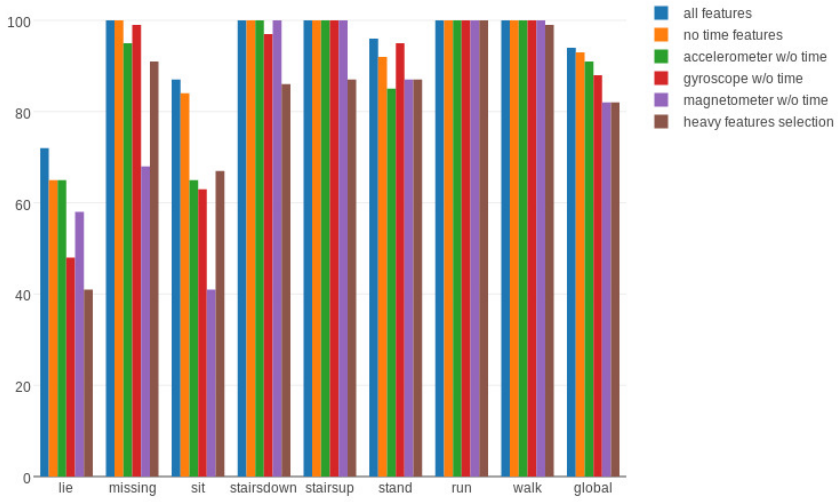


Figure 4 : Prediction Success Rate by Activity and Features Selection for the Decision Tree Classifier

2. Exemple 1

Soit m une masse définie sur le cadre de discernement suivant Θ= {« allonger », « regarder la TV », « manger », « écrire un sms », « parler au téléphone »}

Proposons les notations suivantes pour chaque activité contextuelle :

f_1 = "allonger" , f_2 = « regarder la TV », f_3 = "$manger$" , f_4 = "$écrire\ un\ sms$" , f_5 = "$parler\ au\ telephone$". Supposons que nous souhaitons supprimer un seul élément focal du cadre de discernement.

Nous avons donc le cadre de discernement réécrit comme suit : Θ = $\{f_1, f_2, f_3, f_4, f_5\}$, et nous avons les masses générée par la méthode des classifiers, comme suit :

m ($\{f_1\}$) = 0,07, m ($\{f_2, f_3, f_4, f_5\}$) = 0,09, m ($\{f_4, f_5\}$) = 0,3 m ($\{f_3, f_5\}$) = 0,19, m ($\{f_1, f_2\}$) = 0,1, m ($\{f_2, f_4, f_5\}$) = 0,15 m ($\{f_2, f_5\}$) = 0,1

- L'étape 1 A_1 est retirée de l'assignation de probabilité de base m et la masse de A_1 est affectée aux éléments focaux restants pour générer une nouvelle assignation m_1' Le processus spécifique d'assignation de masse est que S_1 = $\{f_1\}$. La masse de f_1 dans S_1 est de 0,07. Dans les éléments focaux restants, la seule intersection avec f_1 est $\{f_1, f_2\}$. Donc, la masse de f_1 est complètement affectée à $\{f_1, f_2\}$. Une nouvelle assignation de masse m_1' peut être obtenue, ce qui est :

$$\boldsymbol{m_1'(\{f_2, f_3, f_4, f_5\}) = 0.9, m_1'(\{f_4, f_5\}) = 0.3,}$$

$$\boldsymbol{m_1'(\{f_3, f_5\}) = 0.19, m_1'(\{f_1, f_2\}) = 0.17}$$

$$\boldsymbol{m_1'(\{f_2, f_4, f_5\}) = 0.15, m_1'(\{f_2, f_5\}) = 0.1}$$

- Étape 2 En utilisant le coefficient de corrélation, la similarité entre l'assignation de la masse m et m_1' d'origine est r (1) = 0,9948. Ensuite, en exécutant les étapes 1 et l'etape 2 pour tout i = 2,3,4,5,6, 7. A_2 est retiré de l'assignation de masse m et la masse de A_2 est assignée aux éléments focaux restants pour générer un nouvelle assignation de masse m_2'. La masse de f_2, f_3, f_4, f_5 est de 0,025 dans S_2, qui peut être obtenue par $\frac{m(\{f_2,f_3,f_4,f\})}{|\{f_2,f_3,f_4,f_5\}|}$ D'après l'intersection de leurs données d'origine, la masse de f_2 est attribuée à $\{f_1, f_2\}$. La masse de f_3 est affectée à $\{f_1, f_2\}$, $\{f_2, f_4, f_5\}$ et $\{f_2, f_5\}$. La masse de f_3 est affectée à $\{f_3, f_5\}$. La masse de f_4 est affectée à $\{f_4, f_5\}$ et $\{f_2, f_4, f_5\}$. Le f_5 est attribué à $\{f_4, f_5\}$, $\{f_3, f_5\}$, $\{f_2, f_4, f_5\}$, $\{f_2, f_5\}$. Alors, m_2' peut être obtenu, ce qui est

$$\boldsymbol{m_2'(\{f_1\}) = 0.07, m_2'(\{f_4, f_5\}) = 0.3169}$$

$$\boldsymbol{m_2'(\{f_3, f_5\}) = 0.2181, m_2'(\{f_1, f_2\}) = 0.1075}$$

$$\boldsymbol{m_2'(\{f_2, f_4, f_5\}) = 0.1744, m_2'(\{f_2, f\}) = 0.1131}$$

La similarité entre l'assignation de masse m et m_2' est r (2) = 0,9962. De la même manière, nous pouvons obtenir

$$\boldsymbol{m_3'(\{f_1\}) = 0.07, m_3'(\{f_2, f_3, f_4, f_5\}) = 0.2025}$$

$$\boldsymbol{m'_3(\{f_3, f_5\}) = 0.2275, m'_3(\{f_1, f_2\}) = 0.1}$$

$$\boldsymbol{m'_3(\{f_2, f_4, f_5\}) = 0.2625, m'_3(\{f_2, f_5\}) = 0.1375}$$

$$\boldsymbol{m'_4(\{f_1\}) = 0.07, m'_4(\{f_2, f_3, f_4, f_5\}) = 0.2087}$$

$$\boldsymbol{m'_4(\{f_4, f_5\}) = 0.3238, m'_4(\{f_1, f_2\}) = 0.1}$$

$$\boldsymbol{m'_4(\{f_2, f_4, f_5\}) = 0.1738, m'_4(\{f_2, f_5\}) = 0.1237}$$

$$\boldsymbol{m'_5(\{f_1\}) = 0.12, m'_5(\{f_2, f_3, f_4, f_5\}) = 0.1066}$$

$$\boldsymbol{m'_5(\{f_4, f_5\}) = 0.3, m'_5(\{f_3, f_5\}) = 0.19}$$

$$\boldsymbol{m'_5(\{f_2, f_4, f_5\}) = 0.1167, m'_5(\{f_2, f_5\}) = 0.1166}$$

$$\boldsymbol{m'_6(\{f_1\} = 0.07, m'_6(\{f_2, f_3, f_4, f_5\}) = 0.1442}$$

$$\boldsymbol{m'_6(\{f_4, f_5\}) = 0.3375, m'_6(\{f_3, f\}) = 0.2025}$$

$$\boldsymbol{m'_6(\{f_1, f_2\}) = 0.1167, m'_6(\{f_2, f_5\}) = 0.1291}$$

$$\boldsymbol{m'_7(\{f_1\}) = 0.07, m'_7(\{f_2, f_3, f_4, f_5\}) = 0.1192}$$

$$\boldsymbol{m'_7(\{f_4, f_5\}) = 0.3125, m'_7(\{f_3, f_5\}) = 0.2025}$$

$$\boldsymbol{m'_7(\{f_1, f_2\} = 0.1167, m'_7(f_2, f_4, f_5\}) = 0.1791}$$

La similitude est r (3) = 0,9359, r (4) = 0,9689, r (5) = 0,9929, r (6) = 0,9926, r (7) = 0,9935 séparément.

- Étape 3 Trier les A_i de m (dans l'ordre croissant) selon la similarité r. Le résultat est A_3, A_4, A_6, A_5, A_7, A_1, A_2.

Étape 4 Supprimer 1 (puisque nous avons choisi k=1) des éléments focaux de m et la masse des éléments focaux supprimés est assignée à la sortie des éléments focaux en fonction de la stratégie d'assignation. La masse approximée finale est la suivante

$$\boldsymbol{m(\{f_1\}) = 0.07, m_2(\{f_4, f_5\}) = 0.3169}$$

$$\boldsymbol{m(\{f_3, f_5\}) = 0.2181, m(\{f_1, f_2\}) = 0.1075}$$

$$\boldsymbol{m(\{f_2, f_4, f_5\}) = 0.1744, m(\{f_2, f_5\}) = 0.1131}$$

Dans cette section, nous proposons de comparer cette approche avec d'autres approches déjà existantes dans l'état de l'art. La précision de la décision fusionnée est l'une des clés pour

évaluer la méthode approximative. L'impact de notre méthode d'approximation des fonctions de masses sur la décision sera analysé dans l'exemple 2. En outre, nous présentons des résultats de simulation pour comparer les différentes approches d'approximation de fonctions de masses en termes de proximité avec l'original. Pour mesurer la proximité ou la dissemblance entre différentes fonctions, la mesure de distance de Jousselmes [191] entre fonctions est utilisée dans l'exemple 3. Le temps de calcul moyen (plus de 100 exécutions) pour la combinaison originale et la combinaison après l'approximation sont également montrées dans l'exemple 3.

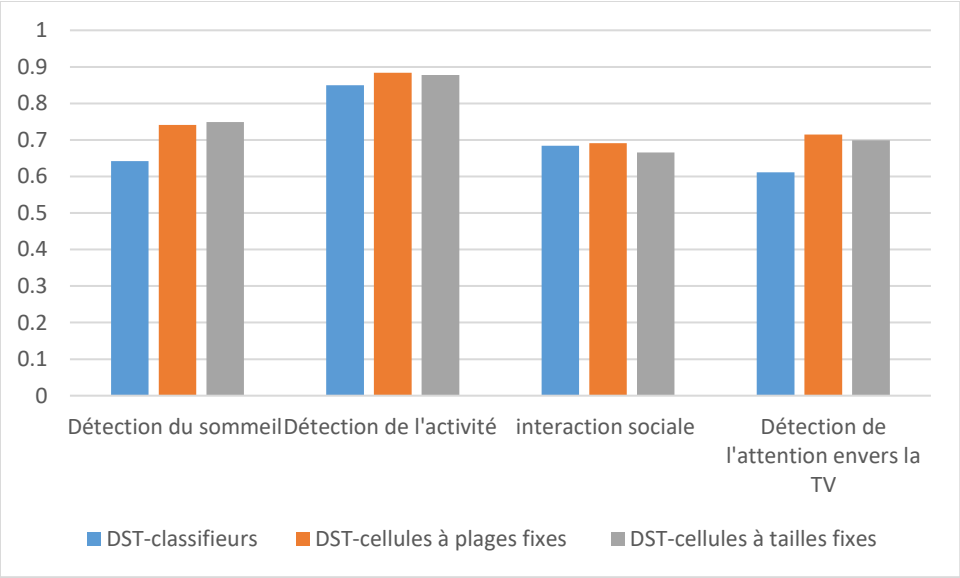


Figure 1 : Prédiction de l'attention envers la TV à l'aide de notre approche de la DST

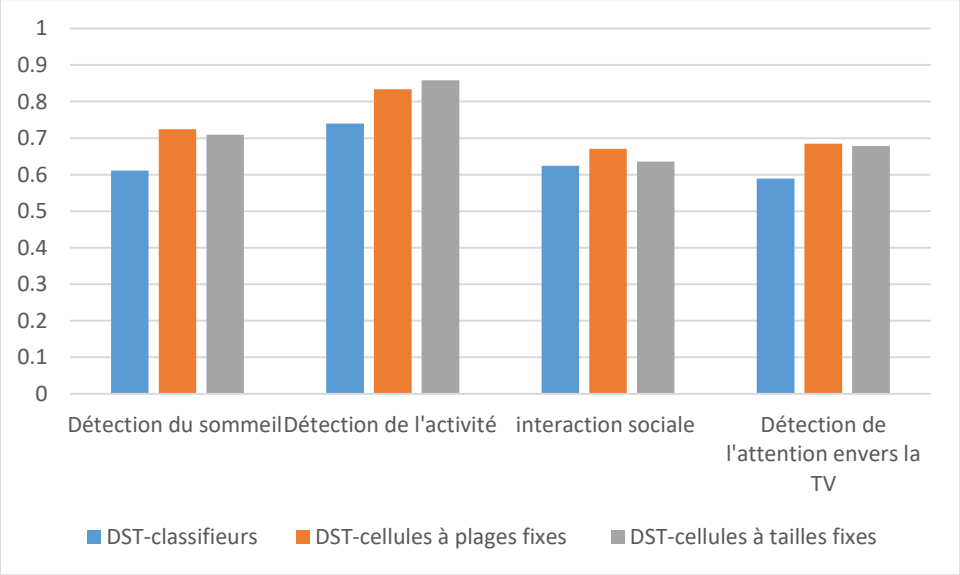


Figure 2 : Précision de reconnaissance de l'attention envers la TV en se basant sur l'approximation faite sur les trois sous-systèmes de prédiction

3. Dilemme utilisabilité vs portabilité

Notre expérience menée avec 33 personnes portant un iPhone, Apple Watch et Apple TV pendant 6 mois s'est déroulée dans un environnement non contrôlé. Les participants n'ont reçu donc aucune instruction par rapport à la manière d'effectuer leurs activités et comment interagir avec les différents objets disponibles.

Nous proposons d'étudier l'impact de l'utilisabilité des objets connectés sur la reconnaissance de l'activité, nous aborderons différents critères allant du choix du modèle, le paramétrage et le cout de calcul et d'apprentissage et son optimisation.

La figure ci-dessous représente une vue globale montrant la différence entre la détection de l'activité physique avec un capteur par rapport sa détection avec un objet connecté.

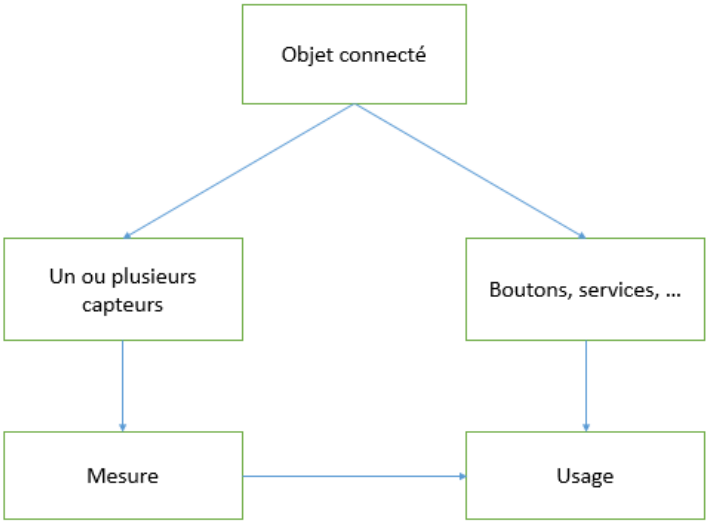


Figure 3 : Les objets connectés et la notion de l'utilisabilité.

En effet, un objet connecté embarque des capteurs de différentes natures : accéléromètre, gyroscope, magnétomètre, …Les mesures qu'apportent ces capteurs contribuent entre autre la reconnaissance de l'activité physique mais pas que. Les téléphones par exemple contiennent des boutons, etc. En utilisant les données renvoyées par ces boutons et en détectant l'activité, nous pouvons remonter à un état biométrique : par exemple : téléphoner en étant assis, taper un sms en étant debout, …C'est ce qu'on appelle l'usage.

Dans la littérature, plusieurs questions se posent autour du traitement de données de l'activité et du processus suivi. La plupart des travaux sont centrés autour de la procédure suivante : (1) collecte de données, (2) extraction de descripteurs ou pas et (3) application des algorithmes d'apprentissage automatique pour la classification.

Prenons deux cas d'études concrets :

- Cas 1 : Télécommander la TV avec son téléphone
- Cas 2 : téléphoner

Que ce soit dans le cas 1 ou cas 2, les participants effectuent leurs activités (debout, assis, manger, …) en se servant du téléphone pour générer un usage.

Voyons donc ce que va donner l'algorithme du CNN-brutes-PSO pour classifier les activités (debout, assis, allonger) dans le cas 1 et le cas 2.

Il s'agit de classifier les activités d'usage (cas 1 et cas 2) ainsi que les trois activités sous-jacentes à ces deux usages, à savoir : debout, assis et allonger.

Les résultats sont présentés dans le graphique suivant :

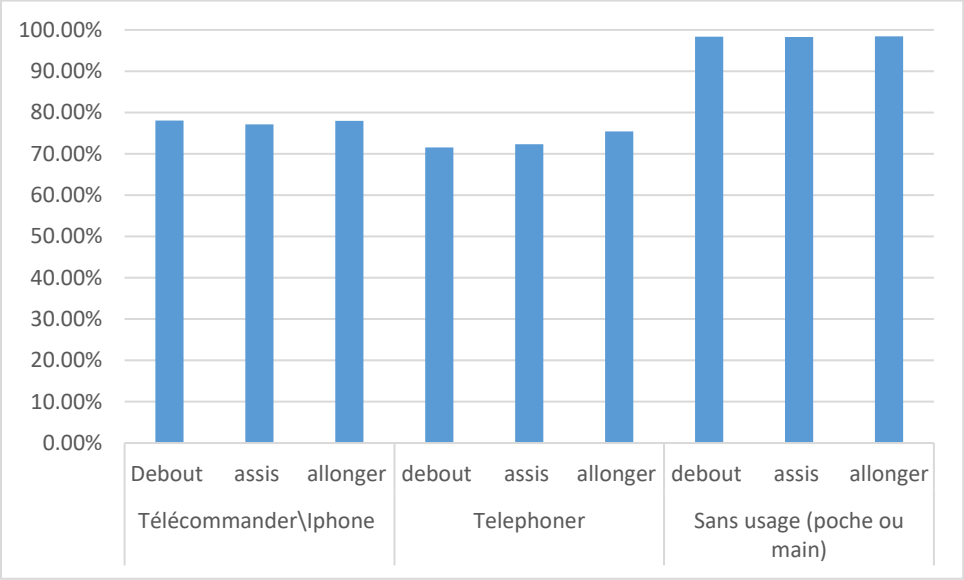


Figure 4 : Précisions de classification des activités « debout », « assis » et « allonger » suivant qu'il ait prise en compte de l'aspect usage ou pas Vs portabilité (poche ou dans la main).

Comme nous pouvons le constater, la classification suivant l'usage donne des performances moins bonnes que la classification sans prise en compte de l'usage. En effet, tout l'intérêt d'utiliser un algorithme performant et surtout bien optimisé pour rendre cette classification précise et rapide.

Annexe E

Optimisation des réseaux de neurones profonds

Annexes

Nous proposons, dans la partie suivante, d'aborder les principaux facteurs pouvant influencer l'entrainement d'un réseau de neurones profond. Nous réaliserons une étude comparative et exploratoire de l'ensemble de ces facteurs, afin de reconnaitre les activités élémentaires et basées sur le contexte. Nous verrons en particulier deux types de facteurs : **i)** Choix de la fonction d'activation et **ii)** Utlisation d'optimiseurs plus rapides définis dans l'état de l'art.

1. Etude comparative pour l'optimisation des réseaux de neurones profonds

L'état de l'art a montré que les réseaux de neurones profonds souffrent principalement de deux problèmes : i) la disparition et/ou explosiion des gradients. Ii) lenteur du procéssus d'entrainement.

Afin de répondre aux problèmes de disparition et/ou d'explosion de grandients, l'état de l'art a proposé d'utiliser différentes fonctions d'activation à incorporer aux différents modèles d'apprentissage profonds. Ainsi, afin de résoudre le problème de lenteur du procéssus d'entrainement, l'état de l'art a proposé plusieurs techniques d'optimisation. Il est donc nécéssaire de choisir à la fois la bonne fonction d'activation ainsi que le bon optimiseur du modèle d'apprentissage profond pour la reconnaissance des activités.

Nous proposons dans cette séction d'explorer ces deux critères en élaborant une étude prospective afin de choisir les meiileurs fonction de masse et optimiseur. Nous commençons par identifier les meilleures fonctions d'activations.

1.1. Choix de la fonction d'activation

L'algorithme de rétro propagation procède en deux passes à travers le modèle : une passe avant, de la couche d'entrée vers la couche de sortie, suivie d'une passe arrière qui propage le gradient d'erreur de la couche de sortie vers la couche d'entrée [176].

Lorsque l'algorithme a déterminé le gradient de la fonction de cout par rapport à chaque paramètre du réseau, il utilise ces gradients pour actualiser chaque paramètre au cours d'une étape de descente du gradient.

Un des problèmes importants des gradients, alors que l'algorithme progresse vers les couches inférieures, c'est qu'ils deviennent souvent de plus en plus petits. Par conséquent, la mise à jour par descente de gradient modifie très peu les poids des connexions de la couche inférieure et l'entrainement ne converge jamais vers une bonne solution. Tel est le problème de disparition des gradients (vanishing gradients, en anglais) [177].

Dans certains cas, l'opposé peut se produire. Les gradients deviennent de plus en plus gros et de nombreuses couches reçoivent des poids extrêmement importants, ce qui fait diverger l'algorithme. Il s'agit du problème d'explosion des gradients (*exploding gradients*, en anglais), qui se rencontre principalement dans les réseaux de neurones récurrents. Plus généralement, les réseaux de neurones profonds souffrent d'instabilité des gradients : différentes couches peuvent apprendre à des vitesses très différentes [178].

Bien que ce comportement ait été observé empiriquement depuis un moment (il est l'une des raisons de l'abandon des réseaux de neurones profonds pendant une longue période), ce n'est qu'autour de 2010 que sa compréhension a fait un bond en avant. Xavier Glorot et Yoshua Bengio [179] ont identifié quelques caractéristiques, notamment l'association de la fonction d'activation logistique (sigmoïde) la plus répandue à l'époque, et la technique d'initialisation des poids également la plus répandue à ce moment-là, à savoir l'initialisation aléatoire avec une loi normale de moyenne zéro et d'écart type 1. En résumé, ils ont montré que, avec cette fonction d'activation et cette technique d'initialisation, la variance des sorties de chaque couche est largement supérieure à celle de ses entrées. Lors de l'avancée dans le réseau, la variance ne cesse d'augmenter après chaque couche, jusqu'à la saturation de la fonction d'activation dans les couches supérieures. Ce comportement est aggravé par le fait que la moyenne de la fonction logistique est non pas de 0 mais de 0.5 (la fonction tangente hyperbolique a une moyenne égale à 0 et se comporte également mieux que la fonction logistique dans les réseaux profonds).

Nous constatons que, lorsque les entrées deviennent grandes (en négatif ou en positif), la fonction sature en 0 ou en 1, avec une dérivée extrêmement proche de 0. Lorsque la rétro propagation intervient, elle n'a pratiquement aucun gradient à transmettre en arrière dans le réseau. En outre, le faible gradient existant est de plus en plus dilué pendant la rétro propagation, à chaque couche traversée. Il ne reste donc quasi plus rien aux couches inférieures

Glorot et Bengio proposent une manière d'atténuer énormément ce problème. Le signal doit se propager correctement dans les deux directions : vers l'avant au moment des prédictions, et vers l'arrière lors de la rétro propagation des gradients. Il ne faut pas que le signal disparaisse, ni qu'il explose et sature. Pour que tout se passe proprement, les auteurs soutiennent que la variance des sorties de chaque couche doit être égale à la variance avant et après le passage au travers d'une couche en sens inverse. Il est impossible de garantir ces deux points, sauf si la couche possède un nombre égal de connexions d'entrée et de sorties. Ils ont cependant proposé un bon compromis, dont la validité a été montrée en pratique : les

poids des connexions doivent être initialisés de façon aléatoire. Cette stratégie d'initialisation est souvent appelée initialisation de Xavier ou initialisation de Glorot.

L'ensemble des illustrations sont réalisées par rapport aux activités physiques basées sur le contexte. **La figure 1** représente l'entrainement du réseau de neurones CNN-brutes en utilisant l'optimisation de Xavier en comparaison avec l'entrainement du réseau DNN-DCT.

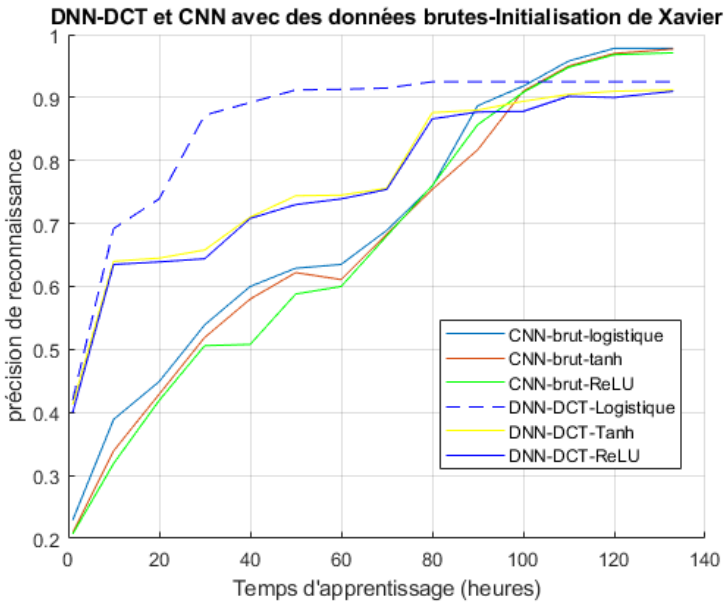


Figure 1: comparaison entre le CNN-brutes et DNN-DCT en utilisant l'initialisation de Xavier.

Le comportement des deux réseaux est très différent selon le type de la fonction d'activation. En effet, la fonction logistique garde pratiquement les mêmes précisions et temps d'apprentissage. Alors que la fonction ReLU permet de réduire le temps d'apprentissage pour les deux algorithmes. En revanche, nous remarquons une baisse de la précision de reconnaissance pour passer de 97.8% (pour le CNN-brutes avec ReLU à 97.1%, c'est un changement dans la deuxième décimale, qui n'est pas très significatif). Le DNN-DCT passe de 92.5% à 91% de précision avec la fonction ReLU), avec un petit gain en temps d'apprentissage.

L'un des enseignements apportés par les travaux de X. Glorot et de Y. Bengio a été que les problèmes de disparition/explosion des gradients étaient en partie dus à un mauvais choix de la fonction d'activation. C'est-à-dire qu'elles devaient constituer un excellent choix. Cependant, il s'avère que d'autres fonctions d'activation affichent un meilleur comportement dans les réseaux de neurones profonds, notamment la fonction ReLU, principalement par ce qu'elle ne sature pas pour les valeurs positives (et aussi par ce qu'elle est plutôt rapide à calculer).

Malheureusement, la fonction d'activation ReLU n'est pas parfaite. Elle souffre d'un problème de mort des ReLU (*dying ReLU,* en anglais) : au cours de l'entrainement, certains neurones

« meurent », c'est-à-dire qu'ils s'arrêtent de produire autre chose que 0. Dans certains cas, il arrive que la moitié des neurones du réseau soient morts à l'issue de l'apprentissage, en particulier si le taux d'apprentissage est grand. Pendant un apprentissage, si l'algorithme de descente de gradient modifie les poids d'un neurone de sorte que la somme pondérée de ses entrées soit négative, ce neurone va commencer à produire 0 en sortie. Il est peu probable que ce neurone reprenne vie, car le gradient de la fonction ReLU vaut 0 lorsque sa sortie est négative.

Afin de résoudre ce problème, une variante de la fonction ReLU a été employée, par exemple leaky ReLU (*leaky* pour fuir en anglais). Cette fonction se définit comme$LeakyReLU(z) = \max(\alpha z, z)$ [180].

L'hyperparamètre α définit le niveau de « fuite » de la fonction : il s'agit de la pente de la fonction pour $z < 0$. En general, il est fixé à 0.01. Grace à cette petite pente, les « leaky ReLU » ne meurent jamais. Ils peuvent entrer dans un long coma, mais ils ont toujours une chance de se réveiller à un moment ou à un autre.

Les travaux de (Xu et al, 2015) [181] comparent plusieurs variantes de la fonction d'activation ReLU et conclut notamment que la variante « leaky » est toujours plus performante que la version stricte. En réalité, en fixant $\alpha = 0.2$ (fuite importante), il semble que les performances soient toujours meilleures qu'avec $\alpha = 0.01$ (fuite légère). Les auteurs ont également évalué la fonction RReLU (Randomized leaky ReLU, en anglais), dans laquelle α est, pendant l'apprentissage, choisi aléatoirement dans une plage donnée et, pendant les tests, fixé à une valeur moyenne. Elle donne de bons résultats et semble agir comme un régularisateur (en réduisant le risque de sur-ajustement du jeu d'apprentissage). Enfin, ils ont également évalué la fonction PReLU (Parametric leaky ReLU, en anglais), qui autorise l'apprentissage de α pendant l'apprentissage (il n'est plus un hyperparamètre mais un paramètre du modèle qui, au meme titre que les autres, peut être modifié par la rétropropagation). Elle surpasse largement la fonction ReLU sur les vastes jeux de données comme les images ou des grandes bases de données de signaux, mais, avec les jeux de données plus petits, elle fait courir un risque de surajustement.

Par ailleurs, Djork-André Clevert [180] propose une nouvelle fonction d'activation appelée ELU (Exponential Linear Unit), qui s'est montrée bien plus performante que toutes les variantes de ReLU dans leurs expérimentations. Le temps d'apprentissage a diminué et le réseau de neurones s'est mieux comporté sur le jeu de test. L'équation 5.2 en donne la définition :

$$ELU_{\alpha}(z) = \begin{cases} \alpha(\exp(z) - 1) & si\ z < 0 \\ z & sinon \end{cases} \quad (5.2)$$

Elle ressemble énormément à la fonction ReLU, mais avec quelques différences majeures :

Premièrement, elle prend en charge les valeurs négatives lorsque $z < 0$, ce qui permet au neurone d'avoir une sortie moyenne plus proche de zéro. Comme nous l'avons expliqué, cela atténue le problème de disparition des gradients. L'hyperparamètre α définit la valeur vers laquelle la fonction ELU tend lorsque z est un grand nombre négatif. Il est généralement fixé à 1, mais nous pouvons l'ajuster au besoin, comme n'importe quel autre hyperparamètre.

Deuxièmement, elle présente un gradient différent de zéro pour $z < 0$, ce qui évite le problème de mort des neurones.

Troisièmement, la fonction est continue en tout point, y compris autour de $z = 0$, ce qui permet d'accelerer la descente de gradient, car il ne rebondit pas autant à gauche et à droite de $z = 0$.

Elle a pour principal inconvénient d'être plus lente à calculer que la fonction ReLU et ses variantes (en raison de l'utilisation de la fonction exponentielle), mais, au cours de la phase d'apprentissage, cette lenteur est composée par une vitesse de convergence plus élevée. En revanche, lors des tests, un réseau ELU sera plus lent qu'un réseau ReLU.

Et enfin, l'introduction de la fonction d'activation SELU, récemment [182] (Scaled ELU, en anglais), qui est simplement la fonction ELU légèrement redimensionnée et avec une valeur particulière pour α :

$$\text{SELU (z)} = 1.0507 * ELU_{1.67326}(z)$$

Cette fonction a la propriété de pousser le réseau de neurones à s'auto-normaliser, c'est-à-dire qu'il va automatiquement tendre à conserver à peu près la même variance pour les sorties de chaque couche, tout au long de l'apprentissage. Le réseau de neurones évite ainsi entièrement le problème de disparition ou d'explosion des gradients, l'entrainement est accéléré et produit de meilleurs résultats.

Avec l'arrivée de la fonction d'activation SELU, le problème de disparition/explosion des gradients est peut-être enfin résolu. Néanmoins, il est encore trop tôt pour l'affirmer avec certitude, et il reste utile d'étudier les autres techniques de normalisation qui étaient les plus employées jusqu'à présent.

La figure 2 suivante montre les résultats de l'implémentation des deux modèles en utilisant l'ensemble des fonctions d'activation non saturantes avec les deux modèles de reconnaissance de l'activité.

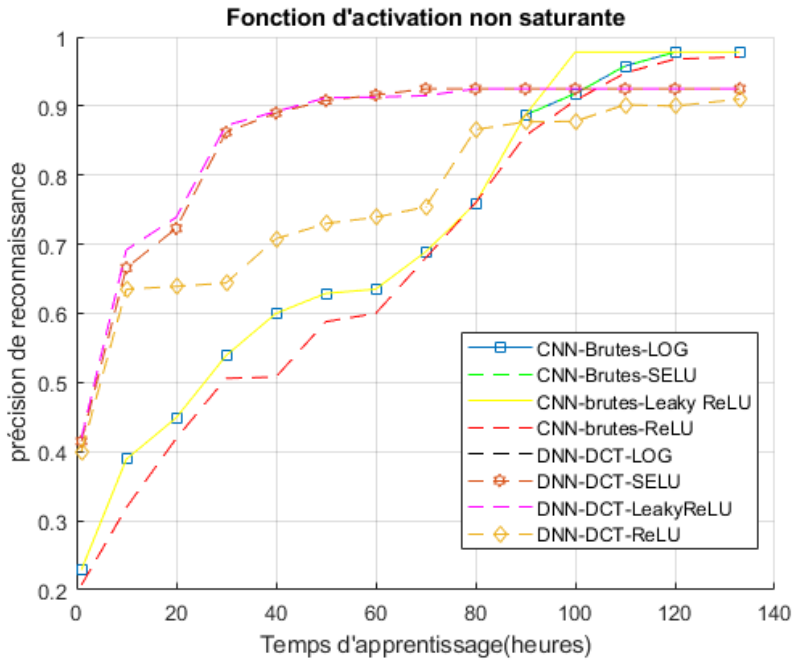


Figure 2 : Evolution du temps d'apprentissage suivant le type de fonction d'activation.

La fonction SELU fait légèrement mieux que les autres concurrentes. Néanmoins, son impact sur le modèle CNN-brutes n'est pas très significatif. Il permet d'avoir un gain d'environ une heure en temps d'apprentissage. Toutefois, le modèle DNN-DCT perd un peu en performance et son temps d'apprentissage ne s'améliore quasiment pas.

Sergey loffe et Christian Szegedy [183] ont proposé une technique appelée normalisation par lots (ou *Batch normalization* (BN) en anglais) pour traiter les problèmes des gradients et, plus généralement, le problème lié au fait que, pendant l'apprentissage, la distribution des entrées de chaque couche évolue en fonction des changements des paramètres des couches précédentes (ce qu'ils appellent le problème de *Internal Covariate Shift*).

Leur technique consiste à ajouter une opération dans le modèle qui se contente de centrer sur zéro et normaliser ses entrées, puis à mettre à l'échelle et à décaler le résultat en utilisant deux nouveaux paramètres par couche (l'un pour la mise à l'échelle, l'autre pour le décalage). Autrement dit, cette opération permet au modèle d'apprendre l'échelle et la moyenne optimales des données. On place généralement une couche BN à la sortie de chaque couche du réseau de neurones, généralement juste avant sa fonction d'activation.

Pour pouvoir centrer sur zéro et normaliser ses entrées, la couche BN a besoin d'évaluer la moyenne et l'écart type de ses entrées. Elle procède en déterminant ces valeurs sur le mini-lot courant (d'où le nom de normalisation par lots).

Pendant les tests, nous ne disposons d'aucun mini-lot pour calculer la moyenne et l'écart type empirique des entrées de la couche BN. A la place, nous utilisons simplement la moyenne et

l'écart type des entrées de la couche BN mesurés sur l'ensemble du jeu de données d'entrainement. Ces valeurs sont en général calculées efficacement au cours de l'entrainement en utilisant une moyenne mobile exponentielle : lorsque la moyenne μ_B d'un mini-lot est mesurée sur le mini-lot courant, l'estimation μ_G de la moyenne globale (sur l'ensemble du jeu de données d'entrainement) est mise à jour à l'aide de l'équation suivante : $\mu_G \leftarrow (1-\eta) \times \mu_G + \eta \times \mu_B$ où η est un hyperparamètre appelé la constante de lissage, typiquement égale à 0.1.

L'estimation de l'écart type global σ_G est mise à jour de la meme façon :

$\mu_G \leftarrow (1-\eta) \times \sigma_G + \eta \times \sigma_G$.

Par conséquent, pour chaque couche normalisée par lots, les paramètres suivants sont appris lors de l'entrainement : γ (échelle), β (décalage), μ_G (la moyenne) et σ_G (l'écart type).

Les auteurs ont démontré que cette technique améliore considérablement tous les réseaux de neurones profonds qu'ils ont testés. Le problème de disparition des gradients est fortement réduit, à tel point qu'ils ont pu utiliser avec succès des fonctions d'activation saturantes, comme la tangente hyperbolique et la fonction logistique. Les réseaux étaient également beaucoup moins sensibles à la méthode d'initialisation des poids. Ils ont également été en mesure d'utiliser des taux d'apprentissage beaucoup plus élevés, accélérant ainsi le processus d'apprentissage. Ils ont notamment remarqué que, appliquée à un modèle de pointe pour la classification des images, la normalisation par lots permet d'arriver à la même précision, mais avec quatorze fois moins d'étapes d'entrainement, et bat le modèle d'origine d'une bonne longueur. En utilisant un ensemble de réseaux de neurones normalisés par lots, ils ont pu améliorer le meilleur résultat publié sur la classification ImageNet : en atteignant un taux d'erreur de validation top 5 de 4.9% (et un taux d'erreur de test de 4.8%), dépassant la précision des évaluateurs humains. Enfin, la normalisation par lots agit également comme un régulateur, diminuant le besoin de recourir à d'autres techniques de régularisation (comme celle du dropout, par exemple).

A noter également que la normalisation par lots ajoute toutefois une certaine complexité au modèle. Par ailleurs, la performance du réseau de neurones au moment de l'exécution est pénalisée : il effectue des prédictions plus lentement, car les calculs requis dans chaque couche sont plus nombreux.

Implémentation des deux modèles en utilisant la normalisation par lot donne les résultats de la **Figure 3** :

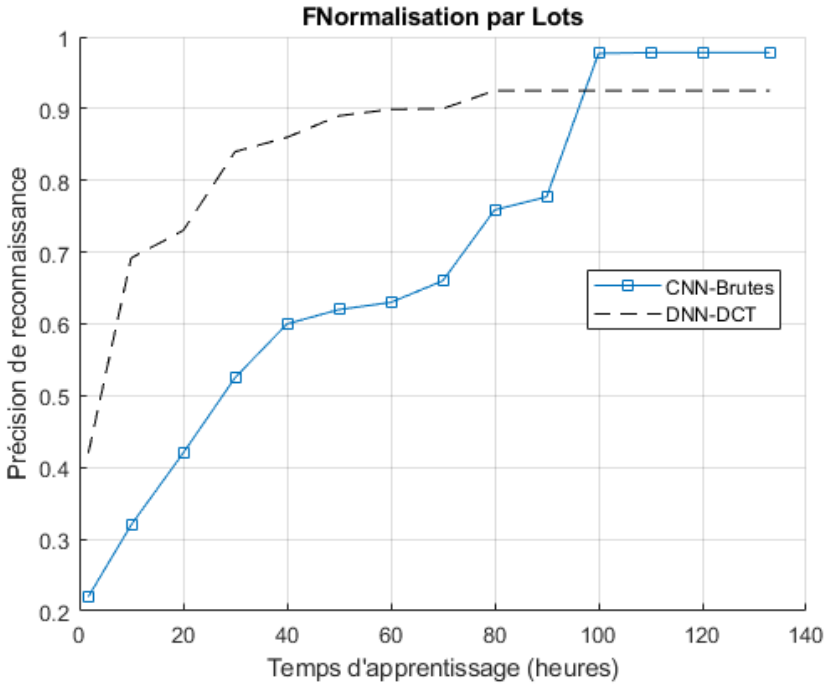


Figure 3 : Evolution du temps d'apprentissage suivant l'utilisation d'une fonction de normalisation par lots.

Ici encore, nous pouvons remarquer deux choses :

(a) Accélération de l'apprentissage des deux modèles avec des gains de 30 minutes et 35 environs pour le modèle CNN-brutes et DNN-DCT

(b) Diminution de la précision du modèle DNN-DCT.

L'optimisation du processus de reconnaissance des activités physiques basées sur le contexte opéré jusqu'à présent et basé sur les trois critères suivants : (1) appliquer une bonne stratégie d'initialisation de poids des connexions, (2) utiliser une bonne fonction d'activation, (3) utiliser la normalisation par lots ont permis d'améliorer le processus de traitement tantôt sur le plan de précision de classification tantôt sur la vitesse de convergence mais cela reste quand même très fragile. Il va falloir rechercher d'autres techniques pouvant aider à améliorer ces résultats.

Nous aborderons dans la partie suivante d'autres techniques d'optimisation issues des derniers travaux de recherche. Nous présenterons, comme nous l'avions fait auparavant, une évaluation empirique de ces méthodes.

1.2. Optimiseurs plus rapides

Nous pouvons également accélérer l'entrainement en utilisant un optimiseur plus rapide que l'optimiseur de descente de gradient ordinaire. Dans cette section, nous allons présenter et implémenter les solutions les plus répandues : optimisation avec inertie, gradient accéléré de Nesterov, AdaGrad, RMSProp et optimisation Adam.

Prenant l'exemple ordinaire, imaginons une boule de bowling qui roule sur une surface lisse en légère pente : elle démarre lentement, mais prend rapidement de la vitesse jusqu'à atteindre une vélocité maximale (s'il y a des frottements de l'air). Voilà l'idée simple derrière l'optimisation avec inertie (momentum optimization) proposée par Boris Polyak. A l'opposé, la décente de gradient ordinaire ferait de petits pas réguliers vers le bas de la pente et mettrait donc plus de temps pour arriver à son extrémité.

Rappelons que la descente de gradient met simplement à jour les paramètres du modèle Θ en soustrayant directement le vecteur gradient de la fonction de cout $J(\theta)$ par rapport aux paramètres (noté $\nabla_\theta J(\theta)$, multiplié par le taux d'apprentissage η. L'équation est :

$\theta \leftarrow \theta - \eta \nabla_\theta J(\theta)$). Elle ne tient pas compte des gradients antérieurs. Si le gradient local est minuscule, elle avance doucement.

L'optimisation avec inertie s'intéresse aux gradients antérieurs : à chaque itération, elle soustrait le gradient local au vecteur d'inertie *m* (multiplié par le taux d'apprentissage η) et actualise le vecteur des paramètres Θ en additionnant simplement ce vecteur d'inertie. Autrement dit, le gradient est utilisé non pas comme un facteur de vitesse mais d'accélération. Pour simuler une forme de frottement et éviter que la vitesse ne s'emballe, l'algorithme introduit un nouvel hyperparamètre β, appelé simplement inertie dont la valeur doit etre comprise entre 0 et 1. Une valeur fréquemment utilisée est 0.9.

$$m \leftarrow \beta m - \eta \nabla_\theta J(\theta)$$

$$\theta \leftarrow \theta + m$$

Nous pouvons vérifier que si le gradient reste constant, la vélocité finale (c'est-à-dire la taille maximale des mises à jour des poids) est égale à ce gradient multiplié par$-\eta/(1-\beta)$. Par exemple, si $\beta = 0.9$, alors la vélocité finale est égale à -10 fois le gradient fois le taux d'apprentissage. L'optimisation avec inertie permet ainsi d'aller jusqu'à dix fois plus rapidement que la descente de gradient ! Elle peut ainsi sortir des zones de faux plat plus rapidement que la descente de gradient. En particulier, lorsque les entrées ont des échelles très différentes, la fonction de cout va ressembler à un bol allongé : la descente de gradient arrive en bas de la pente abrupte assez rapidement, mais il lui faut ensuite beaucoup de temps pour descendre la vallée. En revanche, l'optimisation avec inertie va avancer de plus en plus rapidement vers le bas de la vallée, jusqu'à l'atteindre l'optimum.

Dans les réseaux de neurones profonds qui ne mettent pas en œuvre la normalisation par lots ou la fonction d'activation SELU, les couches supérieures finissent souvent par recevoir des

entrées aux échelles très différentes. Dans ce cas, l'optimisation avec inertie est d'une aide précieuse. Elle permet également de sortir des optima locaux.

La figure 4 compare les deux modèles CNN-brutes et DNN-DCT en termes d'optimisation avec Inertie (définie dans l'état de l'art).

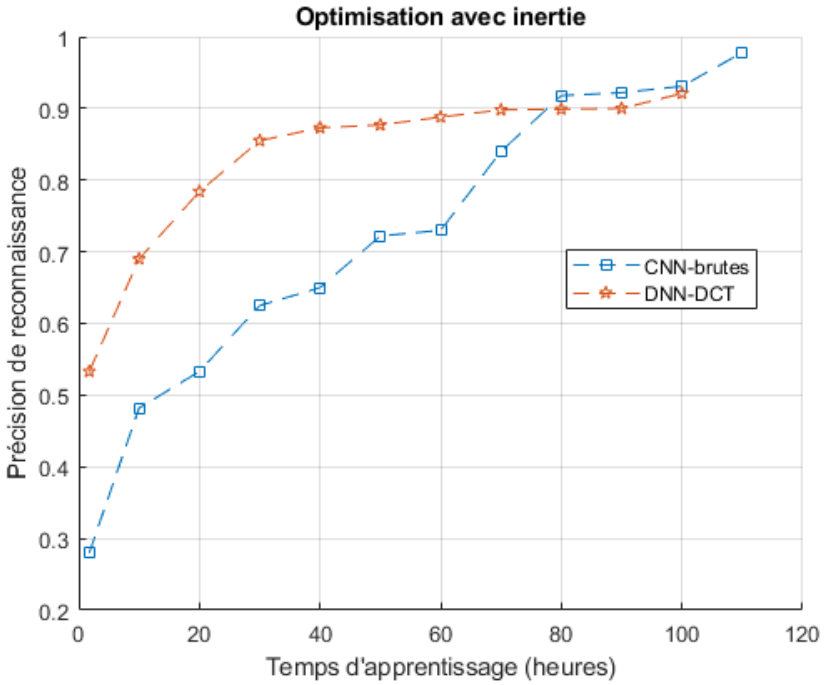


Figure 4 : Evolution du temps d'apprentissage en utilisant la méthode d'optimisation parinertie.

Ici, nous pouvons remarquer que les deux algorithmes ont réagi différemment :

- Le CNN-brutes affiche la même précision de reconnaissance constante 97.8% et progresse encore du temps d'apprentissage pour passer à 5 heures au lieu de 4 heures dans l'expérience précédente.
- Le DNN-DCT quant à lui, perd en performance pour passer de 92.5% à 91.2% mais a largement surpassé le premier modèle : le CNN-brutes.

Yurii Nesterov a proposé une petite variante de l'optimisation avec inertie [184], qui se révèle toujours plus rapide que la version d'origine. L'idée de l'optimisation avec inertie de Nesterov, également appelée gradient accéléré de Nesterov, est de mesurer le gradient de la fonction de cout non pas à position locale mais légèrement en avant dans le sens de l'inertie. Par rapport à l'optimisation avec inertie d'origine, la seule différence est que le gradient est mesuré non plus à θ mais à $\theta + \beta m$

$$m \leftarrow \beta m - \eta \nabla_\theta J(\theta + \beta m)$$

$$\theta \leftarrow \theta + m$$

Ce petit ajustement fonctionne car, en général, le vecteur d'inertie pointe dans la bonne direction (c'est-à-dire vers l'optimum). Il sera donc légèrement plus précis d'utiliser le gradient mesuré un peu plus en avant dans cette direction que d'utiliser celui en position d'origine.

L'implémentation de cette méthode sur nos deux modèles CNN-brutes et DNN-DCT est représentée dans **la figure 5.**

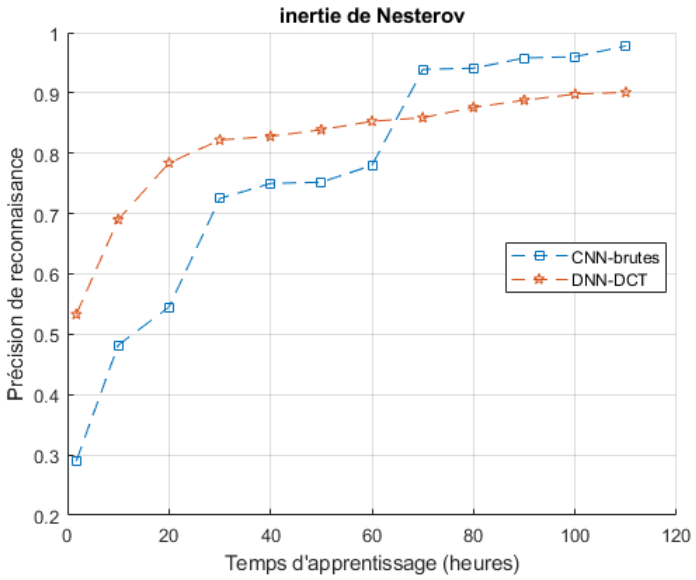


Figure 5 : Evolution du temps d'apprentissage en utilisant l'inertie de Nesterov.

Ici encore, les deux algorithmes ont eu des comportements presque similaires :

- Le CNN-brutes a progréssé en temps d'apprentissage. Il s'est accéléré et arrive à l'optimum rapidement contrairement à l'approche ordinaire de l'inertie.
- Par contre le DNN-DCT a encore perdu en performance et est passé à 90% de précision.

Considérons à nouveau le problème du bol allongé : la descente de gradient commence par aller rapidement vers le bas de la pente la plus abrupte, puis elle va lentement vers le bas de la vallée. Il serait préférable que l'algorithme détecte ce comportement au plus tôt et revoie son orientation pour se diriger un peu plus vers l'optimum global.

L'algorithme AdaGrad [185] met cela en place en diminuant le vecteur de gradients le long des dimensions les plus raides.

$$s \leftarrow s + \nabla_\theta J(\theta) \otimes \nabla_\theta J(\theta)$$

$$\theta \leftarrow \theta - \eta \nabla_\theta J(\theta) \oslash \sqrt{s + \varepsilon}$$

La première étape accumule les carrés des gradients dans le vecteur s. Le symbole $\otimes$ représente la multiplication terme à terme, appelée également produit de Hadamard. Cette forme vectorisée équivaut au calcul de $s_i \leftarrow s_i + (\frac{\partial J(\theta)}{\partial \theta_i})^2$ pour chaque élément s_i du vecteur s. Si la fonction de cout présente une pente abrupte le long de la i^{eme} dimension, alors s_i va augmenter à chaque itération.

La seconde étape est quasi identique à la descente de gradient, mais avec une différence importante : le vecteur de gradients est divisé par la racine carrée de s, plus un minuscule terme de lissage ε qui évite une division par zéro. Le symbole $\oslash$ représente la division terme à terme. Cette forme vectorisée équivaut au calcul de $\theta_i \leftarrow \theta_i - \frac{\eta \partial J(\theta)}{\partial \theta_i} / \sqrt{s + \varepsilon}$ pour tous les paramètres θ_i (simultanément).

En résumé, cet algorithme abaisse progressivement le taux d'apprentissage, mais il le fait plus rapidement sur les dimensions présentant une pente abrupte que pour celles dont la pente est plus douce. Nous avons donc un taux d'apprentissage adaptatif. Cela permet de diriger plus directement les mises à jour résultantes vers l'optimum global. Par ailleurs, l'algorithme exige un ajustement moindre de l'hyperparamètre η pour le taux d'apprentissage.

AdaGrad affiche un bon comportement pour les problèmes quadratiques simples, mais il s'arrête souvent trop tôt lors de l'entrainement des réseaux de neurones, du fait de l'accumulation des carrés des gradients dans le vecteur s. Le taux d'apprentissage est tellement diminué que l'algorithme finit par s'arrêter totalement avant d'atteindre l'optimum global.

L'entrainement des deux modèles CNN-brutes et DNN-DCT avec AdaGrad est représentée à **la figure 6.**

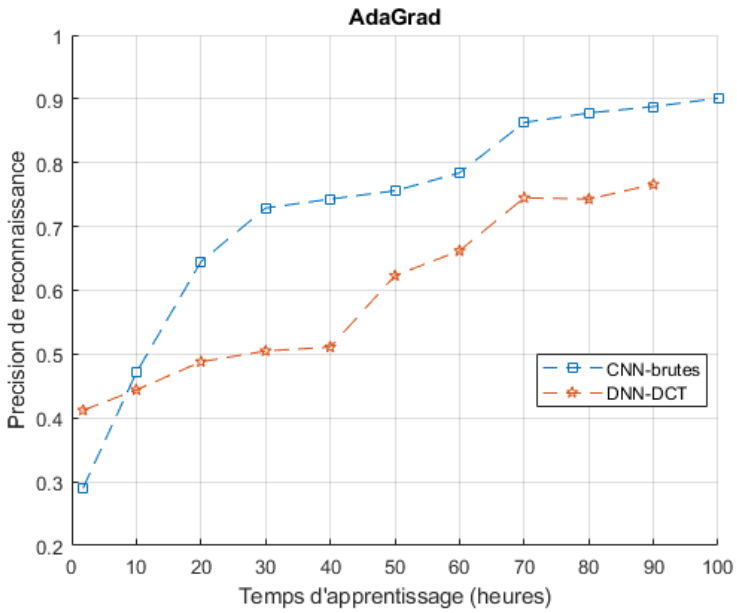


Figure 6 : Evolution du temps d'apprentissage par la méthode AdaGrad.

L'inconvénient d'AdaGrad est de ralentir un peu trop rapidement et de finir par ne jamais converger vers l'optimum global. L'algorithme RMSProp corrige ce problème en cumulant uniquement les gradients issus des itérations récentes (plutôt que tous les gradients depuis le début de l'entrainement). Pour cela, il utilise une moyenne mobile exponentielle au cours de la première étape [186] :

$$s \leftarrow \beta s + (1-\beta)\nabla_\theta J(\theta) \otimes \nabla_\theta J(\theta)$$

$$\theta \leftarrow \theta - \eta \nabla_\theta J(\theta) \oslash \sqrt{s+\varepsilon}$$

Le taux de décroissance β est en général fixé à 0.9. Il s'agit encore d'un nouvel hyperparamètre, mais cette valeur par défaut convient souvent et nous avons rarement besoin de l'ajuster.

Les deux algorithmes ont perdu en performance et en temps d'apprentissage, les modèles ont mal réagit à la méthode d'optimisation par AdaGrad. Nous proposons de voir si la méthode RMSProp peut aider à corriger ce problème (**Figure 7**).

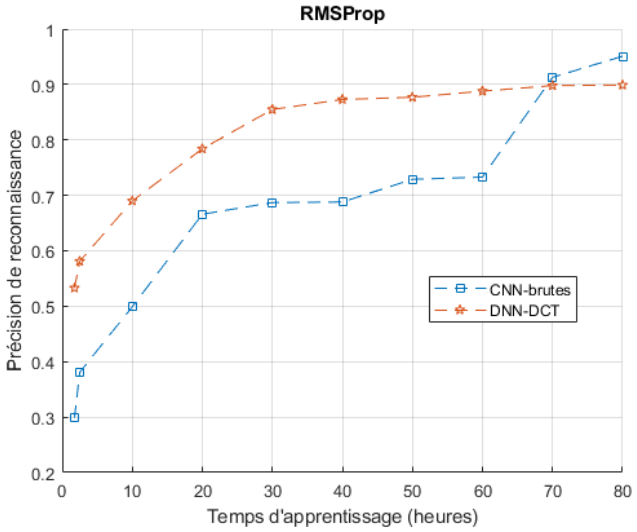


Figure 7 : Evolution du temps d'apprentissage par la méthode RMSProp.

Ici, nous remarquons que le modèle CNN-brutes a repris un peu de précision pour atteindre les 95.1% mais reste loin de son maximum déjà atteint (97.8%).

Quant à l'algorithme du DNN-DCT, lui aussi souffre de la précision assez basse par rapport à son maximum déjà atteint (90% au lieu de 95%).

Adam [187], pour *Adaptative Moment Estimation*, réunit les idées de l'optimisation avec inertie et de RMSProp. Il maintient, à l'instar de la première, une moyenne mobile exponentielle des gradients antérieurs et, à l'instar de la seconde, une moyenne mobile exponentielle des carrés des gradients passés :

1. $m \leftarrow \beta_1 m - (1-\beta_1)\nabla_\theta J(\theta)$
2. $s \leftarrow \beta_2 s + (1-\beta_2)\nabla_\theta J(\theta) \otimes \nabla_\theta J(\theta)$
3. $m \leftarrow \frac{m}{1-\beta_1^t}$
4. $s \leftarrow \frac{s}{1-\beta_2^t}$
5. $\theta \leftarrow \theta + \eta m \oslash \sqrt{s+\varepsilon}$

t représente le numéro de l'itération (en commençant à 1)

Si l'on examine uniquement l'étape 1,2 et 5, on constate une forte similitude avec l'optimisation avec inertie et RMSProp. La seule différence est que l'étape 1 calcule une moyenne mobile exponentielle à la place d'une somme à décroissance exponentielle, mais elles sont en réalité équivalentes, à un facteur constant près (la moyenne mobile exponentielle est simplement égale à 1-$\beta_1$multiplié par la somme à décroissance exponentielle).

Les étapes 3 et 4 représentent en quelque sorte un détail technique. Puisque m et s sont initialisés à zéro, l'algorithme pourrait mettre du temps à démarrer. Pour éviter cela, ces deux étapes aident à dynamiser m et s au début de l'entrainement : si β_1 et β_2 sont égaux à 0.9, alors m et s sont divisés par $1 - 0.9 = 0.1$, donc multipilié par 10, à la première iteration, puis divisé par 1-0.9^2=0.19, donc multipliés par environ 5.3, à la seconde itération, et ainsi de suite. Au bout de quelques dizaines d'itérations, ces étapes deviennent tout à fait négligeables car β_1^t et β_2^t deviennent minuscules.

L'hyperparamètre de décroissance de l'inertie $\beta_1$est en general initialisé à 0.9, tandis que l'hyperparamètre de décroissance de la mise à l'échelle β_2 est souvent initialisé à 0.999. Comme précédemment, le terme de lissage ε a habituellement une valeur initiale minuscule, par exemple10^{-8}.

En réalité, puisque Adam est un algorithme à taux d'apprentissage adaptatif, comme AdaGrad et RMSProp, le réglage de l'hyperparamètre η pour le taux d'apprentissage est moins important. Nous pouvons souvent conserver la valeur par défaut η=0.001, ce qui rend cet algorithme encore plus facile à employer que la descente de gradient.

L'optimisation Adam était considérée comme supérieure aux autres par de nombreux chercheurs, du moins pour l'entrainement de réseaux de neurones profonds, mais un dernier travail de Ashia C.Wilson et al. 2017 [188] a démontré que dans certains cas elle pouvait converger vers des solutions très mauvaises. Leur démonstration s'appliquait d'ailleurs également aux autres méthodes d'optimisation à taux d'apprentissage adaptatif (AdaGrad et RMSProp). Dans la plupart des cas, ces méthodes fonctionnent bien et convergent plus rapidement que les autres, mais il y a un risque sérieux de converger vers une mauvaise solution.

Toutes les techniques d'optimisation décrites précédemment se fondent exclusivement sur les dérivées partielles de premier ordre (jacobien). Dans la littérature sur l'optimisation, on trouve des algorithmes excellents basés sur les dérivées partielles de seconds degrés (héssiens). Malheureusement, ces algorithmes sont très difficiles à appliquer aux réseaux de neurones profonds, car ils ont n^2 hessiens par sortie (où n correspond au nombre de paramètres), à opposer aux seuls n Jacobiens par sortie. Puisque les RNP présentent en général des dizaines milliers de paramètres, la mémoire disponible ne permet pas d'accueillir les algorithmes d'optimisation de second ordre et, même si c'était le cas, le calcul des héssiens serait beaucoup trop long. **La figure 8** présente le résultat de l'implémentation de la méthode Adam.

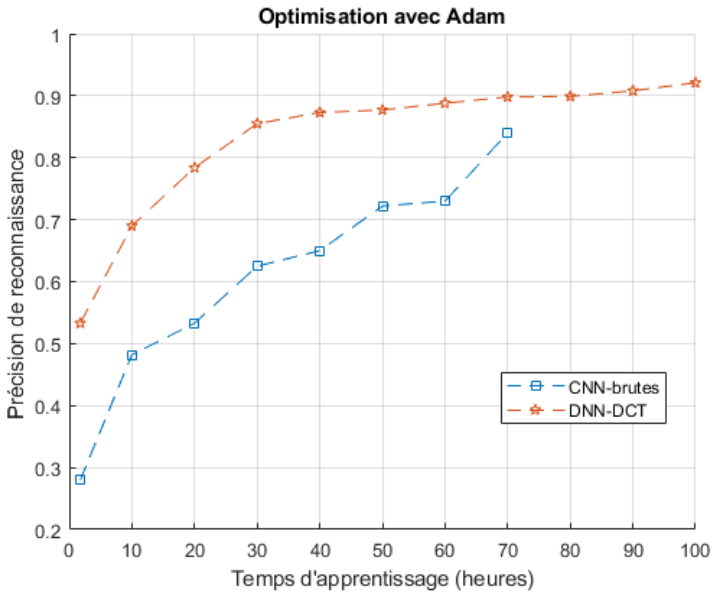


Figure 8 : Evolution du temps d'apprentissage par la méthode Adam.

L'optimisation Adam a permis de reprendre les précisions perdues mais cela reste insuffisant (96% environs). Ainsi, en termes de temps de calcul, ceci reste légèrement moins bonne que les autres méthodes d'optimisation déjà implémentées.

En résumé et afin de respecter le compromis : précision Vs vitesse de convergence, nous avons conservé les paramètres optimaux du **tableau 1.**

Tableau 1 : combinaison optimale d'apprentissage d'un réseau de neurones profonds.

Initialisation	Initialisation de He
Fonction d'activation	SELU
Normalisation	Aucune (grâce à SELU)
Régularisation	Alpha dropout
Optimiseur	Gradient accéléré de Nesterov
Echéancier d'apprentissage	Planification par exponentielle

Ces paramètres permettent de respecter le compromis entre avoir une meilleure précision de reconnaissance et une vitesse d'apprentissage stable.

1.3. Discussion

L'implémentation de ces différentes techniques ainsi que la prise en compte des différents paramétrages et méthodes d'optimisations accélérés ont permis d'expliquer certaines limitations. En effet, ces méthodes d'optimisation permettent soit de garder une performance de classification assez élevée mais ne réduisent pas le temps d'apprentissage d'une manière significative, où bien elles opèrent inversement, c'est-à-dire, elles permettent de réduire le

temps d'apprentissage (jusqu'à 5 heures (le DNN-DCT) et 3 heures 30minutes (le CNN-brutes)) sans un réel impact sur la précision de classification (Adam, la décente de Gradient Stochastique et un peu moins Gradient accéléré de Nesterov).

L'exploration des techniques précédentes inscrites dans **le tableau 4** permet de réduire le temps de calcul à 3h et 53 minutes alors que la précision de classification s'est relativement dégradée (96.3% au lieu de 98.8% initialement). Ceci peut s'expliquer par le fait que, l'utilisation de méthodes telles que la descente de gradient stochastique, Adam ou le Gradient accéléré de Nesterov, nécessite beaucoup d'ajustements manuels. Il est donc nécessaire d'explorer des techniques permettant de bien prendre en compte cet aspect (moins d'hyperparamètres et ajustements manuels).

Ainsi, nos résultats sont comparables à l'état de l'art, notamment les travaux de (Hammerla et al., 2016)[189] où les auteurs ont obtenu une précision de classification de 92% en entrainant un modèle (avec le gradient accéléré de Nesterov) basée sur l'utilisation d'une approche de classification basée sur le modèle LSTM.

Afin de répondre aux limitations de l'optimisation par décente de Gradient (Accéléré de Nesterov et Stochastique) ainsi que les contraintes imposées par la méthode Adam, nous proposons d'utiliser une méthode heuristique tel que l'algorithme d'optimisation de l'essaim de particules (PSO). En effet, contrairement aux méthodes précédentes, cette approche a beaucoup moins d'hyper-paramètres à ajuster, ce qui peut être avantageux pour la vitesse de convergence mais aussi, en l'appliquant directement sur la couche de décision, celle-ci verra sa performance de classification améliorée.

Annexe F

Résultats de classification de quelques activités physiques

Dans cette séction, nous présenterons quelques résultats préliminaires que nous avons effectués pour la détection des activités élémentaires et basées sur le contexte. L'environnement de l'expérience était controlé. L'idée est justement de comparer ces résultats avec les résultats issuent d'un environnement non controlé. Nous avons commencé par comparer deux types de personnes, une personne jeune et une personne agée. L'idée est de classifier quatre type d'activités :marche normale, marche avec claudication, marche lente et marche avec trainement des pieds. Les résultats sont résumés dans les tableaux ci-dessous.

Annexes

1. Prise en compte des axes: X,Y et Z

1.1 XYZ, taille du descripteur 2

démarche	**Normale jeune / autres**		**Normale vieux / autres**		**Claudication / autres**		**Trainement/Lente**	
Matrice de confusion	740	0	527	21	368	3	168	24
	0	184	28	163	2	181	17	160
perf	99,89 %		93,37 %		99.1%		88.89%	

1.2 Axes XYZ, taille du descripteur 4

démarche	Normale jeune / autres		Normale vieux / autres		Claudication / autres		Trainement/Lente	
Matrice de confusion	740	0	528	15	369	0	179	9
	0	184	27	169	1	184	8	175
perf	100 %		94,32 %		99.82%		95.39%	

1.3 Axes XYZ, taille du descripteur 8 :

démarche	Normale jeune / autres		Normale vieux / autres		Claudication / autres		Trainement/Lente	
Matrice de confusion	739	0	547	3	369	0	177	4
	1	184	8	181	1	181	8	180
perf	99,89 %		98,51 %		99.82%		96.75%	

1.4 Axes XYZ, taille du descripteur 16

démarche	**Normale jeune / autres**		**Normale vieux / autres**		**Claudication / autres**		**Trainement/Lente**	
Matrice de confusion	738	0	555	4	369	0	179	2
	2	184	0	180	1	184	6	182

perf	99,78 %	99,46 %	99.82%	97.83%

1.5 Axes XYZ, taille du descripteur 32

démarche	**Normale jeune / autres**		**Normale vieux / autres**		**Claudication / autres**		**Trainement/Lente**	
Matrice de confusion	739	0	554	0	370	1	179	2
	1	184	1	184	0	183	6	182
perf	99,89 %		99,86 %		99.82%		97.83%	

1.6 Axes XYZ, taille du descripteur 48

démarche	**Normale jeune / autres**		**Normale vieux / autres**		**Claudication / autres**		**Trainement/Lente**	
Matrice de confusion	739	0	554	0	369	0	184	1
	1	184	1	184	1	184	1	183
perf	99,89 %		99,86 %		99.82%		99.46%	

1.7 Axes XYZ, taille du descripteur 64

démarche	Normale jeune / autres		Normale vieux / autres		Claudication / autres		Trainement/Lente	
Matrice de confusion	740	0	554	0	370	1	184	1
	0	184	1	184	0	183	1	183
perf	100 %		99,86 %		99.82%		99.46%	

2. Prise en compte de l'axe Y

2.1 Axe-Y, taille du descripteur 2

Démarche	Normale jeune / autres		Normale vieux / autres		Claudication / autres		Trainement/Lente	
Matrice de confusion	715	149	550	184	376	166	164	10

	25	35	5	0	3	18	21	174
perf	81,17 %		74,42 %		69, 49%		91.6%	

2.2 Axe-Y, taille du descripteur 8

démarche	Normale jeune / autres		Normale vieux / autres		Claudication / autres		Trainement/Lente	
Matrice de confusion	710	14	508	91	358	25	182	7
	30	170	47	93	12	159	3	177
perf	95.24 %		81.33 %		93.32%		97.23%	

2.3 Axe Y, taille du descripteur 16

démarche	Normale jeune / autres		Normale vieux / autres		Claudication / autres		Trainement/Lente	
Matrice de confusion	730	15	524	32	369	0	185	11
	10	169	31	152	1	184	0	173
perf	97.29 %		91.47%		99.82%		97.02%	

2.4 Axe Y, taille du descripteur 32

démarche	Normale jeune / autres		Normale vieux / autres		Claudication / autres		Trainement/Lente	
Matrice de confusion	733	5	546	10	369	0	182	5
	7	179	9	174	1	176	3	179
perf	98.7 %		97.43%		98.38%		97.83%	

2.5 Axe Y, taille du descripteur 48

démarche	Normale jeune / autres		Normale vieux / autres		Claudication / autres		Trainement/Lente	
Matrice de confusion	731	3	542	4	370	4	182	8
	9	181	13	180	1	180	3	176

perf	98.7 %	97.7%	99.28%	97.02%

2.6 Axe Y, taille du descripteur 64

démarche	Normale jeune / autres		Normale vieux / autres		Claudication / autres		Trainement/Lente	
Matrice de confusion	726	7	547	6	369	0	180	9
	14	177	8	178	1	184	5	175
perf	97.73 %		98.11%		99.82%		96.21%	

3. Prise en compte de l'axe X

3.1 Axe X, taille du descripteur 2

démarche	Normale jeune / autres		Normale vieux / autres		Claudication / autres		Trainement/Lente	
Matrice de confusion	739	0	541	52	339	65	84	78
	1	184	14	132	31	119	101	106
perf	99.89 %		91.07%		82.67%		51.49%	

3.2 Axe X, tailles du descripteur 4

démarche	Normale jeune / autres		Normale vieux / autres		Claudication / autres		Trainement/Lente	
Matrice de confusion	740	0	526	40	342	66	115	92
	1	184	29	144	28	118	70	92
perf	100 %		90.66%		83.03%		56.1%	

3.3 Axe X, tailles du descripteur 8

démarche	Normale jeune / autres		Normale vieux / autres		Claudication / autres		Trainement/Lente	
Matrice de confusion	736	1	532	45	364	6	126	93
	4	183	23	139	6	178	59	91

perf	99.46 %	90.8%	58.81%	56.1%

3.4 Axe X, taille du descripteur 16

démarche	Normale jeune / autres		Normale vieux / autres		Claudication / autres		Trainement/Lente	
Matrice de confusion	739	0	554	0	369	1	142	60
	1	184	1	184	1	183	43	124
perf	99.89 %		99.86%		99.64%		72.09%	

3.5 Axe X, taille du descripteur 32

démarche	Normale jeune / autres		Normale vieux / autres		Claudication / autres		Trainement/Lente	
Matrice de confusion	739	0	555	4	370	1	136	38
	1	184	0	180	0	183	49	146
perf	99.89 %		99.46%		99.82%		76.42%	

3.6 Axe X, tailles du descripteur 48

démarche	Normale jeune / autres		Normale vieux / autres		Claudication / autres		Trainement/Lente	
Matrice de confusion	739	0	555	0	369	0	153	34
	1	184	0	184	1	184	32	150
perf	99.89%		100%		99.82%		82.11%	

3.7 Axe X, tailles du descripteur 64:

démarche	Normale jeune / autres		Normale vieux / autres		Claudication / autres		Trainement/Lente	
Matrice de confusion	739	3	555	4	369	1	154	40
	1	181	0	184	0	184	31	144

perf	99.57%	100%	99.82%	80.76%

4. Axe XYZ, tailles du descripteur=48, taille de la découpe=128, taille du chevauchement=64:

démarche	Normale jeune / autres		Normale vieux / autres		Claudication / autres		Trainement/Lente	
Matrice de confusion	259	0	194	0	129	7	59	1
	1	64	1	64	1	57	6	63
perf	99.69%		99.61%		95.88%		94.57%	

5. Axe XYZ, tailles du descripteur=48, taille de la découpe=256, taille du chevauchement=128:

démarche	Normale jeune / autres		Normale vieux / autres		Claudication / autres		Trainement/Lente	
Matrice de confusion	259	0	195	0	129	0	61	5
	1	64	0	64	1	64	4	59
perf	99.69%		100%		99.48%		93.02%	

6. Axe XYZ, tailles du descripteur=48, taille de la découpe=512, taille du chevauchement=512:

démarche	Normale jeune / autres		Normale vieux / autres		Claudication / autres		Trainement/Lente	
Matrice de confusion	359	0	268	2	179	1	89	3
	1	89	0	89	0	89	1	86
perf	98.78%		99.44%		99.62%		97.77%	

7. Axe XYZ, tailles du descripteur=48, taille de la découpe=256, taille du chevauchement=256:

démarche	Normale jeune / autres		Normale vieux / autres		Claudication / autres		Trainement/Lente	
Matrice de confusion	279	0	208	2	137	3	65	3
	1	69	0	69	0	69	5	66

perf	99.71%	98.28%	98.56%	94.24%