UNIVERSITÉ PARIS 8
VINCENNES-SAINT-DENIS

Apprentissage Profond Interprétable pour la Reconnaissance d'Activité Humaine Saine et Pathologique à partir de Séries Temporelles Multivariées issues de Capteurs Inertiels

Thèse de doctorat présentée et soutenue par:
Youness EL MARHRAOUI

Encadrant: Prof. Mehdi AMMI
Ecole doctorale n° 224 Cognition, langage, interaction

Abstract

This dissertation investigates the development of interpretable deep learning models for Human Activity Recognition (HAR) using multivariate time series data from wearable inertial sensors. The study primarily targets two critical populations : elderly individuals and pediatric patients, especially within rehabilitation contexts. These populations require continuous monitoring of their physical activities to enhance rehabilitation outcomes, making HAR systems essential for personalized healthcare. The main objectives of this research are to design accurate, interpretable models capable of predicting rehabilitation activities, detecting critical events (such as falls), and providing clinicians with actionable insights.

The first contribution of this work is the integration of self-attention mechanisms into deep learning architectures to improve model interpretability. In healthcare applications, it is essential for clinicians to understand the reasoning behind model predictions, particularly in the detection of critical events like falls. Here, self-attention mechanisms, initially popularized in natural language processing, are adapted for time series data, highlighting the input segments most influential to model decisions. This enhances transparency and trust, crucial for clinical adoption.

The second contribution addresses the challenge of missing data, which frequently occurs in real-time, sensor-based systems. Given the likelihood of data gaps due to signal dropouts or improper usage in elderly patients, this research incorporates self-supervised learning strategies for data imputation. These strategies allow the model to reconstruct missing sensor data segments based on the available data, ensuring system robustness even with partial data, thus improving applicability in real-world scenarios.

Third, this research explores unsupervised and semi-supervised learning approaches to reduce dependence on fully labeled datasets. Labeling rehabilitation activities, particularly for extended monitoring, is time-intensive and requires expert input. Through unsupervised learning, the models can derive meaningful data representations without exhaustive labeling. Additionally, semi-supervised learning fine-tunes the models using a small labeled dataset, enhancing accuracy in detecting and classifying critical rehabilitation activities.

The models were evaluated using two main datasets : the OPPORTUNITY dataset, widely used in HAR research for multimodal sensor data, and the STARR dataset, designed for rehabilitation monitoring. OPPORTUNITY provided a benchmark across diverse activities, while STARR enabled focus on lower-body activities, such as walking phases and post-operative rehabilitation tracking. A combination of accelerometer data from inertial measurement units (IMUs) and video-based skeleton tracking facilitated robust model evaluations.

Results show a 10-15% accuracy improvement over baseline models, especially in detecting critical events and predicting walking phases. Additionally, a 30% error reduction was observed in imputing missing data, making these models highly suitable for real-time clinical applications. The integration of self-attention not only boosted accuracy but also substantially enhanced model interpretability, providing clinicians with clear insights into critical data points in the prediction process. This dissertation marks a significant advancement in HAR for personalized healthcare, particularly in the rehabilitation of elderly and pediatric patients, and lays the groundwork for future deep learning applications in clinical settings.

Keywords : Human Activity Recognition, Deep Learning, Self-Attention, Rehabilitation, Elderly, Pediatric Care, Unsupervised Learning, Semi-Supervised Learning, Missing Data, Fall Detection, Inertial Sensors, Real-Time Applications.

Résumé

Cette thèse se concentre sur le développement de modèles d'apprentissage profond interprétables pour la reconnaissance d'activités humaines (HAR) à partir de séries temporelles multivariées issues de capteurs inertiels portables. Elle s'adresse particulièrement aux besoins de surveillance des personnes âgées et des patients pédiatriques en contexte de rééducation, pour lesquels une évaluation continue des activités physiques est essentielle afin d'améliorer les résultats cliniques. Cette recherche vise à concevoir des modèles précis et interprétables, capables de prédire les activités de rééducation, de détecter des événements critiques (comme les chutes) et de fournir aux cliniciens des informations exploitables.

La première contribution de cette thèse est l'intégration de mécanismes de self-attention dans les architectures d'apprentissage profond, pour améliorer l'interprétabilité des modèles en santé. Ces mécanismes, adaptés des traitements de langage naturel aux données temporelles, permettent de mettre en lumière les segments de données les plus influents dans les décisions du modèle, favorisant une meilleure compréhension clinique des prédictions, notamment pour la détection de chutes.

La seconde contribution aborde le défi des données manquantes, souvent rencontré dans les systèmes de capteurs en temps réel. Des stratégies d'apprentissage auto-supervisé ont été développées pour imputer ces données, assurant ainsi que les modèles restent opérationnels et robustes face aux lacunes de signal, fréquentes chez les populations vulnérables.

En troisième lieu, cette recherche propose des approches d'apprentissage non supervisé et semi-supervisé pour limiter la dépendance aux ensembles de données entièrement étiquetés. Ces méthodes permettent aux modèles d'extraire des représentations utiles des données, même en l'absence d'étiquetage exhaustif, et d'affiner la précision de détection d'activités de rééducation essentielles grâce à un petit sous-ensemble étiqueté.

Enfin, la performance des modèles est évaluée sur deux jeux de données : le jeu de données OPPORTUNITY, pour tester une large gamme d'activités, et le jeu STARR, orienté vers les activités de rééducation des membres inférieurs. Cette évaluation s'appuie sur une combinaison de données d'accéléromètres des unités de mesure inertielle (IMU) et de

suivi squelettique vidéo, offrant une validation robuste et multimodale.

Les résultats démontrent une amélioration de 10 à 15 % de la précision par rapport aux modèles de référence pour la détection d'événements critiques et une réduction de 30 % des erreurs d'imputation des données manquantes. L'intégration des mécanismes de self-attention a non seulement accru la précision, mais aussi l'interprétabilité des modèles, apportant ainsi aux cliniciens des aperçus précieux pour la prise en charge personnalisée en rééducation. Cette thèse représente une avancée significative dans le domaine de la HAR appliquée à la santé, en particulier pour les populations âgées et pédiatriques, et ouvre la voie à de futures applications de l'apprentissage profond en milieu clinique.

Mots-clés : Reconnaissance d'Activité Humaine, Apprentissage Profond, Self-Attention, Rééducation, Personnes Âgées, Patients Pédiatriques, Apprentissage Non Supervisé, Apprentissage Semi-Supervisé, Données Manquantes, Détection des Chutes, Capteurs Inertiels, Applications en Temps Réel.

Remerciements

Ce mémoire de thèse est l'aboutissement d'années de travail, d'apprentissage et de collaboration, et il me tient à cœur de remercier tous ceux qui m'ont soutenu et guidé tout au long de ce parcours.

En premier lieu, je tiens à exprimer ma gratitude envers mon directeur de thèse, Prof. Mehdi AMMI, pour sa confiance, son encadrement et son dévouement. Ses conseils précieux et son expertise ont été des sources d'inspiration et de motivation. Sa rigueur scientifique, combinée à sa patience et son engagement, m'ont permis de progresser et de mener à bien ce projet.

Je souhaite également remercier mes collègues doctorants, avec qui j'ai partagé des moments intenses de travail, des débats stimulants, mais aussi des instants de soutien et de convivialité inestimables. Leur présence et leur amitié ont transformé les défis et les difficultés en expériences enrichissantes. Nous avons partagé des périodes de doutes, de découvertes et d'accomplissements qui ont été marquantes dans cette aventure doctorale. Merci pour leur solidarité, leur écoute et leurs encouragements constants.

À ma famille, dont le soutien inconditionnel m'a permis de mener ce projet jusqu'à son terme, j'adresse mes plus profonds remerciements. À ma mère, en particulier, pour son amour, sa patience et son encouragement indéfectible tout au long de ces années. Son soutien moral et émotionnel a été une source de force durant les moments de doute et d'épuisement, et sans elle, cette réalisation n'aurait pas été possible. Mes pensées vont également à mon père, frère et soeurs, qui m'ont toujours encouragé à poursuivre mes rêves et à donner le meilleur de moi-même.

Enfin, je tiens à remercier également l'organisme EIT Label et le programme BeHealsy, qui m'ont honoré en m'acceptant dans leur programme d'excellence. Leur soutien a été essentiel pour assister mes travaux de recherche, et leur panel de formations avancées en biomédical m'a permis d'approfondir mes compétences et de développer de nouvelles perspectives en adéquation avec mes sujets de recherche. Grâce à ce programme, j'ai eu l'opportunité d'échanger et de collaborer avec des universités et des chercheurs de plusieurs pays européens, une expérience qui a considérablement enrichi mon travail et

mes aspirations scientifiques.

À toutes ces personnes, ainsi qu'à tous ceux qui ont, de près ou de loin, contribué à cette aventure, je tiens à exprimer ma reconnaissance et ma gratitude les plus sincères.

Table des matières

Abstract **i**

Résumé **iii**

Remerciements **v**

Table des matières **vii**

Table des figures **xi**

Liste des tableaux **xiii**

Nomenclature **xv**

1 Introduction **1**

1.1 Défis de la reconnaissance d'activité humaine dans le contexte de la rééducation chez les personnes âgées et les enfants 1

1.2 Contexte de Recherche 2

1.3 Objectifs 3

1.4 Questions de Recherche 4

1.5 Contributions 5

1.6 Plan 6

2 Etat de l'art **9**

2.1 Introduction à la Reconnaissance d'Activité Humaine en Rééducation . . . 10

2.2 Modèles d'Apprentissage Profond pour la HAR 12

2.3 Interprétabilité des Modèles de HAR 13

2.4 Apprentissage Non Supervisé et Auto-Supervisé en RAH 15

2.5 Synthèse des Défis et Perspectives 16

2.6 Conclusion de l'État de l'Art 18

3 Volet 1 : Reconnaissance d'activité humaine pathologique pour la rééducation du bas du corps et l'analyse de la marche 23

3.1 Introduction . . . 23
3.2 Chapitre 1 : Présentation du jeu de données STARR . . . 25
3.2.1 Détails sur le jeu de données STARR . . . 25
3.2.2 Collecte des Données . . . 26
3.2.3 Analyse des Données . . . 28
3.3 Chapitre 2 : Les approches d'apprentissage automatique pour la détection des phases de la marche . . . 31
3.3.1 Introduction . . . 31
3.3.2 Méthodes d'adaptation et de formatage des données temporelles pour la classification des phases de le marche . . . 32
3.3.3 L'impact du formatage des données sur la qualité de détection des phases de marche . . . 33
3.3.4 Conclusion . . . 38
3.4 Chapitre 3 : Reconnaissance d'activité pour l'analyse de la marche et l'évaluation de sa qualité . . . 40
3.4.1 Introduction . . . 40
3.4.2 Modélisation inteprétable pour l'analyse des marqueurs petinents pour l'évaluation de la qualité de la marche . . . 41
3.4.3 Analyse des marqueurs corrélés à la qualité de la marche issues des matrices d'attention . . . 46
3.4.4 Conclusion . . . 54
3.5 Chapitre 4 : Modélisation interprétable pour la reconnaissance des activités et la sélection des signaux d'entrée . . . 57
3.5.1 Introduction . . . 57
3.5.2 Mécanismes d'attention à double niveau pour l'extration des capteurs les plus pertinents . . . 58
3.5.3 Extraction et étude des des signaux d'entrée corrélés avec l'activité de rééducation . . . 62
3.5.4 Conclusion . . . 65
3.6 Chapitre 5 : L'auto-apprentissage pour une détection semi-supervisée des événements de chute . . . 69
3.6.1 Introduction . . . 69
3.6.2 Processus de détection des événements rares grâce à l'auto-apprentissage 70
3.6.3 Quantification des risques de chutes à travers les scores d'attention du modèle auto-entrainé . . . 75

3.6.4 Conclusion 79
3.7 Conclusion 81

4 Volet 2 : Reconnaissance d'activité humaine saine pour la détection de la main manipulatrice et de l'action 87
4.1 Introduction 87
4.2 Chapitre 1 : Présentation du jeu de données OPPORTUNITY++ 89
4.2.1 Détails sur le jeu de données OPPORTUNITY++ 89
4.2.2 Protocole et récolte de données 90
4.2.3 Analyse des données 92
4.3 Chapitre 2 : L'utilisation de l'auto-apprentissage pour le remplissage des données manquantes 95
4.3.1 Introduction 95
4.3.2 Méthode des régions masquées par auto-supervision pour le remplissage des données manquantes 97
4.3.3 Résultats du remplissage des données manquantes par auto-supervision sur la qualité de détection d'activités 103
4.3.4 Conclusion 105
4.4 Chapitre 3 : Reconnaissance de la main manipulatrice lors des activités . . 107
4.4.1 Introduction 107
4.4.2 Méthodes et organisation des données pour une détection agnostique de la main manipulatrice 108
4.4.3 Résultats des approches basées sur les encodeurs pour la détection de la main manipulatrice 112
4.4.4 Conclusion 115
4.5 Conclusion 117

5 Discussion 119

6 Conclusions 123

Bibliographie 127

Table des figures

3.1 Placement des capteurs IMU sur les chevilles des participants 27
3.2 Répartition de la population selon la qualité de la marche (score allant de 1 à 5). 29
3.3 Architecture du modèle ConvLSTM utilisé pour la détection des phases de marche. 33
3.4 Intervalles de confiance en terme de F1-score pour chaque modèle et format de données. 37
3.5 Réseau de neurones pour l'entrée des métadonnées 43
3.6 Réseau de neurones pour l'entrée des séries temporelles 44
3.7 Architecture complète du modèle avec les entrées des métadonnées et des séries temporelles 45
3.8 Taille des partitions (entraînement, test et validation) par classe en termes d'observations (entrée matricielle 3D) 47
3.9 Matrice de confusion de la classification 50
3.10 Moyenne des matrices de poids d'attention pour les entrées des métadonnées 52
3.11 Projection des poids sur les séries temporelles 53
3.12 Distribution de la durée des séquences pondérées selon les étiquettes 54
3.13 Bloc CNN-attention par branche. 60
3.14 L'architecture interprétable proposée pour la détection des activités de rééducation. 61
3.15 Comparaison de l'exactitude des modèles d'apprentissage automatique. . . 63
3.16 Comparaison de l'exactitude des modèles d'apprentissage profond. 64
3.17 Cartes d'attention moyennes selon chaque type d'activité de rééducation pour les branches temporelles avec des tailles de noyaux de 3, 7 et 11 (de gauche à droite). 65
3.18 Cartes d'attention moyennes selon chaque type d'activité de rééducation pour les branches temporelles avec des tailles de noyaux de 3, 7 et 11 (de gauche à droite). 66
3.19 Nombre de patients selon le côté touché par l'hémiparésie. 70

3.20 Score d'évaluation de Berg pour chaque patient. 72
3.21 L'architecture de self-attention spatiale. 73
3.22 Pipeline complet de l'approche d'apprentissage non-contrastif. 74
3.23 Fonctions de perte pour les tâches d'apprentissage non-contrastif et de régression. 75
3.24 Projection des poids sur les signaux IMU pour un événement à risque. (La ligne rouge représente la projection des valeurs élevées des poids). 76
3.25 Projection des poids sur les signaux IMU pour une multitude d'événements à risque avec des poids d'importance différents. (Les lignes rouges représentent la projection des valeurs élevées des poids). 76
3.26 Matrice de confusion pour les instants à risque de chute. 77
3.27 Matrices de confusion pour les différentes durées de fenêtres temporelles (avant référence). 78
3.28 Matrices de confusion pour les différentes durées de fenêtres temporelles (après référence). 78
3.29 Matrices de confusion pour les différentes durées de fenêtres temporelles (intervalle de confiance). 79

4.1 Architecture se l'apprentissage auto-supervisé pour la génération des données manquantes 101
4.2 Matrice de confusion des résultats de classification des activités de locomotion 105
4.3 Architecture de classification pour la détection de la main manipulatrice . 111
4.4 Matrice de confusion pour la détection de la main manipulatrice 113

Liste des tableaux

3.1 Répartition et durée moyenne des enregistrements pour chaque exercice. . . 28
3.2 Statistiques des différentes populations (basées sur le score de Berg) dans le jeu de données . . . 30
3.3 Tableau des scores pour les différentes configurations . . . 34
3.4 Tableau des scores du meilleur modèle pour les différentes activités . . . 36
3.5 Nombre de pas pendant des intervalles de 10 secondes par catégorie . . . 41
3.6 Plage des valeurs des différents hyperparamètres . . . 45
3.7 Scores des différentes architectures concurrentes . . . 49
3.8 Valeurs des hyperparamètres optimisés du modèle . . . 50
3.9 Statistiques sur la durée des séquences pondérées . . . 54
3.10 Nombre d'utilisateurs d'outils de support à la marche. . . . 71
3.11 Hyperparamètres optimisés. . . . 75

4.1 Performance du modèle sur la catégorisation des activités de locomotion . . 104
4.2 Métriques de performances pour les différents ratios du jeu de données d'entrainement . . . 105
4.3 Scores de classification pour la détection de la main manipulatrice . . . 112

Nomenclature

— **HAR** - Human Activity Recognition
— **RAH** - Reconnaissance d'Activité Humaine
— **IMU** - Inertial Measurement Unit
— **STARR** - Smart Technologies for Assisting Rehabilitation and Recovery
— **CNN** - Convolutional Neural Network
— **LSTM** - Long Short-Term Memory
— **ConvLSTM** - Convolutional Long Short-Term Memory
— **NLP** - Natural Language Processing
— **LIME** - Local Interpretable Model-agnostic Explanations
— **SHAP** - SHapley Additive exPlanations
— **HMQA** - Human Movement Quality Assessment

1 | Introduction

1.1. Défis de la reconnaissance d'activité humaine dans le contexte de la rééducation chez les personnes âgées et les enfants

La RAH est un domaine de recherche en plein essor, particulièrement pertinent dans le contexte de la rééducation des populations vulnérables telles que les personnes âgées et les enfants [1]. Ces groupes présentent des défis uniques en raison de leurs besoins physiologiques et cognitifs spécifiques. Chez les personnes âgées, les défis principaux résident dans la détection précoce des risques de chute, la surveillance continue de la mobilité et l'évaluation de l'équilibre au cours du temps [2]. La complexité de ces défis est accentuée par la présence de comorbidités courantes chez les personnes âgées, telles que l'arthrite, les troubles cognitifs ou encore les séquelles d'accidents vasculaires cérébraux [3]. Ces conditions affectent non seulement la démarche mais également l'ensemble des capacités motrices, rendant la reconnaissance précise et fiable des activités humaines particulièrement ardue.

D'un autre côté, les enfants, en particulier ceux en phase de rééducation après des interventions chirurgicales ou des maladies, posent des défis distincts [4]. Leur développement moteur est encore en cours, et les variations dans leurs mouvements peuvent être beaucoup plus imprévisibles que chez les adultes. Les systèmes de HAR doivent donc être capables de s'adapter à cette variabilité et de distinguer les anomalies de développement des comportements normaux liés à la croissance [5]. De plus, l'évaluation de la progression dans la rééducation nécessite des outils qui soient non seulement précis mais aussi suffisamment sensibles pour détecter des améliorations ou des régressions subtiles [6].

Les solutions technologiques actuelles, bien qu'avancées, ne sont pas exemptes de limitations [7]. Les méthodes traditionnelles, telles que les algorithmes basés sur des seuils ou les approches statistiques simples, montrent souvent leurs limites face à la complexité des mouvements et des conditions physiques des patients [8]. De plus, ces méthodes peuvent

manquer de sensibilité et de spécificité, ce qui se traduit par des faux positifs ou des faux négatifs, compromettant ainsi la qualité des soins [9]. L'un des défis majeurs dans ce domaine est donc de développer des systèmes de HAR capables de traiter des données complexes en temps réel, tout en offrant un niveau d'interprétabilité suffisant pour être utilisés de manière fiable par les professionnels de la santé [10].

1.2. Contexte de Recherche

Le contexte de recherche dans lequel s'inscrit cette thèse est marqué par des avancées significatives dans les domaines de l'intelligence artificielle, des technologies de capteurs portables, et de la médecine personnalisée [11]. À mesure que la population mondiale vieillit, les enjeux de santé publique liés à la perte d'autonomie et à la nécessité de soins de longue durée s'intensifient [12]. Les personnes âgées, particulièrement vulnérables aux chutes et aux troubles de la mobilité, représentent un groupe pour lequel les technologies de surveillance et d'intervention doivent être particulièrement adaptées [13].

Dans le cadre de la rééducation, les dispositifs de reconnaissance d'activité humaine sont appelés à jouer un rôle clé dans la surveillance continue des patients, permettant ainsi de détecter précocement les signes de déclin physique ou de complications post-opératoires [14]. Les unités de mesure inertielle (IMU), souvent intégrées dans des dispositifs portables comme les montres connectées ou les orthèses intelligentes, fournissent des données en temps réel sur la cinématique du corps [15]. Ces capteurs, placés sur différentes parties du corps, capturent les mouvements et les postures avec une grande précision. Cependant, la richesse de ces données pose également des défis en termes de traitement et d'interprétation [16].

Les approches traditionnelles de reconnaissance d'activité, qui reposent souvent sur des méthodes statistiques simples ou des algorithmes basés sur des seuils, ont montré leurs limites face à la complexité des données issues des capteurs IMU [7]. Ces méthodes peuvent manquer de sensibilité, particulièrement lorsqu'il s'agit de détecter des variations subtiles dans la démarche ou l'équilibre, ce qui est crucial pour la prévention des chutes chez les personnes âgées [3]. De plus, elles peuvent générer un nombre élevé de faux positifs, conduisant à des interventions inutiles ou à une surcharge de travail pour les cliniciens [9].

Dans ce contexte, les techniques d'apprentissage profond, telles que les réseaux de neurones convolutifs (CNN) et les architectures basées sur la self-attention, émergent comme des solutions prometteuses pour surmonter ces limitations [17]. Les CNN, par exemple, sont capables d'extraire des caractéristiques complexes des signaux IMU, permettant ainsi une reconnaissance plus précise des activités humaines [1]. De plus, l'intégration de mécanismes

de self-attention offre une capacité d'interprétation des décisions prises par les modèles, ce qui est essentiel pour leur application dans des contextes cliniques [18]. Ces modèles peuvent non seulement identifier les moments où le risque de chute est élevé, mais aussi fournir des informations sur les raisons sous-jacentes de ces prédictions, offrant ainsi aux cliniciens des outils pour ajuster les plans de traitement de manière proactive [19].

Le domaine de la rééducation pour les enfants présente des défis similaires, bien que les enjeux soient quelque peu différents [4]. Les enfants en rééducation, notamment ceux souffrant de troubles neuromoteurs ou en convalescence après des interventions chirurgicales, nécessitent une surveillance étroite pour s'assurer que leur développement moteur progresse de manière appropriée [5]. Les systèmes de reconnaissance d'activité doivent être suffisamment flexibles pour s'adapter aux variations dans le développement des enfants, tout en étant capables de détecter des anomalies dans les schémas de mouvement qui pourraient indiquer des complications ou des besoins d'intervention supplémentaires [6].

Par ailleurs, le contexte technologique évolue rapidement avec l'intégration de l'intelligence artificielle dans des systèmes de santé de plus en plus connectés [20]. Les dispositifs portables ne se contentent plus de collecter des données ; ils doivent également interagir avec des infrastructures de santé numérique, où les informations sont partagées et analysées en temps réel pour améliorer la qualité des soins [21]. Dans ce cadre, la capacité des systèmes de reconnaissance d'activité humaine à fonctionner de manière autonome, tout en étant intégrés dans des écosystèmes plus larges de santé numérique, devient un facteur crucial de succès [10].

1.3. Objectifs

Les objectifs de ce travail s'articulent autour de l'optimisation des systèmes de reconnaissance d'activité humaine (HAR) pour la rééducation, notamment auprès des personnes âgées et des enfants, en adoptant une approche raisonnée de l'apprentissage profond. L'accent est mis non pas sur le déploiement du deep learning comme une fin en soi, mais plutôt sur les bénéfices qu'il peut apporter en réduisant la complexité du prétraitement, en minimisant la nécessité d'extraction manuelle de caractéristiques, et en améliorant l'interprétabilité des modèles pour une utilisation clinique.

— **Optimisation des modèles de HAR avec le deep learning :** L'objectif est de concevoir un modèle capable de reconnaître et prédire des activités spécifiques aux contextes de rééducation des personnes âgées et des enfants, tout en tenant compte de leurs différences physiologiques. Grâce à l'apprentissage profond, le modèle devrait offrir des gains en précision et en fiabilité, réduisant la nécessité d'étapes de

prétraitement complexes [4].

— **Accroître l'interprétabilité des modèles de deep learning :** La recherche se concentre sur l'intégration de mécanismes d'attention dans les modèles HAR pour rendre les décisions plus compréhensibles et transparentes. Cette amélioration de l'interprétabilité est essentielle pour les cliniciens, qui doivent pouvoir analyser et interpréter les résultats afin d'adapter les plans de rééducation [19, 18].

— **Explorer des approches non supervisées et auto-supervisées pour les données non étiquetées :** Afin de minimiser le besoin d'annotation exhaustive, ce travail étudiera les techniques d'apprentissage non supervisé et auto-supervisé, telles que l'apprentissage contrastif. L'objectif est de détecter des événements critiques, comme les risques de chute, sans intervention manuelle, et d'améliorer la détection d'activités en tirant parti de méthodes de données masquées, réduisant ainsi les besoins en ressources computationnelles [22, 23].

Ces objectifs traduisent une approche pragmatique et centrée sur l'impact des techniques de deep learning pour les applications de rééducation, visant une utilisation pratique et exploitable dans des environnements cliniques.

1.4. Questions de Recherche

Pour atteindre ces objectifs, plusieurs questions de recherche doivent être explorées :

— **Quelle est la valeur ajoutée des modèles de deep learning** par rapport aux méthodes traditionnelles dans la reconnaissance d'activités humaines dans un contexte de rééducation ? Cette question vise à comparer les performances des nouvelles approches avec celles des techniques plus établies, notamment en termes de précision, de sensibilité et de spécificité [17].

— **Comment les mécanismes de self-attention peuvent-ils améliorer l'interprétabilité** des modèles de HAR, en particulier dans la détection des risques de chute ? Il s'agit ici d'explorer comment ces techniques peuvent rendre les modèles plus transparents et compréhensibles pour les utilisateurs finaux, notamment les professionnels de santé [18].

— **Quelles sont les limites actuelles des systèmes de HAR** pour les personnes âgées et les enfants, et comment les approches d'apprentissage non supervisé peuvent-elles surmonter ces défis ? Cette question examine les obstacles actuels et les opportunités offertes par les nouvelles techniques d'apprentissage pour améliorer la robustesse et la fiabilité des systèmes HAR [22].

1.5. Contributions

Cette thèse apporte des contributions majeures dans le domaine de la reconnaissance d'activités humaines (HAR) pour la rééducation et le suivi de la mobilité de populations vulnérables, telles que les personnes âgées et les patients pédiatriques. Elle propose des avancées à la fois méthodologiques et applicatives en tirant parti de l'apprentissage profond pour améliorer la précision, l'interprétabilité et la robustesse des modèles de HAR dans un contexte médical.

La première contribution consiste en l'intégration de mécanismes de self-attention dans les architectures d'apprentissage profond. Cette innovation vise à rendre les modèles plus interprétables en identifiant les parties des données d'entrée qui influencent le plus les décisions du modèle. Ce niveau d'interprétabilité est crucial en contexte clinique, car il permet aux cliniciens de comprendre les prédictions du modèle, notamment lors de la détection d'événements critiques, tels que les chutes. Les mécanismes de self-attention, adaptés des applications en traitement du langage naturel, permettent ainsi une meilleure transparence, facilitant l'adoption de ces modèles par les professionnels de santé.

La seconde contribution porte sur le développement de techniques de gestion des données manquantes par apprentissage auto-supervisé, répondant aux besoins de robustesse des systèmes en conditions réelles. En raison de la variabilité de la qualité des données provenant des capteurs portables, notamment pour les personnes âgées, les pertes de signal sont fréquentes. Les modèles de cette thèse sont capables d'imputer ces données manquantes en reconstruisant les segments de données indisponibles à partir des informations existantes. Cela garantit que les systèmes de HAR peuvent continuer à fournir des prédictions fiables et stables, même lorsque certaines données sont indisponibles, renforçant ainsi leur efficacité dans des environnements cliniques.

Une troisième contribution réside dans l'exploration des approches d'apprentissage non supervisé et semi-supervisé pour la détection des activités de rééducation et des événements critiques. L'étiquetage manuel des données, en particulier pour des activités spécifiques de rééducation, est un processus coûteux et chronophage. En s'appuyant sur l'apprentissage non supervisé, les modèles développés peuvent extraire des représentations pertinentes des données sans nécessiter un étiquetage exhaustif, tandis que l'apprentissage semi-supervisé permet d'affiner ces représentations en utilisant un petit ensemble de données étiquetées, augmentant la précision de détection des activités clés.

Enfin, cette thèse évalue les performances des modèles proposés à l'aide de deux jeux de données principaux, le jeu de données OPPORTUNITY et le jeu de données STARR. Ces

évaluations démontrent l'efficacité des modèles dans différents contextes de mobilité et de rééducation. Le jeu de données OPPORTUNITY permet de tester la performance des modèles dans des contextes variés, tandis que le jeu STARR se concentre sur la rééducation des membres inférieurs, offrant une validation approfondie des modèles dans des conditions de rééducation réelles. Les résultats obtenus montrent une amélioration de la précision de 10 à 15 % pour la détection des événements critiques, ainsi qu'une réduction de 30 % des erreurs d'imputation des données manquantes, soulignant la pertinence des modèles pour une application en temps réel dans le domaine de la santé personnalisée.

En somme, cette thèse représente une avancée significative en HAR pour la rééducation, offrant des modèles interprétables, robustes et applicables dans des environnements cliniques. Ces contributions posent les bases pour des applications futures en apprentissage profond dans la santé, où la personnalisation et la transparence des systèmes sont essentielles.

1.6. Plan

La structure de ce mémoire est organisée en plusieurs chapitres, chacun apportant une contribution spécifique et développant les concepts clés de la HAR appliquée au domaine de la rééducation.

— **Chapitre 1 : Introduction**
Ce chapitre présente le contexte de la recherche, les enjeux et les défis de la reconnaissance d'activités pour les populations vulnérables, telles que les personnes âgées et les enfants en rééducation. Il introduit les objectifs de la recherche et les questions centrales abordées dans cette thèse, en fournissant une vue d'ensemble des contributions et de l'impact attendu dans le domaine de la santé personnalisée.

— **Chapitre 2 : État de l'art**
Ce chapitre explore les avancées et les défis de la HAR, en particulier dans le contexte de la rééducation. Il couvre les modèles d'apprentissage profond couramment utilisés (CNN, LSTM, et Transformers), l'importance de l'interprétabilité des modèles dans un cadre clinique, et les approches non supervisées et auto-supervisées pour pallier le manque de données étiquetées. La synthèse des défis actuels et des perspectives futures conclut ce chapitre.

— **Volet 1 : Reconnaissance d'activité humaine pour la rééducation du bas du corps et l'analyse de la marche**
Ce premier volet se concentre sur les techniques de HAR appliquées à l'analyse de la marche, à partir de données issues du jeu de données STARR. Il examine en détail

les modèles d'apprentissage profond (CNN, LSTM, ConvLSTM) utilisés pour la détection des phases de marche et l'évaluation des anomalies. Les cartes d'attention sont introduites pour rendre les prédictions des modèles interprétables, en offrant des informations exploitables aux cliniciens pour l'adaptation des programmes de rééducation.

— **Volet 2 : Reconnaissance d'activité pour la détection de la main manipulatrice et de l'action**

Ce second volet élargit l'application de la HAR à la détection des actions de manipulation, en utilisant le jeu de données OPPORTUNITY++. Il présente des techniques d'auto-supervision pour combler les données manquantes et introduit des approches d'apprentissage contrastif pour améliorer la précision sans recourir à un étiquetage exhaustif, particulièrement pertinent pour les données complexes et multimodales de ce domaine.

— **Chapitre 5 : Discussion**

Ce chapitre analyse en profondeur les performances et les limites des modèles développés. Il discute des implications cliniques de chaque méthode et propose des améliorations potentielles, notamment en matière de robustesse et d'interprétabilité des modèles. Les aspects pratiques et les défis rencontrés lors de l'implémentation en conditions réelles sont également abordés.

— **Chapitre 6 : Conclusion**

Ce dernier chapitre résume les contributions principales de la thèse et évalue leur impact sur la HAR appliquée à la rééducation. Il propose des perspectives pour les recherches futures, en mettant l'accent sur le potentiel de ces modèles pour des applications plus autonomes et adaptables dans le domaine clinique.

Ce plan propose une structuration cohérente et progressive des concepts et contributions, guidant le lecteur à travers le développement des idées et les avancées de cette thèse dans le domaine de la reconnaissance d'activités humaines appliquée à la rééducation.

2 | Etat de l'art

L'état de l'art est structuré pour fournir une vue d'ensemble cohérente des techniques et des défis associés à la RAH, avec un accent particulier sur les applications en milieu médical, notamment en rééducation.

Contexte et Enjeux de la HAR : Nous introduisons tout d'abord les enjeux de la HAR dans le domaine de la rééducation, en soulignant son importance pour le suivi des mouvements des patients nécessitant une assistance spécifique. La HAR, facilitée par l'utilisation de capteurs inertiels, se distingue par sa capacité à fonctionner dans des environnements où l'utilisation de caméras est limitée, ouvrant ainsi de nouvelles perspectives dans le suivi médical.

Modèles d'Apprentissage Profond pour la HAR : Cette section explore les modèles d'apprentissage profond, en particulier les réseaux de neurones convolutifs (CNN) et récurrents (RNN), qui sont essentiels pour l'analyse des données temporelles complexes issues des capteurs. Ces techniques permettent d'extraire des caractéristiques riches et pertinentes pour la reconnaissance de mouvements subtils, répondant ainsi aux besoins spécifiques de la rééducation.

Interprétabilité des Modèles : L'interprétabilité des modèles est un aspect clé, surtout dans un contexte clinique. L'intégration de mécanismes d'attention et de modèles explicables contribue à une meilleure transparence des prédictions, facilitant leur adoption par les professionnels de santé. Cette section met en évidence l'importance de rendre les modèles intelligibles afin de garantir une utilisation sécurisée et efficace.

Apprentissage Non Supervisé et Auto-Supervisé : Les approches non supervisées et auto-supervisées, permettant d'exploiter des données non étiquetées, sont explorées pour répondre aux défis posés par la collecte de données annotées en rééducation. Ces méthodes offrent des solutions pour améliorer la flexibilité et la robustesse des systèmes HAR dans des environnements réels, souvent marqués par des données bruitées ou incomplètes.

Synthèse des Défis et Perspectives : Enfin, nous proposons une synthèse des prin-

cipaux défis rencontrés, notamment la variabilité interindividuelle, la gestion des données bruitées et la nécessité d'une interprétabilité accrue. Cette section met en lumière les perspectives futures pour surmonter les limitations actuelles de la HAR dans un cadre clinique, en vue de développer des systèmes plus robustes et adaptés aux environnements médicaux.

Chaque section suit une progression logique qui vise à établir un lien cohérent entre les avancées technologiques et les défis spécifiques de la reconnaissance d'activités humaines dans les applications de santé. Cette structure permet d'explorer les bases théoriques et techniques de la HAR, d'analyser les modèles d'apprentissage adaptés, de mettre en avant l'importance de l'interprétabilité dans un contexte clinique, et de présenter les approches non supervisées pour mieux gérer les données non étiquetées. En conclusion, une synthèse des défis et des perspectives futures est proposée, visant à rendre compte des évolutions nécessaires pour des applications de HAR robustes et adaptées aux besoins du domaine médical.

2.1. Introduction à la Reconnaissance d'Activité Humaine en Rééducation

La reconnaissance d'activité humaine (HAR) a connu un essor considérable au cours des dernières décennies, notamment dans le domaine médical, où elle trouve de nombreuses applications en rééducation des patients. La HAR s'appuie sur des capteurs portables comme les accéléromètres, gyroscopes, et capteurs inertiels pour capter les mouvements du corps de manière continue et en temps réel [24, 25]. Contrairement aux méthodes traditionnelles de suivi en rééducation, basées sur des observations cliniques et des auto-évaluations, souvent influencées par des biais subjectifs, la HAR offre une approche plus objective et précise [26, 27]. En fournissant des données précises, elle permet aux cliniciens de mieux suivre les progrès des patients, notamment ceux souffrant de troubles de la mobilité ou les personnes âgées, qui nécessitent un suivi rigoureux [28, 29].

La technologie HAR permet également une personnalisation accrue des traitements de rééducation. Par exemple, dans les cas de rééducation post-chirurgicale ou de récupération après un accident vasculaire cérébral (AVC), les dispositifs de HAR peuvent surveiller les progrès du patient en continu, offrant des retours immédiats aux cliniciens sur l'efficacité des exercices et des protocoles thérapeutiques [30, 31]. Les données recueillies permettent de proposer des exercices plus adaptés aux capacités de chaque individu, et l'usage des capteurs pour évaluer en temps réel la qualité des mouvements assure un meilleur ajustement

du traitement [32, 33]. Des recherches ont d'ailleurs montré que cette personnalisation favorise une meilleure adhésion des patients aux protocoles de rééducation et contribue à améliorer les résultats thérapeutiques [29, 32].

En outre, la HAR apporte un soutien significatif en matière de prévention et de sécurité des patients. Les capteurs et les systèmes d'apprentissage automatique peuvent détecter des événements critiques tels que les chutes, particulièrement chez les personnes âgées, et évaluer les facteurs de risque liés aux rechutes post-traumatiques [29, 32]. La capacité prédictive de ces dispositifs permet non seulement de renforcer la sécurité des patients en alertant les soignants en cas de chute imminente, mais elle offre aussi la possibilité d'adapter les exercices de rééducation pour minimiser les risques [34]. Cette détection proactive aide les cliniciens à anticiper et à prévenir des incidents potentiellement dangereux, ce qui contribue à un environnement de soins plus sécurisé.

Malgré ses nombreux avantages, la HAR en rééducation présente certains défis. L'un des principaux problèmes est la variabilité interindividuelle dans les mouvements, qui peut être accentuée chez des populations spécifiques comme les personnes âgées ou les enfants, dont les mouvements peuvent être atypiques ou irréguliers [34, 35]. Les algorithmes de HAR doivent donc être capables de distinguer ces mouvements complexes, souvent entachés de bruit, pour éviter des erreurs de classification [36, 37]. La qualité des capteurs et la précision des modèles d'analyse sont donc cruciales pour garantir la fiabilité des systèmes. Des capteurs de mauvaise qualité ou des signaux bruités peuvent compromettre l'efficacité du suivi, et des recherches récentes s'attachent à développer des méthodes pour filtrer et traiter ces données, en combinant des capteurs inertiels à d'autres sources d'information comme les caméras ou le GPS [38, 39].

Les avancées en apprentissage profond et en fusion de données multimodales ouvrent de nouvelles perspectives pour la HAR, permettant d'améliorer encore la précision et la robustesse des systèmes. Par exemple, la fusion de données issues de capteurs multiples (comme les accéléromètres et gyroscopes combinés à des caméras ou des dispositifs GPS) permet d'obtenir une analyse plus riche et complète des mouvements, réduisant ainsi les marges d'erreur [38, 39]. Ces systèmes multimodaux sont particulièrement prometteurs dans les environnements de rééducation, où il est souvent nécessaire de capturer des détails fins des mouvements pour en évaluer la qualité et l'efficacité dans des exercices de rééducation spécifiques [40, 41]. Ainsi, la HAR en rééducation devient un domaine en pleine croissance, capable de transformer les traitements et le suivi des patients grâce à une approche plus intelligente et adaptable.

De plus, les recherches en cours explorent de nouvelles applications de la HAR en réédu-

cation, telles que l'usage de modèles de reconnaissance d'activité pour la détection des signes de régression dans la mobilité du patient, et la gestion à distance des exercices de rééducation, permettant aux patients de suivre des protocoles de rééducation depuis leur domicile sous la supervision des cliniciens via des interfaces connectées [42, 43]. Ces innovations permettent de rendre les soins plus accessibles, en particulier pour les patients dans des zones isolées ou ayant des limitations de déplacement, et de réduire les coûts de soins en diminuant la nécessité des consultations en présentiel.

Enfin, la HAR en rééducation peut contribuer à enrichir les bases de données cliniques, offrant aux chercheurs et cliniciens des informations inédites sur les progrès en mobilité, l'adhésion aux programmes de rééducation, et les prédictions de risques chez des populations variées. Ces données peuvent être utilisées pour former des modèles d'apprentissage automatique encore plus performants, ce qui peut alimenter une boucle d'amélioration continue dans les soins et les technologies de rééducation [44, 45].

2.2. Modèles d'Apprentissage Profond pour la HAR

L'émergence des modèles d'apprentissage profond a considérablement amélioré les performances de la reconnaissance d'activités humaines (HAR) en s'appuyant sur des capteurs portables tels que les accéléromètres et gyroscopes. Les modèles basés sur l'apprentissage profond, notamment les réseaux de neurones convolutifs (CNN) et les réseaux de neurones récurrents (RNN), sont largement utilisés dans la HAR pour extraire automatiquement des caractéristiques à partir des données brutes sans nécessiter d'interventions manuelles complexes [28, 36]. Les CNN, en particulier, sont bien adaptés à la détection de motifs spatiaux dans les séries temporelles multivariées générées par les capteurs [39, 40]. Par exemple, Chen et Xue [39] ont utilisé des CNN pour la reconnaissance d'activités à partir de capteurs inertiels et ont montré que ces réseaux étaient capables de capturer des motifs locaux pertinents pour la classification des mouvements. D'autre part, les RNN, et plus spécifiquement les réseaux LSTM (Long Short-Term Memory), sont utilisés pour capturer les dépendances temporelles à long terme dans les données de mouvement, rendant ces modèles particulièrement efficaces pour la reconnaissance d'activités complexes et continues comme la marche, la course, ou les transitions entre ces activités [46, 36]. De plus, l'association des CNN et LSTM en architectures hybrides permet de tirer parti des forces de chaque modèle pour améliorer encore la précision des systèmes HAR, en combinant l'extraction de caractéristiques spatiales avec la modélisation des relations temporelles [46, 28].

Les architectures avancées telles que les Transformers et les mécanismes d'attention ont

également été appliquées avec succès dans la HAR, permettant non seulement d'améliorer la performance, mais aussi de rendre les modèles plus interprétables [29, 46]. Les Transformers, utilisés initialement dans le traitement du langage naturel, sont de plus en plus adoptés dans la HAR pour capturer des relations à longue distance entre les événements de la série temporelle [46]. Par exemple, Shavit et Klein [46] ont montré que l'utilisation des Transformers pour la HAR à partir de données inertiels permettait d'améliorer les performances tout en réduisant la complexité du modèle par rapport aux méthodes traditionnelles. Par ailleurs, l'introduction des mécanismes d'attention permet aux modèles de se concentrer sur les parties les plus pertinentes des données d'entrée, améliorant ainsi l'interprétabilité des décisions prises par le modèle [29, 36]. Ces avancées en apprentissage profond offrent de nouvelles perspectives pour le développement de systèmes HAR plus robustes et adaptatifs, capables de s'ajuster à la variabilité des données et aux contextes spécifiques tels que la rééducation des patients [27, 31]. Grâce à ces modèles, il est désormais possible de mieux comprendre les schémas de mouvement et de fournir des retours en temps réel aux cliniciens et aux patients, améliorant ainsi la qualité des soins et le suivi des traitements.

2.3. Interprétabilité des Modèles de HAR

L'émergence des modèles d'apprentissage profond a transformé le domaine de la reconnaissance d'activités humaines (HAR), en particulier pour les applications basées sur des capteurs portables tels que les accéléromètres et gyroscopes. En effet, l'apprentissage profond permet d'extraire automatiquement des caractéristiques pertinentes des données brutes, sans nécessiter de processus de prétraitement complexes, ce qui simplifie l'implémentation de systèmes efficaces de HAR [28, 36]. Parmi les modèles populaires, les réseaux de neurones convolutifs (CNN) se distinguent pour leur capacité à détecter des motifs spatiaux dans les séries temporelles multivariées issues des capteurs [39, 40]. Les CNN ont été utilisés avec succès pour classifier des mouvements de base en exploitant leur aptitude à détecter des motifs locaux, rendant ces réseaux idéaux pour des tâches de classification rapide et précise des activités humaines [39].

En parallèle, les réseaux de neurones récurrents (RNN) et leurs variantes comme les réseaux LSTM (Long Short-Term Memory) sont appliqués pour capturer les relations temporelles dans les données de mouvement, s'avérant particulièrement utiles pour des activités complexes et continues telles que la marche ou la transition entre des états physiques [46, 36]. Les LSTM, en particulier, ont l'avantage de modéliser des dépendances temporelles à long terme, un élément crucial pour analyser des mouvements séquentiels

complexes. L'association de CNN et de LSTM en architectures hybrides permet de tirer parti des forces de chaque type de réseau pour améliorer encore la précision et la robustesse des systèmes HAR, en combinant la détection de motifs spatiaux avec une modélisation temporelle avancée [46, 28].

Les architectures de pointe comme les Transformers, initialement développées pour le traitement du langage naturel, trouvent également leur place dans la HAR. En raison de leur capacité à capturer des relations à longue distance dans les séquences, les Transformers se montrent particulièrement efficaces dans l'analyse de séries temporelles de données de capteurs [46]. Shavit et Klein [46] ont illustré l'efficacité des Transformers en HAR, montrant qu'ils permettent d'améliorer les performances tout en réduisant la complexité du modèle, comparativement aux méthodes traditionnelles. Par ailleurs, les mécanismes d'attention intégrés aux Transformers offrent des avantages en termes d'interprétabilité, permettant aux modèles de se concentrer sur les parties les plus pertinentes des données d'entrée pour mieux justifier leurs décisions [29, 36].

L'introduction des mécanismes d'attention dans la HAR a apporté des avantages significatifs. L'attention permet aux réseaux de pondérer différentes parties de l'entrée, mettant en avant les zones les plus informatives pour chaque tâche de classification. Par exemple, pour des activités à haut risque de chute, le modèle peut se concentrer plus intensément sur les segments des données où la perte d'équilibre est détectée, permettant d'améliorer la précision des prédictions [29, 44]. De plus, la combinaison de données issues de capteurs multiples permet de traiter des variations dans les mouvements ou les contextes environnants, un aspect essentiel pour l'adaptation des modèles aux différents besoins des patients [27, 31]. Ainsi, ces approches en apprentissage profond ont permis de rendre les modèles HAR plus robustes et adaptatifs, permettant aux systèmes de s'ajuster à la variabilité des données, notamment en contexte de rééducation.

Les récents progrès dans l'architecture des réseaux profonds, incluant des approches telles que les réseaux de neurones récurrents bidirectionnels (Bi-RNN) et les réseaux résiduels pour les réseaux LSTM, offrent également de nouvelles solutions pour optimiser les modèles HAR. Les réseaux résiduels permettent d'atténuer les problèmes de gradient, ce qui rend les architectures profondes plus stables et améliore leur capacité de généralisation aux différents types de mouvements [47, 48]. Par ailleurs, la combinaison de ces avancées avec des techniques de fusion de données provenant de capteurs multimodaux (par exemple, les signaux inertiels fusionnés avec les flux vidéo ou les données biométriques) améliore la précision des systèmes de HAR et leur capacité à capturer la complexité des schémas de mouvement [40, 49].

En conclusion, les modèles d'apprentissage profond, et en particulier les architectures CNN, RNN, LSTM et Transformers, ont démontré une efficacité accrue dans les applications HAR, permettant d'obtenir une compréhension plus fine des schémas de mouvement humain. En fournissant des retours en temps réel aux cliniciens et aux patients, ces modèles améliorent la qualité des soins et facilitent un suivi précis des traitements de rééducation. Ces progrès technologiques dans les architectures de réseaux de neurones permettent d'imaginer de nouvelles applications pour la HAR, en rééducation et dans d'autres domaines de santé où l'analyse de mouvements détaillés et en temps réel est cruciale [45, 50].

2.4. Apprentissage Non Supervisé et Auto-Supervisé en RAH

L'apprentissage non supervisé et auto-supervisé a connu un essor considérable dans le domaine de la RAH. Ces approches exploitent efficacement de vastes volumes de données non étiquetées, ce qui les rend particulièrement adaptées aux contextes où l'annotation manuelle s'avère coûteuse, chronophage et parfois difficilement réalisable par des experts, tels que le suivi longitudinal de patients en rééducation fonctionnelle ou l'analyse des mouvements quotidiens des personnes âgées à domicile.

L'apprentissage non supervisé permet de découvrir de manière autonome des motifs récurrents, des structures latentes et des régularités cachées dans les signaux sensoriels bruts, sans aucune supervision explicite ni étiquette. Cette capacité intrinsèque est essentielle pour identifier des comportements anormaux ou des événements critiques, tels que les chutes, où les données positives sont rares et les patterns déviants difficiles à anticiper a priori [89]. Des techniques variées, notamment les autoencodeurs (classiques ou variationnels), les réseaux génératifs antagonistes (GAN) et les méthodes de clustering (comme le k-means ou les approches basées sur la densité), ont été appliquées avec succès en HAR. Elles visent à apprendre des représentations latentes riches et compressées, qui peuvent ensuite être exploitées en aval pour des tâches de classification supervisée fine ou de détection d'anomalies en temps réel [80]. Par exemple, l'utilisation d'autoencodeurs empilés (stacked autoencoders) a démontré une efficacité notable dans la réduction de dimensionnalité des données issues de capteurs multi-axes (accéléromètres, gyroscopes, magnétomètres), tout en préservant les caractéristiques temporelles et fréquentielles essentielles à la discrimination fine des activités quotidiennes, comme la marche, la montée d'escaliers ou la station assise [64].

L'apprentissage auto-supervisé, quant à lui, a permis des progrès majeurs en HAR en tirant parti de signaux non étiquetés pour pré-entraîner des modèles robustes et hautement généralisables, même lorsque seules de très faibles quantités de données annotées sont disponibles par la suite. Contrairement à l'apprentissage non supervisé pur, cette famille de méthodes génère automatiquement des signaux de supervision à partir des données elles-mêmes, via des tâches proxy (pretext tasks). Les méthodes contrastives, en particulier SimCLR et BYOL (Bootstrap Your Own Latent), constituent des exemples emblématiques : elles apprennent des représentations discriminantes et invariantes en maximisant la similarité (via une perte contrastive comme NT-Xent) entre plusieurs vues augmentées d'un même signal (par exemple, rotations temporelles, masquage de segments, ajout de bruit gaussien ou permutation de fenêtres), tout en minimisant implicitement ou explicitement la similarité avec des signaux provenant d'instances différentes [91]. Ces techniques se révèlent particulièrement efficaces dans les scénarios réalistes où les données de capteurs inertiels (IMU) sont abondantes grâce au port de dispositifs portables ou embarqués, mais restent rarement labellisées en raison des contraintes d'annotation en conditions écologiques [76].

De plus, en intégrant des données multimodales, telles que les flux de capteurs inertiels (accélération et rotation angulaire) et les séquences vidéo issues de caméras ambiantes ou portées, l'apprentissage auto-supervisé enrichit considérablement les représentations apprises. Il permet de capturer de manière conjointe des informations contextuelles (scène et interactions), temporelles (dynamique du mouvement) et spatiales (posture et trajectoire) complémentaires, ce qui renforce la robustesse du modèle face aux variations intra- et inter-sujets, aux bruits sensoriels ou aux changements d'environnement [59, 96].

Ces avancées ouvrent la voie au développement de systèmes de HAR plus flexibles, évolutifs et adaptables, capables de s'ajuster dynamiquement aux variations des conditions d'acquisition (fréquence d'échantillonnage, placement des capteurs, interférences) ainsi qu'aux besoins spécifiques des utilisateurs finaux, qu'il s'agisse de monitoring médical personnalisé, d'assistance à domicile ou d'applications sportives.

2.5. Synthèse des Défis et Perspectives

Malgré les avancées significatives de la reconnaissance d'activités humaines (HAR) grâce aux progrès en apprentissage profond et à l'utilisation de capteurs portables, plusieurs défis restent à surmonter pour une adoption généralisée et une utilisation optimisée dans le domaine médical, en particulier pour la rééducation. L'un des obstacles majeurs réside dans la variabilité interindividuelle des données de mouvement, particulièrement chez des

populations telles que les personnes âgées ou les patients en rééducation, qui peuvent présenter des schémas de mouvement atypiques ou irréguliers [51, 52]. Cette variabilité rend la généralisation des modèles complexe, et une dégradation des performances est souvent observée lorsque ces systèmes sont appliqués à des groupes de patients aux caractéristiques diverses. La prise en compte de cette variabilité reste un enjeu crucial pour améliorer la robustesse et l'adaptabilité des modèles HAR dans des contextes de rééducation.

La gestion des données bruitées et incomplètes constitue un autre défi central. Les capteurs inertiels, en particulier, sont sujets à des erreurs de calibration et à des décalages, ce qui peut compromettre la précision des modèles HAR [53]. Bien que des techniques de débruitage et d'imputation aient été explorées, elles n'offrent pas toujours une solution complète pour garantir une performance stable en conditions réelles. Les méthodes de correction et de prétraitement de ces données doivent encore être perfectionnées, en intégrant potentiellement des approches de fusion de données pour combiner les informations de capteurs multiples et ainsi minimiser les erreurs [54]. En outre, l'amélioration de la qualité des données recueillies est essentielle pour que les modèles HAR puissent produire des prédictions fiables et précises dans des environnements non contrôlés.

L'intégration de l'apprentissage auto-supervisé et non supervisé représente une opportunité prometteuse pour surmonter la limitation liée à la dépendance aux données étiquetées dans les environnements HAR. Ces approches permettent de réduire la charge de l'annotation manuelle, mais nécessitent des ajustements spécifiques pour être pleinement efficaces dans les contextes médicaux, où la précision et la fiabilité des prédictions sont essentielles [55, 56]. L'exploitation des données non étiquetées dans la formation des modèles peut permettre de capturer des motifs subtils dans les mouvements, tout en augmentant la capacité du système à généraliser dans des contextes variés, ce qui est fondamental pour des applications comme la rééducation personnalisée.

L'interprétabilité des modèles HAR demeure un enjeu majeur pour une adoption clinique de ces technologies. En effet, les cliniciens et thérapeutes ont besoin de comprendre les décisions des modèles pour pouvoir les intégrer en toute confiance dans les processus de soins [57, 58]. À cet égard, les mécanismes d'explicabilité, comme les réseaux de neurones attentionnels et les techniques d'interprétation locale (par exemple, LIME et SHAP), sont des voies prometteuses. Ces techniques permettent de rendre les décisions des modèles plus transparentes, ce qui favorise leur acceptabilité par les professionnels de santé. L'amélioration des approches d'explicabilité est donc cruciale pour renforcer la confiance des cliniciens dans l'utilisation des modèles HAR, en particulier dans des scénarios critiques comme la détection de chutes ou l'évaluation des progrès en rééducation.

Les perspectives futures pour les systèmes HAR en rééducation incluent également l'optimisation de la fusion de données multimodales. En intégrant des informations issues de capteurs inertiels, de vidéos et de données biométriques, il est possible d'obtenir des modèles plus complets et précis, capables de mieux capturer la complexité des mouvements humains [59, 60]. Cette fusion multimodale peut enrichir les systèmes HAR, en leur permettant d'analyser les données sous différents angles, et en offrant une meilleure prise en compte des contextes dans lesquels les mouvements sont réalisés. Cela est particulièrement pertinent dans les environnements de rééducation, où chaque détail de mouvement peut influencer l'évaluation de la progression du patient et l'adaptation de son programme de soins.

D'autres défis incluent l'adaptation en temps réel des modèles aux changements dans le comportement des patients et aux variations des conditions d'acquisition des données. Le développement de modèles plus flexibles, capables de s'adapter dynamiquement aux nouvelles données, est essentiel pour suivre l'évolution des besoins des patients en rééducation. Ces modèles devraient également pouvoir intégrer des méthodes d'apprentissage continu pour s'améliorer au fur et à mesure qu'ils collectent davantage de données, permettant une personnalisation accrue des traitements.

En conclusion, bien que des avancées notables aient été réalisées en HAR pour la rééducation, des défis majeurs subsistent, notamment en ce qui concerne la gestion des données bruitées, la variabilité interindividuelle et l'interprétabilité des modèles. Les approches futures devront se concentrer sur le développement de modèles plus robustes et adaptables, capables d'exploiter des données multimodales et de fournir des explications claires aux professionnels de santé. La recherche en HAR a donc un potentiel significatif pour améliorer la qualité et la personnalisation des soins de rééducation, et pour contribuer de manière plus large aux systèmes de santé intelligents.

2.6. Conclusion de l'État de l'Art

Les avancées récentes en apprentissage profond, en fusion de données multimodales et en interprétabilité de l'intelligence artificielle ont significativement amélioré la reconnaissance d'activités humaines (HAR) pour les applications en rééducation. Ces technologies permettent de surveiller de manière plus précise et objective les mouvements des patients, de personnaliser les programmes de rééducation et de renforcer la sécurité grâce à la détection précoce des événements critiques, tels que les chutes. Les capteurs portables, couplés aux architectures de réseaux neuronaux tels que les CNN, LSTM et Transformers, offrent désormais une analyse détaillée des données de mouvement, ouvrant de nouvelles

possibilités pour le suivi continu et en temps réel de la rééducation.

Cependant, plusieurs défis demeurent pour que ces technologies puissent être pleinement intégrées dans les soins. L'un des principaux obstacles reste la variabilité interindividuelle des données de mouvement, qui complique la généralisation des modèles et limite leur applicabilité à des groupes de patients diversifiés. L'intégration des données bruitées et incomplètes représente également un défi majeur, surtout dans des environnements non contrôlés. Bien que certaines techniques de débruitage et d'imputation aient montré des résultats prometteurs, il reste essentiel de poursuivre les recherches pour garantir une précision stable en conditions réelles.

L'apprentissage non supervisé et auto-supervisé offre des solutions potentielles pour exploiter de grandes quantités de données non étiquetées, réduisant ainsi la dépendance aux données annotées. Cependant, ces approches doivent être adaptées aux exigences des contextes médicaux, où la précision des prédictions est cruciale pour la sécurité des patients. De plus, l'interprétabilité des modèles de HAR demeure essentielle pour une adoption clinique. Les mécanismes d'attention, les techniques d'explicabilité locales comme LIME et SHAP, ainsi que les réseaux de neurones attentionnels, sont des pistes de recherche prometteuses pour offrir aux cliniciens une meilleure compréhension des décisions des modèles.

Les perspectives futures dans la HAR pour la rééducation incluent le développement de modèles plus robustes et adaptatifs, capables de s'ajuster dynamiquement aux conditions de données changeantes et de mieux gérer la variabilité des patients. L'optimisation de la fusion de données multimodales, combinant les informations issues des capteurs inertiels, des vidéos et des données biométriques, permettra d'obtenir une vision plus complète des schémas de mouvement, facilitant ainsi une rééducation personnalisée. Par ailleurs, les modèles d'apprentissage continu pourraient jouer un rôle clé en permettant aux systèmes de s'améliorer avec le temps, en fonction des nouvelles données collectées.

En somme, les technologies de HAR ont un potentiel immense pour transformer les pratiques en rééducation, mais leur adoption clinique nécessite des efforts soutenus pour répondre aux défis liés à la variabilité, à l'interprétabilité et à la gestion des données. Les recherches futures devront se concentrer sur le développement de solutions scalables, robustes et interprétables pour permettre une utilisation généralisée de la HAR dans les environnements médicaux, avec pour objectif ultime d'améliorer la qualité des soins et d'offrir un suivi personnalisé aux patients en rééducation.

Volet 1 : Reconnaissance d'activité humaine pathologique pour la rééducation du bas du corps et l'analyse de la marche

3 | Volet 1 : Reconnaissance d'activité humaine pathologique pour la rééducation du bas du corps et l'analyse de la marche

3.1. Introduction

La reconnaissance d'activité humaine est devenue une composante clé dans le domaine de la rééducation, notamment pour l'évaluation et l'amélioration de la mobilité du bas du corps. L'analyse de la marche est particulièrement cruciale pour les patients souffrant de troubles moteurs, tels que les personnes âgées ou celles ayant subi des blessures ou des interventions chirurgicales au niveau des membres inférieurs [28]. En utilisant des dispositifs portables équipés de capteurs inertiels tels que les accéléromètres et les gyroscopes, il est désormais possible de surveiller en continu la marche des patients et de détecter des anomalies dans leurs mouvements [61]. Ces technologies permettent non seulement de suivre les progrès au cours du traitement, mais également de fournir des feedbacks en temps réel, ce qui est essentiel pour personnaliser les interventions et les ajuster aux besoins individuels [59].

Dans le contexte de la rééducation du bas du corps, la HAR permet de capturer des données précises sur les différents aspects de la marche, comme la cadence, la longueur de pas, l'équilibre et les transitions entre les différentes phases de marche [62]. L'utilisation de modèles d'apprentissage profond, tels que les réseaux CNN et les réseaux LSTM , permet d'analyser ces signaux de manière plus approfondie, en identifiant les caractéristiques spécifiques de la marche qui peuvent indiquer des déficiences ou des améliorations au fil du temps [63]. Par ailleurs, l'analyse automatisée de la marche via des systèmes basés

sur les capteurs offre des avantages en termes de précision et d'objectivité, par rapport aux méthodes d'évaluation clinique traditionnelles souvent subjectives et basées sur des observations manuelles [64].

Malgré ces avancées, des défis persistent, notamment en ce qui concerne l'interprétabilité des modèles et la gestion de la variabilité interindividuelle, qui peut complexifier la généralisation des modèles de HAR à des populations diversifiées [65]. De plus, l'intégration de données multimodales, telles que les flux vidéo et les données des capteurs inertiels, offre de nouvelles opportunités pour améliorer la précision des analyses de la marche, notamment en combinant des informations sur la posture et les mouvements globaux du corps [46]. Ces approches pourraient ouvrir la voie à des systèmes de rééducation plus intelligents et plus personnalisés, capables d'aider les cliniciens à ajuster les programmes de rééducation en fonction des performances spécifiques de chaque patient, tout en assurant une surveillance continue des risques potentiels, comme les chutes ou les déséquilibres [31].

La suite de ce volet est organisée selon le plan suivant : On débutera par une présentation des données utilisées, issues du projet STARR, comprenant des mesures de capteurs inertiels et des données vidéo pour capturer les phases de marche et les paramètres de rééducation. Cette section est suivie d'une explication des méthodes de détection des phases de marche, avec un accent particulier sur les modèles d'apprentissage profond tels que les architectures CNN et LSTM. Ensuite, chaque méthode est analysée pour démontrer son efficacité dans l'identification des anomalies de marche et l'adaptation des programmes de rééducation.

La section suivante explore l'impact des choix de formatage des données, ainsi que l'utilisation des architectures ConvLSTM pour exploiter les dépendances temporelles et spatiales dans les signaux. Enfin, les résultats obtenus sont discutés en détail, mettant en lumière les améliorations potentielles pour des applications cliniques et les limites des modèles utilisés, avec des suggestions pour de futures recherches et applications dans le domaine de la rééducation

3.2. Chapitre 1 : Présentation du jeu de données STARR

Le projet STARR (Smart Technologies for Assisting Rehabilitation and Recovery) a pour objectif de créer un cadre permettant l'évaluation et l'amélioration de la rééducation physique des patients à travers des dispositifs intelligents. Le jeu de données STARR est né de la collecte de données issues de capteurs portables placés sur différentes parties du corps, tels que les accéléromètres, les gyroscopes et les magnétomètres, ainsi que des systèmes de capture de mouvement vidéo. Ces données ont été recueillies auprès de patients en rééducation après des traumatismes ou des interventions chirurgicales du bas du corps, en particulier pour des pathologies liées à la mobilité, comme les troubles de la marche et de l'équilibre. Les participants au projet étaient suivis pendant leur processus de rééducation dans un environnement clinique contrôlé, et les données incluaient des informations sur les phases de marche, les changements de posture, ainsi que les efforts physiques réalisés pendant des sessions d'exercices supervisées [28, 59]. Ce jeu de données a été enrichi par l'ajout de métadonnées concernant l'état clinique des patients, comme les scores de mobilité et les observations des physiothérapeutes, afin d'améliorer la qualité des analyses.

Le jeu de données STARR a servi de base pour les quatre projets de recherche majeurs présentés dans ce chapitre, chacun apportant une contribution spécifique dans l'analyse des activités humaines pour la rééducation du bas du corps. Ces projets partagent le même jeu de données pour assurer la comparabilité des résultats et l'intégration des différentes approches méthodologiques. Le premier projet explore l'utilisation des réseaux de neurones convolutifs (CNN) pour la classification des phases de marche, tandis que le second se concentre sur l'emploi des réseaux LSTM pour capturer les dépendances temporelles complexes dans les séquences de mouvement [64, 65]. Le troisième projet s'appuie sur des méthodes d'apprentissage auto-supervisé afin d'exploiter les données non étiquetées dans la détection des anomalies de la marche [46]. Enfin, le quatrième projet utilise une approche multimodale en combinant des données de capteurs inertiels et vidéo pour améliorer la précision de la détection des phases critiques de la rééducation [31]. Grâce à l'uniformité des données issues du projet STARR, les résultats de ces contributions offrent une vision cohérente et intégrée des avancées en matière de HAR pour la rééducation.

3.2.1. Détails sur le jeu de données STARR

Le jeu de données STARR (Smart Technologies for Assisting Rehabilitation and Recovery) est conçu pour permettre une analyse complète et détaillée des activités humaines, avec un

accent particulier sur la rééducation du bas du corps et l'analyse de la marche. Le protocole d'inclusion des patients dans cette étude est rigoureux : il comprend des personnes ayant subi des traumatismes ou des interventions chirurgicales sur les membres inférieurs, ainsi que des patients souffrant de troubles de la marche liés à l'âge ou à des pathologies spécifiques, telles que l'arthrite ou les suites d'un AVC. Les patients inclus dans l'étude ont été évalués à plusieurs étapes de leur rééducation, ce qui permet de suivre l'évolution des performances motrices dans le temps. Le protocole exige également que les patients soient en mesure de marcher sur une distance déterminée, avec ou sans aide, afin de permettre une acquisition correcte des données de marche.

Le matériel utilisé pour la collecte des données inclut une série de capteurs inertiels portables (accéléromètres, gyroscopes et magnétomètres) placés sur différentes parties du corps, notamment les chevilles, les genoux et les hanches des patients. Ces capteurs permettent de capturer les mouvements de manière détaillée en trois dimensions (axes x, y, z) et sont synchronisés avec des caméras vidéo haute résolution qui enregistrent les séquences de marche des patients. Le dispositif inclut également des systèmes de capture de mouvement basés sur la vision, qui offrent une vue d'ensemble de la posture et des gestes des patients pendant leurs exercices de rééducation. De plus, des capteurs de pression sont intégrés aux chaussures des patients pour analyser la répartition du poids et les phases d'appui pendant la marche.

Les types de données récoltées sont variés et permettent une analyse exhaustive de la marche. Les données brutes des capteurs inertiels fournissent des informations sur l'accélération, la vitesse angulaire et les changements d'orientation. Les capteurs de pression, quant à eux, fournissent des informations sur l'intensité et la répartition des forces au niveau des pieds lors des différentes phases de la marche. Ces données sont ensuite segmentées pour isoler les cycles de marche et identifier les anomalies ou les améliorations des performances au fil des séances de rééducation.

3.2.2. Collecte des Données

Les participants de cette étude ont été sélectionnés parmi les patients du département de neurologie du centre de rééducation Jacques Calvé, situé à la Fondation Hopale à Berck-sur-Mer, en France. L'étude s'est concentrée sur des individus en phase chronique post-AVC, afin de constituer une cohorte représentative de cette étape spécifique du processus de rééducation. Le recrutement a été élargi pour inclure des patients en hospitalisation complète ou journalière dans le centre de rééducation. En outre, d'anciens patients résidant à proximité du centre ont également été inclus pour enrichir la cohorte. Cette

stratégie de recrutement exhaustive a permis de conduire une étude monocentrique sur une période d'un an, impliquant un total de 26 participants. La diversité de cette population, regroupant des individus à différents stades de la récupération post-AVC, contribue à la richesse et à l'applicabilité des résultats obtenus. Les participants ont été informés et ont donné leur consentement écrit. Leur participation a été approuvée par le Comité de Protection des Personnes (CPP) Sud-Est V (IRB : 6705, 22 mars 2018), garantissant le respect des normes éthiques en matière de collecte de données et de conduite de la recherche.

Dans le cadre du projet STARR, un dispositif innovant de surveillance de l'activité physique a été conçu et développé par une équipe d'experts du laboratoire CEA-LIST, à Palaiseau, France. Ce dispositif combine une Unité de Mesure Inertielle (IMU) et une semelle instrumentée équipée de capteurs de pression, le tout intégré dans un boîtier électronique conçu pour l'acquisition des signaux. Le système fonctionne en plaçant la semelle instrumentée à l'intérieur d'une chaussure standard, tandis que l'IMU et le boîtier électronique sont fixés à la cheville pour une capture optimale des données. Les capteurs de pression mesurent les forces exercées au niveau des pieds, tandis que l'IMU capture les accélérations et les vitesses angulaires. Ce système utilise un taux d'échantillonnage de 40 Hz et une échelle complète de ± 160 m/s^2 pour l'accéléromètre et de ± 2000 deg/s pour le gyroscope, bien que les mesures spécifiques observées se situent autour de ± 40 m/s^2 pour l'accéléromètre et de ± 260 deg/s pour le gyroscope. Cette configuration permet une surveillance détaillée et complète des activités physiques, offrant une compréhension holistique des mouvements et des variations de pression pendant les différentes phases de la marche et de la rééducation.



FIGURE 3.1 – Placement des capteurs IMU sur les chevilles des participants

3.2.3. Analyse des Données

Dans le cadre de cette investigation, portant sur des individus présentant des troubles de la marche, l'analyse des données est essentielle pour évaluer la mobilité fonctionnelle et l'équilibre des participants. L'outil d'évaluation principal utilisé est le test de Berg [54], largement reconnu pour évaluer les capacités d'équilibre des individus, en particulier ceux présentant des troubles de la marche. Pour cette étude, sept activités de rééducation majeures ont été adoptées : marche de 6 minutes, position assise sans support, station debout pieds joints, lever et marcher, position sur un pied, montée d'escaliers, et allongement vers l'avant. Ces activités représentent le spectre des actions que nous cherchons à identifier à l'aide de nos modèles d'apprentissage profond. La variation des volumes de données entre ces catégories est attribuée aux différences de durée des activités respectives, comme détaillé dans le tableau 3.1. Ce tableau présente également la distribution des différentes classes d'activités.

TABLE 3.1 – Répartition et durée moyenne des enregistrements pour chaque exercice.

Activité	Durée		Distribution (%)
	Moyenne	Écart type	
Marche 6 minutes	6min 31s	13s	52.3%
Position assise sans support	49s	12s	17.6%
Debout pieds joints	1min 02s	3s	8.7%
Se lever et marcher	50s	29s	6.6%
Position sur un pied	53s	31s	6.5%
Monter les escaliers	46s	27s	5.3%
Allongement vers l'avant	1min 11s	6s	3%

Les participants ont été catégorisés en cinq groupes distincts, en fonction de l'évaluation réalisée par les professionnels de santé, qui ont mesuré la gravité de leurs troubles de la marche. Cette évaluation s'appuie sur une échelle allant de 1 (faible mobilité) à 5 (mobilité normale), permettant une classification précise des différents niveaux de troubles observés. La population est divisée en ces cinq catégories, comme illustré dans la Figure 3.2, reflétant la diversité des mouvements des participants.

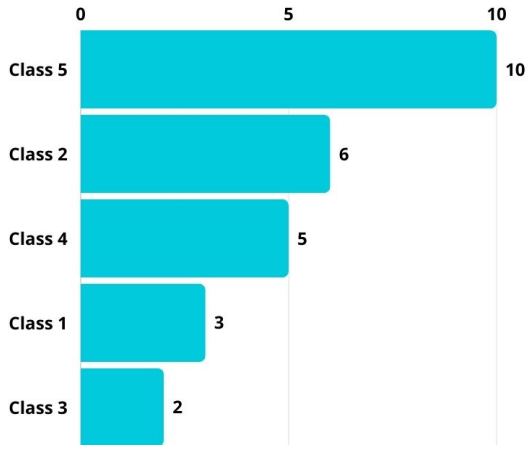

FIGURE 3.2 – Répartition de la population selon la qualité de la marche (score allant de 1 à 5).

Table 3.2 présente un aperçu de la population étudiée, avec chaque statistique exprimée sous la forme moyenne±écart-type. Outre l'âge, la taille, le poids et d'autres indicateurs, la table inclut le score de Berg, un test d'évaluation de l'équilibre basé sur une série d'exercices physiques. La base de données STARR se compose de patients présentant divers troubles de la marche, avec un accent particulier sur ceux montrant des trajectoires de récupération favorables. De plus, au cours de la période de 6 minutes, les participants ont parcouru une distance moyenne de 290 m ± 143 m, avec des distances allant de 38 m à 540 m. Il est important de noter que les participants utilisaient différents types de supports de marche, certains marchant sans aide tandis que d'autres s'appuyaient sur un ou deux dispositifs d'assistance, en particulier ceux avec des scores d'évaluation de la marche plus faibles.

Variable	Score 1	Score 2	Score 3	Score 4	Score 5
Nombre de patients (sexe : H/F)	3(F=3)	7(H=5, F=2)	2(H=1, F=1)	4(H=4)	9(H=5, F=4)
Hémiparésie gauche/droite	G=1, D=2	G=3, D=4	D=2	G=2, D=2	G=4, D=5
Âge (années)	58.66 ± 24.54	61.85 ± 11.48	57.50 ± 9.19	64.75 ± 12.36	59.44 ± 10.81
Taille (cm)	154.0 ± 2.6	170.7 ± 6.8	173.0 ± 14.1	181.2 ± 4.9	170.3 ± 5.8
Poids (kg)	61.6 ± 10.4	75.1 ± 21.7	81.5 ± 3.5	85.5 ± 9.0	77.0 ± 16.1
Période post-AVC (mois)	56.3 ± 89.7	19.5 ± 37.3	9.5 ± 4.9	45.7 ± 56.0	5.4 ± 7.4
Durée de l'exercice (minutes)	6.5 ± 0.2	6.0 ± 0.5	6.3 ± 0.1	6.2 ± 0.2	6.2 ± 0.3
Score de Berg	40.3 ± 9.4	47.1 ± 2.0	48.0 ± 4.2	50.0 ± 4.9	53.5 ± 3.0

TABLE 3.2 – Statistiques des différentes populations (basées sur le score de Berg) dans le jeu de données

3.3. Chapitre 2 : Les approches d'apprentissage automatique pour la détection des phases de la marche

3.3.1. Introduction

La marche est une activité quotidienne que la plupart des individus accomplissent sans difficulté. Toutefois, elle diffère d'une personne à l'autre [66] et peut être classée selon la manière dont le pied entre en contact avec le sol (talon, milieu du pied, avant-pied). D'autres aspects, tels que la vitesse de la marche, le nombre de pas par mètre, et la variabilité du temps de balancement, caractérisent également le processus de marche. Ces indicateurs ont été introduits par [67] pour comparer différentes démarches et offrir des points de distinction clairs, notamment pour les personnes âgées. Ces dernières sont particulièrement vulnérables aux troubles de la marche, qui peuvent nuire à leur posture et leur équilibre [68].

Face à cette réalité, des outils tels que les cannes ou d'autres formes de soutien deviennent indispensables pour aider ces personnes à se déplacer, impactant ainsi leur démarche. Dans ce cadre, la distinction entre les participants de notre étude repose sur deux facteurs principaux : le score de Berg, qui évalue la capacité d'une personne à marcher et le degré de difficulté qu'elle rencontre, ainsi que l'usage de dispositifs d'assistance. À partir de ces critères, les professionnels de santé proposent des exercices adaptés pour améliorer la démarche des patients. Cependant, pour des recommandations plus personnalisées, ils ont besoin de plus d'indicateurs liés à la marche des patients.

Afin de répondre à cette demande, ce chapitre vise à détecter les deux principales phases de la marche, marquées par le contact du pied au sol (touchdown) et son retrait (takeoff). Ces phases permettent d'extraire des informations essentielles telles que la largeur de la foulée, la vitesse de marche, et le temps de balancement. La modélisation de ces mouvements est un sujet d'étude de plus en plus courant, et différentes méthodes ont été appliquées pour détecter les séquences où le pied touche ou quitte le sol. Ce travail est crucial pour des applications telles que la reconnaissance d'activités humaines (HAR) et est largement utilisé dans le domaine médical pour évaluer l'évolution de la marche [69, 70].

3.3.2. Méthodes d'adaptation et de formatage des données temporelles pour la classification des phases de le marche

Cette section décrit la méthode employée pour la détection des phases de contact du pied avec le sol au cours de la marche humaine. Nous décrivons ici l'acquisition des données, le prétraitement, les méthodes de représentation, ainsi que l'architecture d'apprentissage profond ConvLSTM.

Les variables cibles ont été générées à partir des données de pression issues de semelles connectées, en identifiant les moments de contact avec le sol à l'aide de seuils calculés. Ces seuils ont été ajustés et post-traités pour réduire le bruit et les anomalies. Les enregistrements sont collectés via deux capteurs IMU placés autour des chevilles des participants, capturant l'accélération et la vitesse angulaire pour les pieds gauche et droit, résultant en 12 composantes (accélération et vitesse angulaire sur les axes X, Y et Z).

Nous avons exploré trois configurations principales pour le format des données, selon leur adéquation aux différentes architectures de réseaux neuronaux.

- **Matrice bidimensionnelle (WS, 6)** : Dans ce format, chaque fenêtre de taille WS (window size) est une matrice avec 6 colonnes représentant les composantes de l'accélération et de la vitesse angulaire pour les axes X, Y, Z des deux pieds. Ce format est bien adapté aux architectures CNN classiques, qui exploitent les corrélations entre ces six dimensions.
- **Entrées multiples pour chaque composante axiale** : Cette configuration fournit une série temporelle individuelle pour chaque composante du signal (par exemple, l'accélération ou la vitesse angulaire) dans des branches distinctes du modèle. Cela permet au modèle d'apprendre des motifs spécifiques à chaque composante.
- **Deux canaux (WS, 3, 2)** : Ce format ajoute une profondeur aux données, similaire à une image RGB en vision par ordinateur. Les trois axes de l'accélération et de la vitesse angulaire sont séparés en deux canaux : un pour l'accéléromètre et un pour le gyroscope. Cette approche capture les relations spatiales entre ces deux capteurs.

Le modèle principal que nous avons utilisé est basé sur une architecture **ConvLSTM**, qui combine des réseaux convolutifs (CNN) et des réseaux à mémoire longue courte (LSTM). Cette architecture est conçue pour capturer à la fois les dépendances spatiales et temporelles des séries multivariées.

Couches ConvLSTM : Les couches ConvLSTM appliquent des convolutions dans l'es-

pace tout en maintenant la dynamique des états cachés et des cellules de LSTM. Elles permettent de traiter les données séquentielles tout en préservant les motifs spatiaux, une caractéristique importante pour les signaux multivariés comme ceux des capteurs IMU.

Couches de déconvolution : Pour améliorer la représentation spatiale et temporelle des caractéristiques extraites par les couches ConvLSTM, une couche de déconvolution est appliquée. Cela permet un sur-échantillonnage des caractéristiques pour reconstruire les motifs d'origine après leur transformation par les couches ConvLSTM.

L'architecture complète, composée de deux couches ConvLSTM suivies de couches de déconvolution et d'une couche de décision finale, est illustrée dans la Figure 3.3.

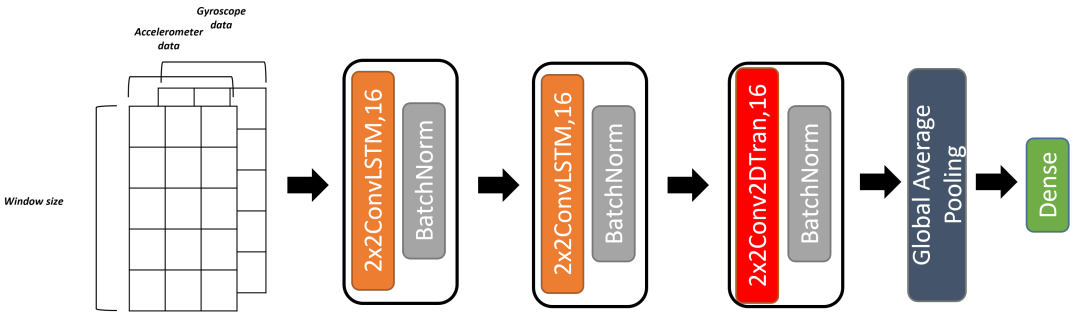

FIGURE 3.3 – Architecture du modèle ConvLSTM utilisé pour la détection des phases de marche.

L'architecture ConvLSTM proposée présente un compromis intéressant entre complexité et capacité de généralisation. Elle tire parti des couches convolutionnelles pour exploiter les motifs spatiaux tout en utilisant les LSTM pour capturer les dépendances temporelles dans les données de capteurs. Grâce à cette approche, nous avons pu traiter efficacement des séries temporelles multi-variables et obtenir des résultats prometteurs dans la détection des phases de marche, comme discuté dans les sections suivantes.

3.3.3. L'impact du formatage des données sur la qualité de détection des phases de marche

Puisque nous suivons le protocole d'évaluation LOSO, chaque modèle sera évalué en fonction de sa performance dans la détection du contact du pied avec le sol pour le même patient, à savoir le patient n° 27. Pour notre meilleur modèle, nous fournirons plus de détails sur sa performance pour différentes activités afin mesurer à quel point il est adapté aux différents scénarios de mouvements et de tester sa capacité agnostique aux données descriptives des patients.

Ces résultats incluent les métriques d'évaluation pour chaque combinaison testée. Ainsi,

72 modèles neuronaux ont été entraînés et évalués, et seront présentés dans la première partie de cette section. Nous montrerons également les résultats détaillés du meilleur modèle afin d'avoir un aperçu de sa performance pour différents types d'activités et pour chaque pied.

Pour mesurer l'efficaté des modèles multi-inputs, nous avons entraîné plusieurs modèles de chaque catégorie en utilisant différentes transformations de données afin de mesurer l'impact du traitement des données. Ensuite, nous avons sélectionné plusieurs métriques pour les comparer, telles que : Précision, F-score, Rappel, Précision positive et Spécificité. Le tableau suivant présente ces scores pour chaque modèle entraîné.

$$\text{Précision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3.1}$$

$$\text{Rappel} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.2}$$

$$\text{F}_1 = 2 \cdot \frac{\text{précision} \cdot \text{rappel}}{\text{précision} + \text{rappel}} = \frac{\text{TP}}{\text{TP} + \frac{1}{2}(\text{FP} + \text{FN})} \tag{3.3}$$

$$\text{Spécificité} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{3.4}$$

Note : TP, TN, FP et FN représentent respectivement les Vrais Positifs, les Vrais Négatifs, les Faux Positifs et les Faux Négatifs.

WS	Transformation	Architecture	Acc	F1	Rec	Pre	Spe
3	None	**ConvLSTM-Deconv**	**97.01%**	**0.965**	**0.969**	**0.961**	**0.971**
		Multi-input LSTM	96.97%	0.964	0.968	0.961	0.970
		Multi-input CNN	96.95%	0.964	0.968	0.961	0.970
		Two channels CNN	96.91%	0.964	0.974	0.954	0.965
		Deep CNN	96.87%	0.964	0.978	0.950	0.961
		LSTM	96.62%	0.960	0.965	0.956	0.967

TABLE 3.3 – Tableau des scores pour les différentes configurations

En tenant compte de l'accuracy des différentes architectures, notre modèle composé de couches ConvLSTM et d'une couche de déconvolution a obtenu de meilleurs résultats que

les autres modèles. Cela montre que ce modèle est adapté à l'apprentissage sur des données brutes, car les performances chutent facilement lorsqu'une transformation est appliquée. De plus, il bénéficie de l'extraction de motifs spatiaux et temporels entre les signaux ainsi qu'entre chaque pas de temps de chaque signal. La représentation des données est également un atout spécial qui met en évidence les corrélations entre les deux types de signaux (données de l'accéléromètre et du gyroscope). Les autres scores montrent que pour d'autres tailles de fenêtre, notre modèle et le modèle LSTM multi-entrées sont les meilleurs, ce dernier pouvant surpasser le modèle ConvLSTM. Il a également obtenu le même meilleur score en utilisant des données transformées par DCT, mais comme l'objectif est de créer un modèle qui s'entraîne bien sur des données brutes, notre modèle reste le meilleur dans cette catégorie.

D'autres architectures neuronales basées sur les couches LSTM [71, 72, 73] ont également très bien performé sur notre ensemble de données, telles que le LSTM multi-entrées suivi de couches DNN. En dehors des configurations avec des transformées FFT, ce modèle a obtenu un classement élevé dans la plupart des cas. En effet, cette architecture est particulièrement adaptée aux séries temporelles, surtout lorsque nous traitons plusieurs séries temporelles dépendantes. Le modèle apprend des motifs temporels à partir de chaque signal en les traitant de manière indépendante, puis les sorties des couches LSTM sont aplaties et concaténées, et le vecteur résultant bénéficie d'un apprentissage supplémentaire par les couches denses.

Nous pouvons également observer que plus la taille de la fenêtre diminue, plus les modèles deviennent précis. Cela est principalement dû au fait que la taille de la fenêtre joue un rôle majeur dans la création de cibles plus ou moins précises [74, 75]. En réduisant la taille de la fenêtre, nous réduisons également la marge d'erreur en ayant moins de valeurs à choisir dans une seule fenêtre de la variable cible, ce qui permet au modèle d'apprendre les motifs de différentes tranches du signal contenant moins de contradictions. Ces contradictions peuvent résulter de tranches de données similaires ou très proches, mais en raison de la condition de fréquence, les cibles peuvent avoir des valeurs différentes. Par conséquent, le modèle améliore ses performances avec l'augmentation de la granularité des données en obtenant une étiquette pour des fenêtres de données plus petites. Cependant, il y a un compromis à prendre en compte lors de l'exécution de ce type d'approche. En termes d'effort computationnel et de temps, la diminution de la taille de la fenêtre peut demander plus de ressources et de temps pour s'exécuter, que ce soit pour le formatage des données, l'entraînement du modèle et la prédiction. Ainsi, en fonction des ressources disponibles et du temps nécessaire, nous devrons choisir une taille de fenêtre adaptée tout en nous assurant que la précision correspondante reste acceptable.

En plus des modèles d'apprentissage profond entraînés, nous avons également entraîné des classificateurs Random Forest et XGBoost pour une comparaison à grande échelle. Ces modèles ont également été formés sur des données fenêtrées (brutes, transformées avec FFT et DCT) et avec différentes valeurs de taille de fenêtre. Cependant, ces modèles n'acceptent pas les matrices 2D en tant qu'entrée unique, nous avons donc utilisé l'ACP (Analyse en Composantes Principales) pour réduire la dimension de la matrice d'entrée et obtenir un vecteur unidimensionnel. Lors de l'entraînement des modèles RF et XGBoost pour les différents ensembles de configurations, nous avons observé le même comportement que pour les modèles de réseaux neuronaux, à savoir que la précision de détection des phases de contact pied-sol augmentait en réduisant la valeur de la taille de la fenêtre. Cependant, cette amélioration a atteint une limite pour des valeurs inférieures à 6, la précision n'a plus progressé et le score de précision le plus élevé enregistré était de 95,43%. Le même constat s'applique au modèle DNN que nous avons conçu pour la même représentation des données et que nous avons testé pour les mêmes transformations différentes et les valeurs de taille de fenêtre, il a également stagné avec une meilleure précision de 94,46%.

Connaissant que l'ensemble de données que nous avons utilisé pour notre étude comprend des enregistrements pour différents types d'activités, nous avons choisi d'exécuter le processus d'évaluation pour chaque activité et chaque pied afin d'avoir une meilleure idée de la performance du modèle dans différentes circonstances. Le tableau 3.4 présente les valeurs des métriques choisies pour chacune de ces configurations.

Activité	Pied	Acc	F1	Recall	Precision	Spécificité
6minMarche	Gauche	96,18%	0,973	0,974	0,972	0,929
	Droit	94,78%	0,965	0,958	0,972	0,879
LeverEtMarcher	Gauche	94,87%	0,895	0,907	0,883	0,961
	Droit	94,62%	0,906	0,931	0,882	0,951
Step	Gauche	97,00%	0,937	0,915	0,959	0,987
	Droit	95,95%	0,905	0,919	0,891	0,970
1minJoint	Gauche	99,65%	0,727	0,666	0,800	0,998
	Droit	99,39%	0,764	0,896	0,666	0,995
1piedÉquilibre	Gauche	99,15%	0,700	0,736	0,666	0,994
	Droit	98,79%	0,451	0,700	0,333	0,990
2minBassin	Gauche	99,98%	0,800	1,000	0,666	0,999
	Droit	99,82%	NaN	NaN	NaN	0,998

TABLE 3.4 – Tableau des scores du meilleur modèle pour les différentes activités

Ce tableau reflète la performance du modèle dans chaque type d'activité. Certaines de ces activités présentent des données déséquilibrées en raison de la nature de l'exercice pratiqué, ce qui peut influencer considérablement les scores. En revanche, les trois premiers

exercices sont plus susceptibles de se produire dans la vie réelle, car ils incluent la marche ou du moins l'alternance entre les pieds gauche et droit. De plus, pour la dernière activité, la personne garde les deux pieds au sol presque tout le temps (sauf en cas de perte d'équilibre), donc nous obtenons souvent une seule classe, qui est "0", ce qui rend impossible le calcul de certaines métriques en raison de l'absence de vrais positifs et parfois de faux positifs également. En outre, pour les trois activités principales, ces scores supplémentaires sont assez similaires aux scores globaux et ne chutent pas pour une activité principale ou une autre.

Un autre aspect important de cette étude est l'analyse des intervalles de confiance de nos modèles en fonction des différentes configurations adoptées. D'une part, cela nous permettra d'étudier à quel point les prédictions des différents modèles sont certaines. D'autre part, cela nous donnera une idée de la cohérence de chaque modèle et du degré de reproductibilité pour chaque configuration. Pour cela, nous avons lancé un nouvel ensemble d'expériences où chaque modèle a été entraîné 5 fois pour chaque configuration unique. Bien que nous ayons limité notre portée d'expériences à une taille de fenêtre de 3, puisque la taille de la fenêtre est un atout majeur pour de meilleures prédictions. La figure 3.4 montre les intervalles des scores obtenus pour différentes configurations.

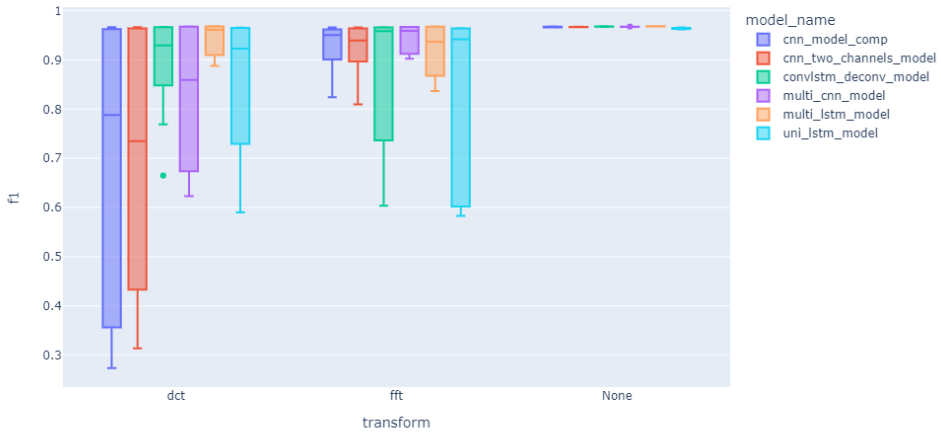

FIGURE 3.4 – Intervalles de confiance en terme de F1-score pour chaque modèle et format de données.

Le figure 3.4 contient des boîtes à moustaches de la métrique d'évaluation F1-score. Ce graphique a une caractéristique assez visible : pour les données brutes, l'intervalle de confiance est très petit, ce qui signifie que l'entraînement du même modèle sur la même configuration plusieurs fois donne le même résultat. Les scores ne baissent pas d'une expérience à l'autre, ce qui conduit à un modèle reproductible avec une grande confiance

d'obtenir la bonne sortie. En revanche, dans les cas où les transformations FFT et DCT ont été appliquées aux données d'entrée, tous les scores présentent de larges marges d'erreur, rendant le modèle moins confiant dans ses prédictions et peu susceptible de reproduire les mêmes résultats s'il est entraîné à nouveau. Sur la base de ces chiffres, des modèles comme le ConvLSTM et le LSTM multi-canal ont obtenu les meilleurs scores tout en atteignant des niveaux élevés de confiance. De plus, ils ont généralement la plus faible déviation standard sur les différents essais. En revanche, les modèles basés sur CNN semblent être plus compatibles avec la transformation FFT et obtiennent des classements élevés dans cette configuration.

3.3.4. Conclusion

Dans ces travaux, nous avons démontré l'importance du formatage des données dans le cadre de la détection des phases de marche à partir de capteurs inertiels. Le projet a exploré différentes stratégies de formatage des données brutes issues des capteurs IMU pour améliorer la performance des modèles d'apprentissage profond dans l'analyse des séries temporelles multivariées. Nous avons proposé trois configurations principales de formatage : la matrice bidimensionnelle (accélération et vitesse angulaire sur trois axes), les entrées multiples pour chaque composante axiale, et un format à deux canaux séparant les données d'accélération et de gyroscope. Ces approches de formatage ont permis d'exploiter les corrélations spatiales et temporelles entre les signaux, améliorant ainsi la détection des phases de marche.

La contribution majeure de ce travail réside dans l'intégration de ces formats de données dans une architecture ConvLSTM, qui combine les réseaux de neurones convolutifs et les réseaux à mémoire longue courte (LSTM). Cette combinaison permet de capturer les motifs spatiaux des données de capteurs tout en conservant la dynamique temporelle. L'application des couches de déconvolution a également permis d'améliorer la reconstruction des motifs spatiaux après transformation par les couches ConvLSTM, offrant une meilleure précision dans la détection des phases critiques de la marche, telles que le contact du pied avec le sol et son retrait.

Les résultats montrent que la configuration des données brutes, sans transformation préalable, a permis d'obtenir des performances supérieures en termes de précision, de rappel et de F1-score. L'étude a également mis en évidence l'impact des transformations FFT et DCT sur les performances des modèles, avec des marges d'erreur plus élevées et une reproductibilité réduite lors de l'application de ces transformations. En définitive, nos résultats soulignent que le formatage des données brutes est non seulement essentiel pour la per-

formance des modèles, mais aussi pour garantir la robustesse et la fiabilité des prédictions dans des contextes variés, notamment pour les applications en rééducation.

Intérêts techniques de la détection des phases de marche : D'un point de vue technique, la détection des phases de marche offre plusieurs avantages cruciaux. Premièrement, elle permet une mesure précise de variables spatio-temporelles essentielles, telles que la durée de chaque phase de contact du pied avec le sol, la longueur des foulées, et la cadence. Ces informations peuvent être utilisées pour identifier les asymétries ou irrégularités dans la marche, ce qui est particulièrement utile pour évaluer les troubles moteurs et suivre l'évolution des patients au fil du temps. En analysant les phases de la marche, il est également possible de détecter des signes précoces de chute ou d'autres complications liées à des troubles de l'équilibre. Deuxièmement, le formatage adapté des données brutes permet d'optimiser l'entraînement des modèles d'apprentissage profond, en facilitant l'extraction des caractéristiques pertinentes des signaux multi-dimensionnels. Enfin, ces techniques de détection des phases permettent de personnaliser les programmes de rééducation en fonction du comportement individuel du patient, en fournissant des recommandations adaptées à chaque étape du processus de récupération. Ainsi, la précision et l'adaptabilité offertes par ces outils font d'eux des éléments-clés pour améliorer l'efficacité et la sécurité des interventions de rééducation.

Applications réelles dans le domaine de la rééducation : Le modèle développé dans le cadre de cette étude présente un potentiel significatif pour des applications cliniques en rééducation. En particulier, il peut être utilisé pour surveiller les progrès des patients lors de programmes de rééducation fonctionnelle après des blessures ou des chirurgies du bas du corps, telles que des fractures ou des prothèses de hanche. En offrant une analyse détaillée des phases de marche, cet outil permet aux cliniciens de suivre en temps réel les performances des patients et de détecter d'éventuels déséquilibres ou anomalies dans leur démarche. De plus, il peut être utilisé pour ajuster les plans de traitement en fonction des besoins spécifiques des patients, en améliorant la personnalisation des soins. L'intégration de cet outil dans les centres de rééducation pourrait également faciliter la mise en place de programmes de rééducation à distance, permettant aux patients de poursuivre leur rééducation chez eux tout en restant sous la surveillance des professionnels de santé grâce à des dispositifs portables.

3.4. Chapitre 3 : Reconnaissance d'activité pour l'analyse de la marche et l'évaluation de sa qualité

3.4.1. Introduction

La modélisation des types de mouvements liés à la marche est de plus en plus étudiée, et différentes méthodes ont été appliquées pour détecter les séquences où le pied touche le sol et où il ne le touche pas [31, 40, 63, 76]. Ce travail est essentiel pour des tâches telles que la reconnaissance d'activité humaine et est largement utilisé dans le domaine médical dans le but d'évaluer l'évolution de la marche [46, 64]. La majorité des défis auxquels ce domaine de recherche est confronté sont liés aux données. Par exemple, le type de capteurs à utiliser est primordial pour l'étude ; certains travaux se basent sur des données d'image et d'autres sur des données de capteurs inertiels [37, 39]. Par conséquent, une phase de pré-traitement est nécessaire pour compléter, corriger, redimensionner et filtrer les données ou pour extraire des caractéristiques et des motifs de ces signaux pour un développement ultérieur [36, 38, 77]. De plus, le formatage des données est également l'un des principaux défis de cette étude ; la manière dont chaque horodatage est classifié peut différer d'une méthode à l'autre en mettant en avant la dépendance temporelle ou spatiale, en concaténant des mesures provenant de différents signaux ou toute autre représentation des données pouvant avoir un sens pour le classificateur [46, 78, 79].

L'évaluation de la qualité des mouvements humains (HMQA) a suscité un intérêt croissant dans la littérature scientifique [76, 31]. Le HMQA ne doit pas être confondu avec la HAR, qui est une autre technique d'analyse des mouvements axée sur la reconnaissance de l'activité effectuée [39, 40]. Traditionnellement, le HMQA a été réalisé à l'aide d'approches basées sur des règles ou des modèles [64, 65]. Les approches basées sur des règles utilisent des déclarations conditionnelles prédéfinies pour évaluer les propriétés des mouvements humains [79]. Les approches basées sur des modèles utilisent des algorithmes tels que le Dynamic Time Warping (DTW) et les modèles de Markov cachés (HMM), qui effectuent des correspondances de motifs à l'aide de mouvements enregistrés précédemment [36, 80]. Ces méthodes sont faciles à mettre en œuvre, mais le nombre d'exécutions d'exercices pouvant être utilisées pour créer des règles ou modèles est limité. Cela réduit l'évolutivité de ces approches et rend difficile le développement de solutions personnalisées. Pour surmonter ces limitations, un intérêt croissant a été porté à l'utilisation de l'apprentissage automatique pour le HMQA [46, 81].

3.4.2. Modélisation inteprétable pour l'analyse des marqueurs petinents pour l'évaluation de la qualité de la marche

La marche est une activité complexe influencée par de nombreux facteurs tels que l'équilibre, la force musculaire et la coordination. La qualité de la marche, en particulier chez les patients en rééducation, est cruciale pour évaluer l'efficacité des thérapies et ajuster les protocoles de rééducation. L'objectif de cette étude est de proposer une méthode d'analyse des données issues de semelles connectées et de métadonnées des patients pour détecter les schémas liés à la qualité de la marche. Cette analyse permet de prédire les scores de marche en fonction des différents paramètres de mouvement recueillis. Pour ce faire, nous avons développé un cadre méthodologique basé sur des séries temporelles et des modèles d'apprentissage automatique.

Les données ont été collectées à partir de 43 participants ayant effectué un exercice de marche de 6 minutes. Chaque participant portait des semelles connectées mesurant la pression plantaire et les mouvements, enregistrant des données d'accélération et de pression sur les axes X, Y, et Z. Cependant, après filtrage des données incomplètes ou bruyantes, le jeu de données final se compose de 26 patients. Chaque série temporelle a été segmentée en fenêtres glissantes de 400 points, soit environ 10 secondes, pour capturer les transitions entre différentes phases de la marche.

Les séries temporelles ont ensuite été transformées en matrices 3D pour rendre les signaux compatibles avec notre modèle d'apprentissage profond. Cette transformation permet d'extraire des caractéristiques spatiales en superposant les signaux, facilitant ainsi la détection des corrélations entre les signaux. Le tableau 3.5 présente le nombre de pas effectués par catégorie pendant les fenêtres de 10 secondes.

Nombre de pas	Score 1	Score 2	Score 3	Score 4	Score 5
Min	3.08	5.32	6.17	7.75	2.92
Max	4.60	7.71	6.47	8.43	9.73
Moyenne	4.03	6.24	6.32	7.91	7.33
Écart-type	0.83	0.75	0.21	0.36	2.21

TABLE 3.5 – Nombre de pas pendant des intervalles de 10 secondes par catégorie

Pour évaluer la qualité de la marche, les experts en rééducation ont attribué un score allant de 1 à 5, basé sur une annotation manuelle des enregistrements vidéo capturant l'exercice de marche de 6 minutes. Ce score est déterminé en tenant compte de la vitesse de marche, du nombre de pas, de la position du centre de pression, et du score de Berg, qui évalue l'équilibre. La variable cible multi-classe a été transformée en un vecteur "one-hot"

pour répondre aux exigences de notre modèle d'apprentissage profond.

Outre les données de pression temporelles, des métadonnées sur les patients ont été collectées, incluant le sexe, le type de support utilisé lors de la marche et la pathologie. Ces variables catégorielles ont été encodées sous forme de vecteurs "one-hot". Des caractéristiques supplémentaires ont été extraites des données des semelles, telles que les valeurs minimales, maximales, et moyennes des positions X et Y, ainsi que la vitesse de déplacement V_{x_i}, calculée comme suit :

$$V_{x_i} = \frac{X_i - X_{i+1}}{T_e} \tag{3.5}$$

où T_e représente l'intervalle d'acquisition. D'autres indices tels que l'indice de déplacement moyen et l'indice de vitesse ont également été calculés pour fournir une vue plus complète de la qualité de la marche.

Compte tenu de la distribution déséquilibrée des classes, comme le montre la Figure 3.2, nous avons utilisé des techniques d'augmentation des séries temporelles pour enrichir le jeu de données et améliorer la robustesse du modèle. L'augmentation des données a été réalisée en ajoutant du bruit blanc aux séries temporelles, en générant des signaux gaussiens avec une distribution de probabilité définie par la fonction :

$$p_G(s) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} \tag{3.6}$$

où μ et σ sont respectivement la moyenne et l'écart-type du signal s. Cette méthode garantit que les signaux augmentés conservent les caractéristiques statistiques du signal original tout en augmentant la diversité des données pour l'entraînement du modèle.

Nous avons élaboré une stratégie d'entraînement guidée par deux objectifs principaux : l'interprétabilité des résultats et l'extraction efficace d'informations à partir de séries temporelles. Pour atteindre ces objectifs, nous avons développé deux blocs d'architectures basées sur le mécanisme d'auto-attention, chacun étant conçu pour traiter des types de données spécifiques. Notre jeu de données se compose de deux catégories principales : des séries temporelles multivariées provenant des semelles connectées, comprenant les composantes axiales des signaux, et les métadonnées descriptives des patients, telles que l'âge, le poids, la taille, la pathologie, et d'autres variables cliniques.

L'analyse des données des patients vise à identifier les combinaisons de variables qui influencent de manière significative le processus de décision du modèle. Pour cela, nous

avons implémenté un mécanisme d'auto-attention, en particulier une variante basée sur des convolutions. Cette approche nous permet de découvrir des corrélations potentielles parmi les métadonnées des patients, aidant ainsi le modèle à évaluer la qualité de la marche. Le poids des différentes entrées est ensuite calculé, et la carte de poids W_p est appliquée à l'entrée X_p par multiplication, produisant un vecteur pondéré $X_{weighted} = X_p W_p$ qui reflète l'importance de chaque entrée selon leur signification respective.

L'architecture de ce sous-module, illustrée à la Figure 3.5, se compose de trois blocs de couches convolutives unidimensionnelles, générant des cartes de caractéristiques distinctes représentant les requêtes, les clés et les valeurs. L'attention est ensuite calculée à l'aide de la fonction Softmax :

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3.7}$$

où $\sqrt{d_k}$ représente la dimension des vecteurs clé et requête. Le produit final est ensuite ajusté par un paramètre γ avant d'être ajouté à l'entrée initiale.

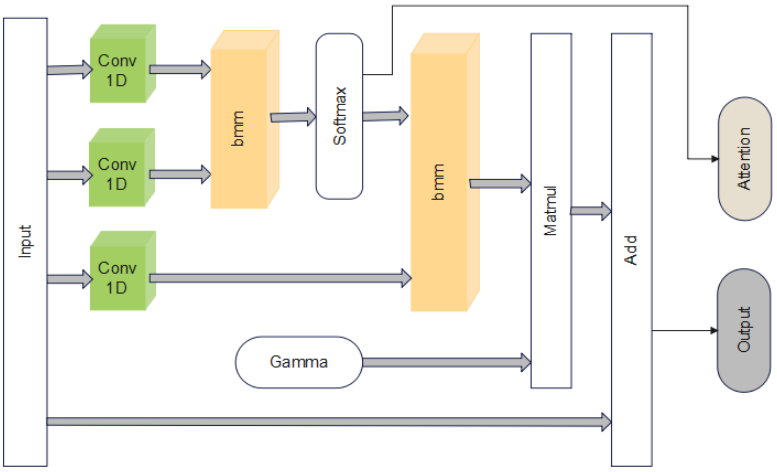

FIGURE 3.5 – Réseau de neurones pour l'entrée des métadonnées

L'architecture que nous avons développée pour traiter les séries temporelles multivariées joue un rôle essentiel dans la détection des interdépendances au sein des signaux. Plus précisément, nous utilisons les données agrégées de pression des semelles connectées, mesurées au centre de chaque pied. Notre approche consiste à segmenter les données en segments de taille égale, qui sont ensuite formatés en matrices 2D empilées. Cette transformation permet l'application d'un mécanisme d'auto-attention convolutif adapté, capable de capturer à la fois les corrélations intra-signal et inter-signal. Ainsi, cette méthode permet d'extraire des informations pertinentes provenant de différentes régions temporelles au

sein d'une série temporelle, tout en identifiant des motifs de corrélation entre les quatre composantes de chaque signal.

La Figure 3.6 illustre l'architecture de ce segment du modèle d'apprentissage profond. Bien que similaire au cadre des métadonnées, les couches convolutives 2D sont utilisées ici pour traiter les données multivariées. Cette approche permet de capturer des séquences temporelles importantes qui influencent directement les décisions du modèle.

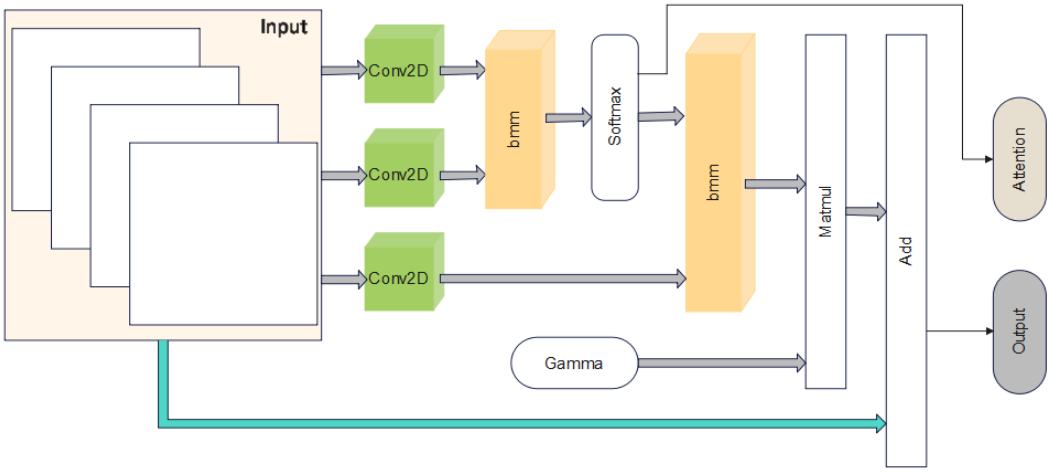

FIGURE 3.6 – Réseau de neurones pour l'entrée des séries temporelles

Les sorties des modules précédents sont aplaties et concaténées. De plus, nous dupliquons le vecteur des métadonnées du patient et l'ajoutons à la sortie concaténée pour améliorer la représentation d'une seule entrée. Cet enrichissement de l'entrée intègre désormais à la fois les informations sur le profil du patient et sa performance lors de l'exercice de marche de 6 minutes. Ensuite, une couche de projection est appliquée pour réduire la taille de la carte de caractéristiques, produisant ainsi une représentation des données plus compacte et concise. Enfin, le vecteur résultant est passé dans une couche de décision avec une taille de sortie de 5, correspondant aux cinq classes définies dans notre cas d'utilisation. L'architecture complète du modèle est illustrée à la Figure 3.7.

Notre cadre a été implémenté en utilisant la bibliothèque *PyTorch* et exécuté sur un ordinateur portable avec les spécifications suivantes : processeur Intel(R) Core(TM) i9 11e génération (2,60 GHz), 32 Go de RAM, et une carte graphique NVIDIA RTX A4000. Globalement, les expériences d'entraînement ont nécessité environ 3,5 jours pour s'achever. Afin d'assurer une évaluation impartiale, chaque expérience a été répétée 10 fois, et les scores ont été dérivés des moyennes calculées.

Pour optimiser les performances de notre architecture, nous avons ajusté les paramètres en fixant une liste de valeurs pour chaque hyperparamètre. Nous avons utilisé une approche

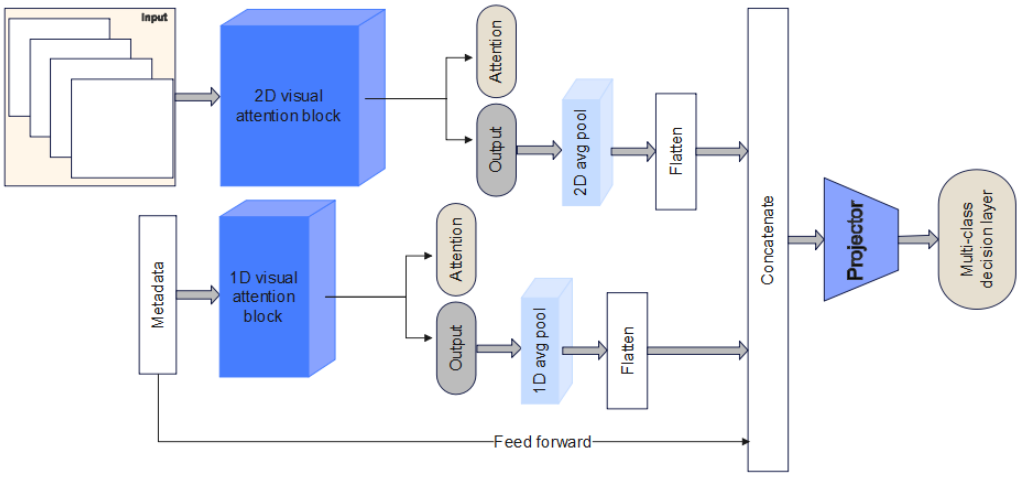

FIGURE 3.7 – Architecture complète du modèle avec les entrées des métadonnées et des séries temporelles

Hyper-paramètre	Plage de valeurs
Taille de batch	[1, 4, 8, 16, 32]
Taux d'apprentissage	[1e-3, 5e-3, 1e-4, 5e-4, 1e-5, 5e-5, 1e-6, 5e-6]
Nombre d'époques	[10, 15, 20, 25, 30, 40]
Taille du noyau CNN 2D	[1, 2, 4, 6, 8]
Taille du noyau CNN 1D	[1, 2, 3, 4]
Taille de la fenêtre de pool	[2, 3, 4]

TABLE 3.6 – Plage des valeurs des différents hyperparamètres

de recherche en grille (*grid-search*) pour tester une série de valeurs, comme détaillé dans le Tableau 3.6.

Pour des fins de comparaison, nous avons entraîné divers modèles avec une multitude de structures mettant en œuvre différentes approches, telles que des extracteurs de caractéristiques temporelles et spatiales. Conformément à l'objectif de l'étude, qui est d'explorer des architectures d'apprentissage profond explicables, nos efforts se sont concentrés sur les modèles basés sur l'attention. Les architectures testées sont listées comme suit :

— **Baseline DNN :** Ce premier modèle sert de référence pour la comparaison des résultats. Il a été entraîné sur différentes configurations, incluant des données de pression agrégées de chaque pied et, lors d'un second essai, des données agrégées du centre de gravité du corps humain. Le modèle a également été testé avec et sans chevauchement des fenêtres temporelles et avec ou sans augmentation de données.
— **RNN & LSTM :** Étant donné que ces modèles sont réputés pour leur capacité à traiter des séries temporelles, nous les avons appliqués exclusivement sur nos don-

nées temporelles. À ce stade, nous avons observé une amélioration des performances lors du traitement de séquences chevauchantes.

- **Auto-attention :** Dans les premières expériences, nous avons appliqué un mécanisme d'auto-attention directement sur les séries temporelles pour étudier son effet sur les données temporelles.
- **LSTM + auto-attention :** Nous avons combiné les atouts des LSTM avec les aspects interprétables de l'auto-attention. Le LSTM est appliqué sur les données temporelles et l'auto-attention sur les métadonnées.
- **Auto-attention à deux canaux :** Cette structure applique le même pipeline de traitement aux séries temporelles et aux métadonnées, en extrayant les poids d'importance des deux types de données via le mécanisme d'auto-attention.
- **Auto-attention temporelle :** Ce modèle crée des représentations clés, requêtes et valeurs avec des couches LSTM pour les séries temporelles, tandis que l'auto-attention standard est utilisée pour les métadonnées.
- **Auto-attention temporelle et spatiale :** Nous avons appliqué une auto-attention temporelle à chaque composante de signal individuellement, puis concaténé les sorties pondérées, sur lesquelles nous avons appliqué une auto-attention spatiale. Pour les métadonnées, nous avons utilisé une structure d'auto-attention standard.

Afin de déterminer les hyperparamètres les plus adaptés à notre modèle, nous avons partitionné notre ensemble de données en trois sous-ensembles : entraînement, validation et test. Nous avons alloué 60% des données pour l'entraînement, 20% pour la validation, et réservé les 20% restants pour l'évaluation du modèle. À l'aide de la bibliothèque `tsaug`, nous avons généré des séries temporelles similaires en ajoutant du bruit blanc, garantissant que les nouvelles séries temporelles maintiennent les mêmes caractéristiques de distribution. Pour chaque classe, nous avons recréé une ou plusieurs séquences similaires de signaux. La distribution des données allouées entre les partitions est illustrée dans la Figure 3.8.

3.4.3. Analyse des marqueurs corrélés à la qualité de la marche issues des matrices d'attention

Dans cette section, nous fournissons une vue d'ensemble complète du processus d'évaluation de notre modèle, ainsi que des évaluations des architectures concurrentes. Nous effectuons également une analyse approfondie des résultats, en mettant l'accent sur les interprétations médicales pour obtenir des informations sur la performance et l'efficacité du modèle.

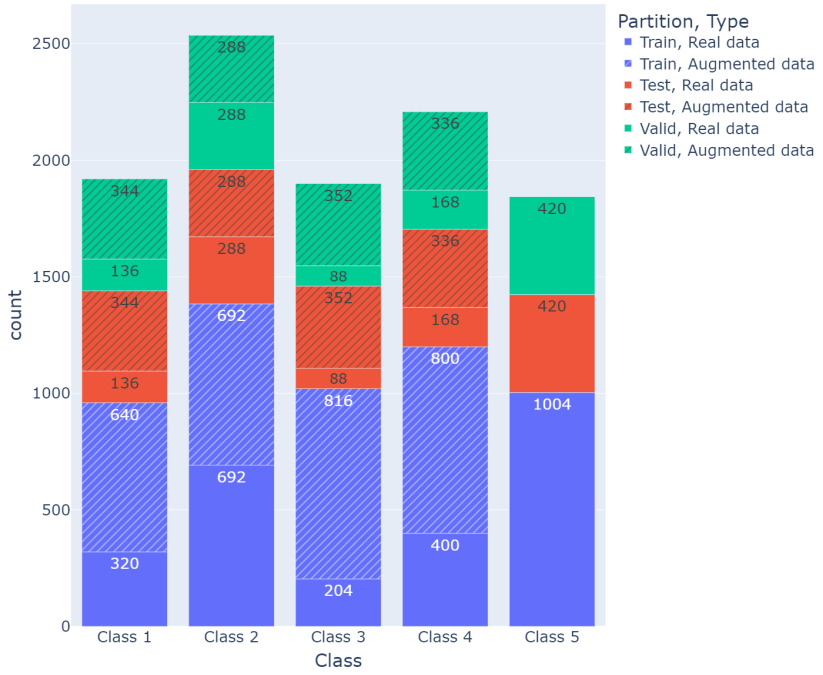

FIGURE 3.8 – Taille des partitions (entraînement, test et validation) par classe en termes d'observations (entrée matricielle 3D)

Dans le Tableau 3.7, il est évident que le modèle de référence obtient un score de précision notable de 92%. À travers ces premières expériences centrées sur le modèle de base et ses variantes, nous avons validé deux hypothèses. Tout d'abord, nous avons confirmé la nécessité d'utiliser des fenêtres chevauchantes et des techniques d'augmentation pour améliorer les performances du modèle. Étant donné les données limitées et les ratios de classes déséquilibrés, le modèle de base a rencontré des difficultés à généraliser à travers les différentes catégories. L'augmentation des données d'entraînement a permis de réduire considérablement les problèmes de surapprentissage (*overfitting*). De plus, la comparaison de deux types d'agrégation des données—spécifiquement, la position du centre d'équilibre du corps humain versus la position du centre de pression de chaque pied—montre que cette dernière fournissait des informations plus riches au modèle, ce qui a conduit à une meilleure détection.

Cependant, des défis sont apparus lors des expérimentations avec le mécanisme d'auto-attention standard. L'application de ce mécanisme, que ce soit sur les séries temporelles (ST) ou les métadonnées, a entraîné des problèmes de convergence pendant le processus d'optimisation, souvent bloquant le modèle dans des minima locaux. L'incorporation de LSTM avec le mécanisme d'attention pour créer un mécanisme d'auto-attention temporelle n'a pas permis de résoudre ce problème. En revanche, l'utilisation de LSTM pour les

séries temporelles et de l'auto-attention pour les métadonnées a facilité un entraînement plus fluide, avec un score de précision de 86%. Cependant, les matrices de poids extraites pour les métadonnées se sont avérées très bruitées et difficiles à interpréter. De plus, extraire des informations supplémentaires à partir des données temporelles elles-mêmes est resté un défi.

L'analyse des résultats montre que le modèle de base avec l'agrégation des données au niveau des pieds et l'utilisation de données chevauchantes a permis d'atteindre la meilleure performance avec une précision de 92%. Cela confirme l'importance de ces étapes de prétraitement dans la détection des motifs pertinents pour évaluer la qualité de la marche.

Cependant, les mécanismes d'attention appliqués sans ajustements appropriés se sont avérés inefficaces, avec des scores de précision aussi bas que 25% pour certaines configurations. Cela indique que, bien que l'attention soit prometteuse, sa mise en œuvre dans les modèles HAR nécessite davantage de réglages pour garantir une convergence stable et des matrices de poids interprétables.

Enfin, l'application du mécanisme LSTM avec attention sur les métadonnées s'est révélée relativement efficace, atteignant une précision de 86%, ce qui montre que l'attention peut effectivement extraire des informations utiles des données descriptives des patients. Cependant, l'extraction de motifs plus significatifs dans les données temporelles nécessite des méthodes plus robustes pour éviter le surapprentissage.

Notre modèle a atteint la meilleure précision de 94,9% en utilisant l'ensemble des valeurs d'hyperparamètres détaillées dans le Tableau 3.8. En termes de précision, le modèle a correctement identifié les classes 1 à 3 et 5, bien que 120 échantillons de la classe 4 aient été classés à tort dans la classe 5, comme illustré dans la matrice de confusion de la Figure 3.9. Ces résultats démontrent que notre modèle a surpassé les réseaux neuronaux récurrents, les LSTM et le modèle de base DNN en termes de performance de détection. Cependant, cette amélioration s'est faite au détriment de l'interprétabilité. Néanmoins, les matrices de poids obtenues sont plus susceptibles de fournir des interprétations claires quant à l'importance des entrées pour la détection.

Les erreurs commises par le modèle concernent principalement une catégorie spécifique et un seul patient, "P24". Notamment, toutes les erreurs correspondent à ce patient, dont les valeurs des métadonnées ressemblent étroitement à celles d'un patient de classe 5. Cela suggère que notre modèle accorde une grande importance aux métadonnées. D'un point de vue médical, nous pouvons réinterpréter la classification en cinq catégories comme un problème de classification en trois catégories, où la classe 1 représente une marche difficile, les classes 2 et 3 indiquent une marche de qualité moyenne, et les classes 4 et 5 désignent

Configuration	Modèle	Type de données	Augmentation	Précision
Baseline	DNN	Données du centre du corps	—	65%
			Données chevauchantes	71%
		Données des pieds	—	81%
			Données chevauchantes	92%
			Bruit blanc	90%
	RNN	Données des pieds	Données chevauchantes	88%
	LSTM	Données des pieds	Données chevauchantes	72%
Mécanisme d'attention	Auto-attention standard (SA)	Données du centre du corps	Chevauchement + bruit	32%
		Données des pieds	Chevauchement + bruit	29%
Attention sur métadonnées	LSTM (ST) + SA (métadonnées)	Données des pieds	Chevauchement + bruit	86%
Attention à deux canaux	SA (ST) + SA (métadonnées)	Données des pieds	Chevauchement + bruit	25%
Attention temporelle	SA temporelle (ST) + SA (métadonnées)	Données des pieds	Chevauchement + bruit	32%
Attention temporelle + spatiale	SA temporelle + spatiale (ST)	Données des pieds	Chevauchement + bruit	39%

TABLE 3.7 – Scores des différentes architectures concurrentes

Hyperparamètre	Taille du batch	Taux d'apprentissage	Nombre d'époques	Taille du noyau 2D	Taille du noyau 1D	Taille du pool
Valeur optimisée	1	1e-5	20	1	1	4

TABLE 3.8 – Valeurs des hyperparamètres optimisés du modèle

une marche fluide. Dans cette configuration, le modèle atteindrait des scores parfaits pour toutes les classes.

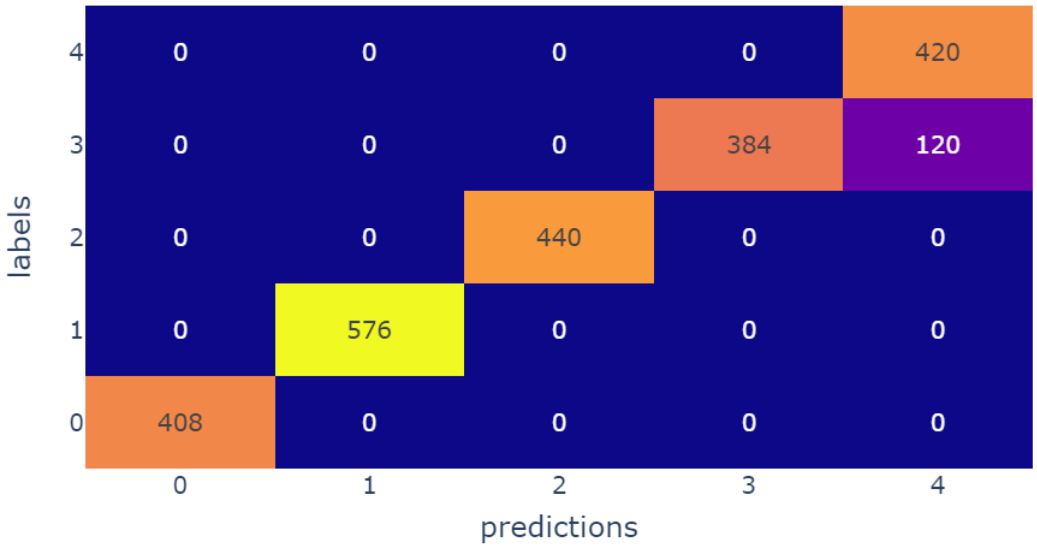

FIGURE 3.9 – Matrice de confusion de la classification

L'analyse des erreurs du modèle montre une concentration des erreurs sur la classe 4, souvent confondue avec la classe 5. Cette confusion peut être attribuée à la similitude des caractéristiques entre ces deux classes, notamment pour certains patients spécifiques. Le cas du patient "P24" est particulièrement révélateur, ce dernier ayant été systématiquement mal classé en raison de métadonnées proches de celles d'un patient de classe 5. Cela souligne l'importance des métadonnées dans les prédictions du modèle.

En réinterprétant les résultats avec une classification en trois catégories (marche difficile, marche moyenne et marche fluide), les performances du modèle se rapprochent de la perfection. Cette approche alternative permet de simplifier l'évaluation des capacités du modèle, en mettant en lumière l'excellente performance sur les patients présentant des troubles de la marche plus prononcés.

L'examen des matrices de poids d'attention dérivées des métadonnées des patients pour

chaque observation révèle que le modèle a appris à attribuer des degrés d'importance variables aux différents aspects du profil des patients. De manière intéressante, pour certains patients, la matrice de poids d'attention reste cohérente entre les différentes observations, ce qui suggère que le modèle a développé une représentation spécifique au patient basée sur les métadonnées. Cette cohérence souligne la capacité du modèle à reconnaître les patients en fonction de leurs caractéristiques uniques, en particulier celles capturées dans les métadonnées.

Ce phénomène éclaire les erreurs observées lors de la phase d'évaluation, où toutes les erreurs de classification sont attribuées à un seul patient. Il est plausible que le modèle, en s'appuyant fortement sur les métadonnées, ait mal évalué la classification de cet individu en particulier, l'assignant potentiellement à une classe supérieure (par exemple, classe 5) alors qu'il aurait dû être catégorisé dans une classe inférieure (par exemple, classe 4).

L'analyse des matrices de poids d'attention révèle également des informations précieuses sur les caractéristiques des métadonnées qui influencent de manière significative le processus de prise de décision du modèle. Notamment, des variables telles que la "Longueur du centre de pression" montrent de fortes corrélations avec d'autres variables des métadonnées. Cela suggère que cette caractéristique particulière joue un rôle important dans le processus de classification. D'un point de vue clinique, la longueur de l'excursion du centre de pression est un indicateur clé de la qualité de la marche et de l'évolution de la rééducation. Une excursion plus longue du centre de pression indique généralement un déroulement du pied plus homogène pendant la phase d'appui du cycle de marche, caractéristique d'une marche plus fluide et plus stable. À l'inverse, les individus présentant des excursions du centre de pression plus courtes peuvent afficher une répartition irrégulière ou inégale du poids, potentiellement indicatrice d'anomalies ou de pathologies de la marche.

De même, la variable "Valeur moyenne du centre de pression selon l'axe y" montre également des corrélations, en particulier avec elle-même. Cela met en évidence l'importance de la position du centre de pression dans l'analyse de la marche. Les variations de la position moyenne du centre de pression le long de l'axe y peuvent révéler des tendances à exercer plus de pression sur certaines régions du pied, comme les orteils ou le talon. Sur le plan clinique, ce paramètre fournit des informations précieuses sur les schémas de répartition du poids pendant la marche, aidant à l'évaluation des anomalies de la marche ou des écarts par rapport à la biomécanique normale. Dans notre cas, cela se réfère à une ou plusieurs des observations suivantes :

— des petits pas ou un mouvement du pied parétique ramené au niveau de l'autre

pied, sans sépassement ;

— des déficits de flexion de la hanche et du genou, entraînant le saut de certaines phases de marche et obligeant à poser le pied à plat directement, sans déroulé complet ;

— l'utilisation d'une aide technique permettant d'éviter d'appliquer toute la pression sur le côté parétique.

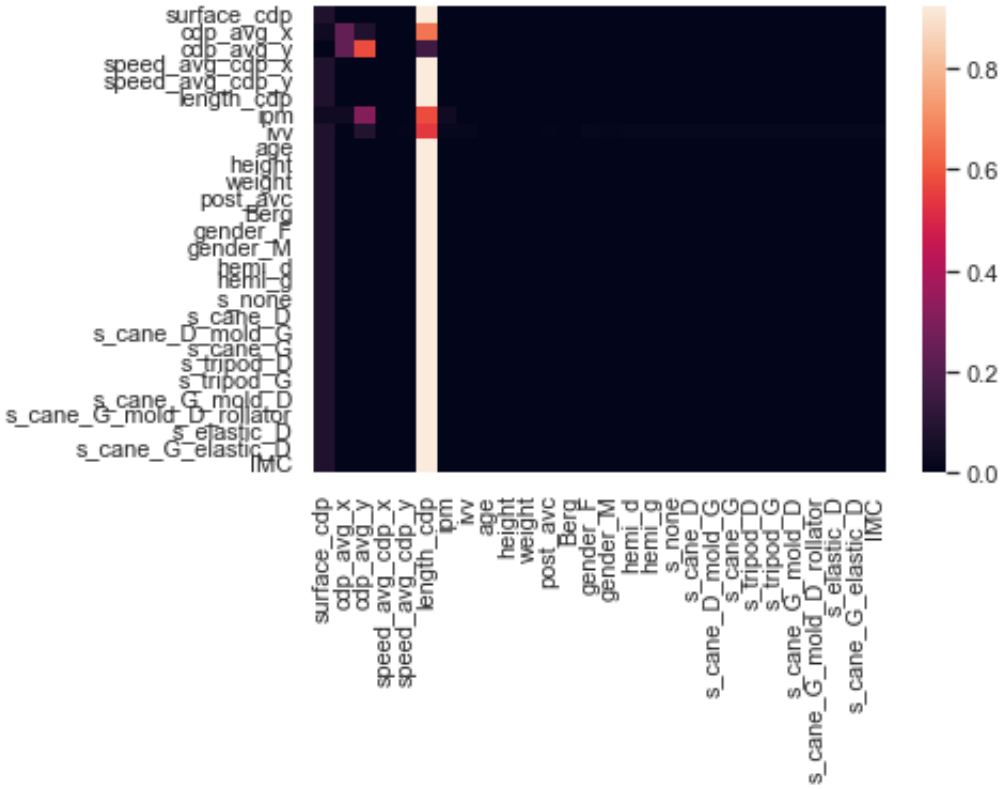

FIGURE 3.10 – Moyenne des matrices de poids d'attention pour les entrées des métadonnées

À travers la projection des poids sur les séries temporelles (voir Figure 3.11), nous observons un schéma périodique dans la détection, reflétant la nature cyclique inhérente aux événements répétitifs tels que la marche.

En agrégeant la moyenne de ces matrices de poids le long de l'axe vertical, nous obtenons des informations sur les zones les plus cruciales du signal. Cette visualisation agrégée, illustrée dans la Figure 3.11, met en évidence des régions pondérées caractérisées par des pics périodiques de valeurs. En exploitant des couches d'attention visuelle, les couches convolutionnelles extraient des motifs spatiaux communs aux quatre composantes du signal de pression. Par conséquent, les intervalles mis en évidence encapsulent des informations sur les changements abrupts de valeurs ou des plages où le signal enregistre une activité si-

gnificative. Notamment, les événements constituant le cycle de marche, tels que le contact du talon ou le retrait des orteils, se manifestent généralement par des activités de pression brèves et intenses des pieds. Dans le domaine médical, la durée, l'intensité et la fréquence de ces phases sont des facteurs clés pour évaluer la qualité de la marche d'un individu. Il est remarquable que notre modèle montre des largeurs d'intervalle variables en fonction des différentes activités, suggérant sa capacité à discerner les séquences cruciales du cycle de marche sans connaissance préalable de ses phases. Cette adaptabilité souligne le potentiel du modèle à identifier des motifs de marche importants associés à des types spécifiques d'activités.

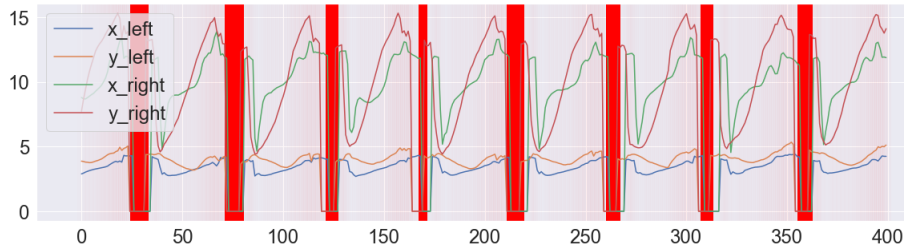

FIGURE 3.11 – Projection des poids sur les séries temporelles

La durée des séquences pondérées complètes varie entre les patients, comme illustré dans la Figure 3.12, qui montre la distribution de ces intervalles. Pour obtenir un signal binaire couvrant toute la série temporelle, nous avons pris en compte uniquement les poids dépassant la moyenne des valeurs correspondant à chaque échantillon d'entrée. Dans l'ensemble de test, la majorité des séquences temporelles pondérées comportent entre un et huit points temporels, avec de nombreux petits intervalles étroitement groupés et traités comme un seul intervalle.

De manière remarquable, chaque patient présente une série distincte d'intervalles temporels, un schéma fortement corrélé à leur classe respective de qualité de marche. Notamment, la Figure 3.12 indique une diminution de la durée des intervalles pondérés avec la baisse de la classe de qualité de marche. Cette tendance s'explique par le fait que les personnes souffrant de troubles de la marche mettent plus de temps à accomplir diverses actions au sein du cycle de marche. Par conséquent, les actions associées à une pression significative sur les semelles, telles que le contact du talon ou le retrait des orteils, ont tendance à durer plus longtemps. Ainsi, la représentation graphique révèle une tendance à la baisse de la longueur des séquences pondérées, en cohérence avec le score de qualité de la marche de chaque patient.

De plus, la vitesse de marche est fortement influencée par les problèmes de motricité, entraînant des vitesses plus faibles parmi les populations ayant des scores plus bas. No-

TABLE 3.9 – Statistiques sur la durée des séquences pondérées

Statistiques	Nombre	Moyenne	Min	25%	50%	75%	Max
Valeur	18976	7.42	1	2	5	11	40

tamment, les vitesses de marche moyennes par population sont alignées avec les scores de qualité de marche : score 1 (0.21 m/s), score 2 (0.48 m/s), score 3 (0.75 m/s), score 4 (0.91 m/s) et score 5 (1.18 m/s). Ces valeurs soulignent la corrélation entre la durée des séquences pondérées et la classe cible, fournissant un contexte supplémentaire aux tendances observées.

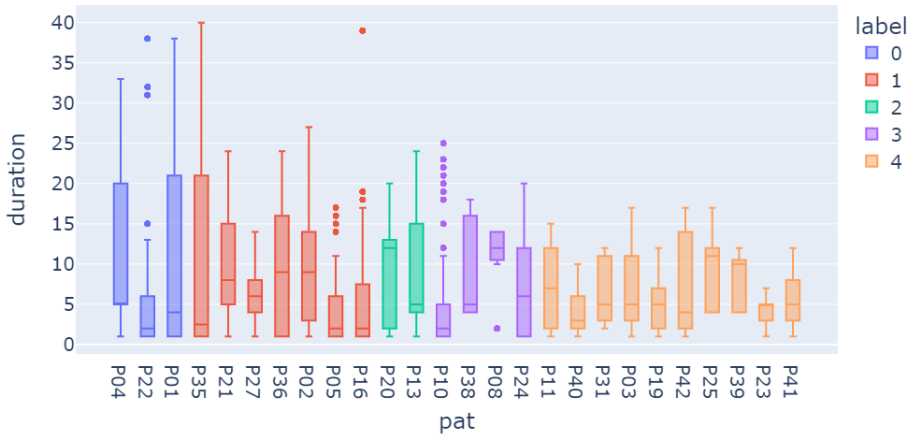

FIGURE 3.12 – Distribution de la durée des séquences pondérées selon les étiquettes

3.4.4. Conclusion

Variables d'entrée et hypothèses : Les données de pression provenant des semelles connectées, qui constituent les principales entrées de notre modèle, encapsulent des informations périodiques que notre modèle identifie habilement à travers différentes séquences. En particulier, le modèle montre une grande capacité à détecter des segments de signaux caractérisés par des pics prononcés ou des changements brusques de valeurs, correspondant à des séquences de marche reconnues. Il est remarquable que cette capacité de détection des signaux ait été acquise de manière autonome par le modèle, sans connaissance préalable des phases spécifiques de la marche [17].

Par ailleurs, nous intégrons des métadonnées issues du profil des patients, telles que l'âge, le sexe, le type de pathologie, l'utilisation de supports à la marche, les métriques de pression et le score de Berg. Bien que certaines de ces variables puissent présenter des corrélations évidentes avec le score de marche, leur importance dans les matrices de poids

varie. Par exemple, l'âge, souvent perçu comme un facteur influençant les schémas de marche, peut ne pas apparaître de manière significative en raison des réponses hétérogènes des individus aux traitements de rééducation, indépendamment de leur âge. Ainsi, les troubles de la marche liés à l'âge, notamment ceux associés à un risque accru de chutes, pourraient avoir un impact plus subtil sur les caractéristiques de la marche, expliquant ainsi leur faible présence dans les matrices de poids [82].

De même, la variable représentant le délai post-AVC (variant de 1 à 160 mois) peut ne pas apparaître systématiquement dans les matrices de poids. La période initiale de six mois après l'AVC, considérée comme une phase de récupération, est généralement dédiée à la correction des irrégularités de marche. Au-delà de cette période, ces irrégularités sont souvent perçues comme des séquelles, les patients adaptant leur démarche pour optimiser leurs capacités. Par conséquent, un délai post-AVC supérieur à six mois peut ne pas influer directement sur les caractéristiques de la marche, ce qui explique son importance réduite dans les matrices de poids [83].

Interprétabilité du modèle d'apprentissage profond : En comparant les modèles de type "boîte noire" (black box) aux modèles dits "boîte blanche" (white box), l'accent est mis sur la transparence et l'interprétabilité des modèles. Les modèles "boîte noire" fonctionnent de manière opaque, ne permettant pas aux utilisateurs de comprendre les décisions et les prédictions du modèle. À l'inverse, les modèles "boîte blanche" offrent généralement une transparence et une interprétabilité accrues. Bien que la performance soit un critère important pour comparer ces deux types de modèles, les modèles interprétables privilégient souvent la compréhension des processus décisionnels au détriment d'une légère baisse de performance.

Dans cette étude, l'obtention de scores de classification élevés s'est avérée être un défi considérable, en particulier avec l'utilisation du mécanisme d'attention standard. Ainsi, la principale contribution de ce travail réside dans la réalisation de performances élevées avec des modèles interprétables, même en présence de données limitées. De plus, la capacité d'extraire des caractéristiques spécifiques et d'identifier des phases critiques du cycle de marche qui influencent les décisions du modèle renforce la pertinence de cette contribution [84].

Utilisation des méthodes d'IA dans l'assistance médicale : L'intelligence artificielle (IA) est de plus en plus utilisée dans divers domaines de la santé, allant de l'analyse des images médicales pour la détection des maladies à l'aide au diagnostic, en passant par l'interprétation des signes vitaux et des résultats de laboratoire. Cependant, notre étude se concentre plus spécifiquement sur le développement d'une infrastructure permettant

de soutenir les patients à domicile et d'assurer un suivi continu. À notre connaissance, cette recherche est la première à évaluer la qualité de la marche à l'aide de modèles d'apprentissage profond, en particulier à travers une approche interprétable facilitant l'analyse des sorties du modèle [85].

Ce travail démontre l'efficacité des modèles d'attention visuelle pour classer la qualité de la marche. Pour étendre l'applicabilité de cette solution à des institutions plus grandes ou à des environnements similaires, une surveillance continue et un raffinement supplémentaire sont nécessaires. Un ensemble de données plus large est également crucial pour garantir la généralisabilité du modèle à une population diversifiée de patients. En outre, la surveillance de la qualité des données et des métriques de précision du modèle est nécessaire pour éviter toute dérive du modèle au fil du temps [86].

Notre modèle est également capable d'identifier des phases clés du cycle de marche qui influencent de manière significative le processus de décision. Il peut également être utilisé pour explorer des indicateurs ou des facteurs dans le profil d'un patient pouvant affecter ses capacités de marche. Sur la base de ces contributions, nous prévoyons d'élargir cette étude à une cohorte plus large afin d'exploiter les forces de ce modèle pour la détection des phases de marche associées aux irrégularités de marche [87].

3.5. Chapitre 4 : Modélisation interprétable pour la reconnaissance des activités et la sélection des signaux d'entrée

3.5.1. Introduction

Le besoin d'outils automatisés pour le suivi et l'évaluation des performances des patients en rééducation est devenu essentiel, notamment dans les situations où un suivi continu et en temps réel est nécessaire [56]. Cela est particulièrement vrai dans les cas de patients victimes d'accidents vasculaires cérébraux (AVC) ou d'autres affections entraînant des troubles de la marche. Ces patients nécessitent des programmes de rééducation personnalisés et adaptés à leurs besoins spécifiques. Dans ce cadre, la HAR permet aux thérapeutes de suivre les progrès des patients de manière objective, en se basant sur des données quantitatives plutôt que sur des observations subjectives [88].

Les approches basées sur l'intelligence artificielle, notamment les réseaux neuronaux profonds, sont de plus en plus utilisées pour améliorer l'analyse des signaux inertiels en vue de la reconnaissance d'activités. En particulier, les modèles hybrides combinant des architectures CNN-BiLSTM (réseaux de neurones convolutifs et réseaux neuronaux à mémoire à long court terme bidirectionnels) se sont révélés très efficaces pour traiter à la fois les données temporelles et fréquentielles [63, 77]. Ces modèles permettent d'extraire des caractéristiques complexes des signaux et d'identifier des schémas spécifiques correspondant à différentes activités physiques [78, 46]. En outre, l'utilisation de mécanismes d'attention intégrés dans ces architectures permet d'améliorer l'interprétabilité des modèles, offrant ainsi aux professionnels de santé des informations compréhensibles sur les facteurs influençant la performance des patients [89, 90].

Dans cette étude, nous nous intéressons à la reconnaissance des phases de marche chez les patients en rééducation, en utilisant des données issues de capteurs inertiels et en les transformant à la fois dans les domaines temporel et fréquentiel [54]. En exploitant les techniques d'IA interprétable, nous cherchons à offrir des outils plus transparents et efficaces aux thérapeutes, facilitant ainsi la prise de décision et l'adaptation des programmes de rééducation. De plus, cette approche permet d'identifier les aspects les plus pertinents des données pour chaque activité, renforçant ainsi la personnalisation des traitements [91].

3.5.2. Mécanismes d'attention à double niveau pour l'extration des capteurs les plus pertinents

L'hétérogénéité révèle un déséquilibre des données, ce qui peut entraîner un risque d'overfitting lors de l'entraînement des modèles. Il devient donc nécessaire de mettre en place des techniques d'augmentation des données pour atténuer cet effet. Dans ce cadre, nous avons choisi d'utiliser la méthode *RandomOverSampler*, qui a pour objectif de rééquilibrer la distribution des classes dans un jeu de données déséquilibré. Cette technique fonctionne en augmentant le nombre d'instances dans les classes minoritaires (dans ce contexte, les activités sous-représentées) par réplication aléatoire d'échantillons existants. Le processus consiste à sélectionner de manière aléatoire des échantillons dans les classes minoritaires et à les dupliquer jusqu'à atteindre un niveau de balance souhaité. Cela garantit que le modèle est suffisamment exposé à ces activités pendant l'entraînement, améliorant ainsi sa capacité à généraliser de manière plus robuste et précise. À la suite de cette opération, chaque activité représente 14,29% du jeu de données, ce qui équivaut à environ 17 195 éléments par classe.

Les signaux provenant des capteurs inertiels produisent des séries temporelles continues, et les données d'activité suivent ainsi ce format. La première étape de la reconnaissance d'activités humaines consiste à segmenter les données de ces capteurs. Généralement, une fenêtre glissante est utilisée pour diviser les données en fenêtres de taille fixe. Pour optimiser la performance du modèle, un processus expérimental est mené pour déterminer la taille de fenêtre et le pourcentage de chevauchement les plus appropriés. Le nombre de caractéristiques prises en compte est de 12, incluant les données d'accélération et de vitesse angulaire, pour chacun des trois axes, recueillies au niveau des pieds gauche et droit. Ensuite, afin de préparer les données d'entrée X et de sortie y, différentes transformations sont appliquées aux données de chaque fenêtre, qui sont ensuite concaténées en un seul vecteur ajouté à X.

Dans notre quête de performance optimale, nous avons exploré plusieurs techniques de transformation des données en utilisant des métriques statistiques telles que la moyenne, la variance, le maximum, le minimum et l'écart-type. Les méthodes les plus avancées soulignent l'importance de transformer les données temporelles en domaine fréquentiel pour améliorer les performances.

Afin d'évaluer de manière complète plusieurs approches, nous avons tout d'abord utilisé des algorithmes classiques de *Machine Learning* (ML). Pour la classification, nous avons évalué plusieurs modèles, notamment les *Support Vector Machines* (SVM), le *Random Forest Classifier*, la régression logistique, ainsi que les classificateurs XGBOOST. Les don-

nées collectées à partir des capteurs inertiels étant des séries temporelles, elles possèdent une structure unidimensionnelle où les variables à proximité immédiate sont fortement corrélées au fil du temps. Il est donc crucial d'extraire ces caractéristiques locales.

Les réseaux neuronaux convolutifs (CNN) ont la capacité d'extraire ces caractéristiques locales grâce à leurs champs réceptifs locaux. Par ailleurs, les architectures basées sur des réseaux neuronaux récurrents (RNN) sont plus adaptées pour extraire les dépendances temporelles au sein des séries. Cependant, ces architectures sont souvent perçues comme des "boîtes noires" en raison de leur manque d'interprétabilité. Or, lorsqu'il s'agit de données médicales, il est essentiel de pouvoir interpréter les prédictions, d'où la nécessité d'étudier les corrélations entre les entrées et les sorties prédites.

Les limitations des modèles traditionnels d'apprentissage profond en reconnaissance d'activités humaines résident dans leur incapacité à capturer efficacement à la fois les caractéristiques locales et les dépendances à long terme inhérentes aux données de mouvement. Toutefois, notre modèle proposé surmonte ces limitations en utilisant une architecture hybride qui intègre des données à la fois du domaine temporel et du domaine fréquentiel. À travers notre exploration de l'état de l'art, nous avons découvert que l'utilisation de l'architecture multibranche CNN-BiLSTM améliore la précision [54]. Initialement, la même entrée est transmise à chacune des trois branches. Les trois premières branches reçoivent les données d'origine dans le domaine temporel, tandis que les trois autres branches reçoivent les données transformées du domaine temporel au domaine fréquentiel à l'aide de la FFT (Transformée de Fourier Rapide). Pour ces six branches parallèles, l'entrée passe par différentes tailles de filtre : 3, 7 et 11. L'utilisation de différentes tailles de filtre permet de capturer diverses dépendances locales temporelles de différentes longueurs. Chaque branche comprend une couche de mécanisme d'attention, suivie de deux couches convolutionnelles 1D (Conv1D) pour extraire les caractéristiques locales des données séquentielles. L'entrée est acceptée sous la forme (128, 12), où 128 correspond à la taille de la fenêtre et 12 aux composantes du signal, incluant l'accélération et la vitesse angulaire pour les deux pieds.

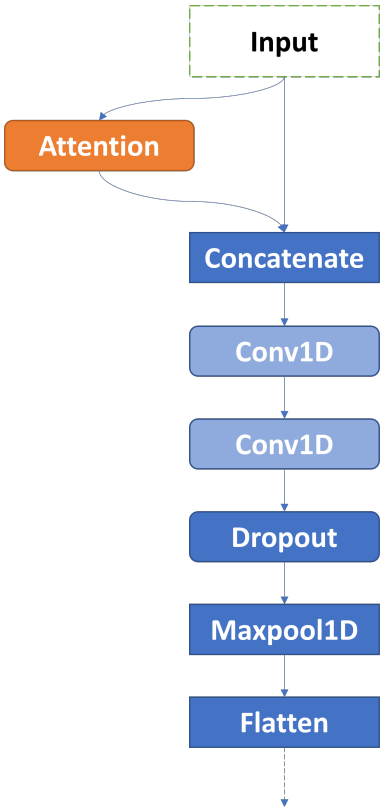

FIGURE 3.13 – Bloc CNN-attention par branche.

Chaque branche traite des entrées de longueur 128, divisées en quatre sous-séquences de 32. L'activation ReLU introduit des non-linéarités dans les couches Conv1D. Après la deuxième couche Conv1D, un taux de dropout de 0.5 est appliqué pour prévenir le sur-apprentissage (overfitting). La dimensionnalité est réduite à l'aide d'une couche de MaxPooling 1D avec une taille de fenêtre de deux. Cette séquence de couches, illustrée dans la Fig. 3.13, varie en taille de noyau pour capturer des motifs spatiaux de différentes longueurs. La Fig. 3.14 illustre l'architecture globale qui combine les sorties de toutes les six branches. Ces sorties sont concaténées, aplaties, puis transmises à une seconde couche d'attention, suivie de deux couches BiLSTM pour capturer les dépendances à long terme. La sortie finale du BiLSTM est transmise à une couche dense, normalisée à l'aide de la normalisation par lot (Batch Normalization), puis à une couche de classification avec une activation Softmax pour prédire les classes d'entrée. L'architecture est entraînée pour minimiser la perte d'entropie croisée catégorielle en utilisant l'optimiseur Adam. Les poids issus des couches d'attention, mis en évidence dans les figures, sont utilisés pour une analyse d'interprétabilité supplémentaire.

Ce modèle sera comparé à deux autres architectures dérivées, l'une où la normalisation par lot (Batch Normalization) est appliquée après chaque activation ReLU, ce qui contri-

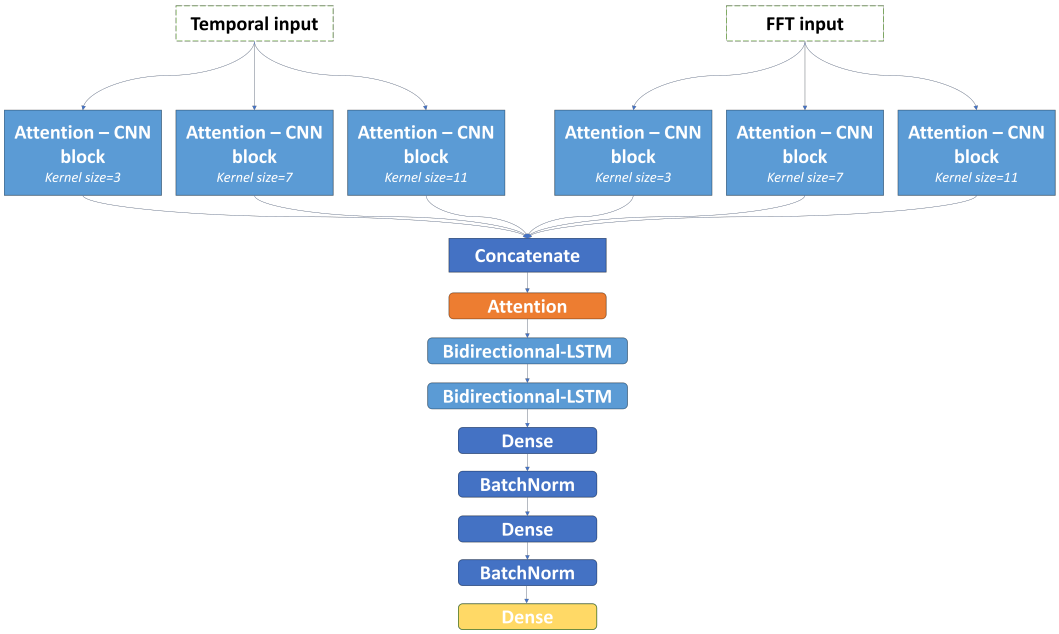

FIGURE 3.14 – L'architecture interprétable proposée pour la détection des activités de rééducation.

bue à stabiliser le processus d'entraînement et conduit à une convergence plus rapide. En effet, la normalisation au sein de chaque couche devrait aboutir à un processus d'entraînement plus stable, empêchant le sur-apprentissage (overfitting) et améliorant la généralisation aux données non vues. L'autre modèle utiliserait « TimeDistributed » appliqué à chaque couche CNN. En effet, les données de séries temporelles sont séquentielles, avec un ordre temporel important pour la compréhension des motifs et la réalisation de prédictions. L'utilisation de « TimeDistributed » permet au modèle d'apprendre des motifs et des caractéristiques significatives à travers la dimension temporelle, tout comme les CNN apprennent les caractéristiques spatiales dans les données d'image. Ces architectures manquent de l'aspect interprétabilité, mais nous souhaitons comparer les performances de nos modèles et vérifier si l'interprétabilité du modèle compromet ses performances.

L'efficacité de l'architecture de modèle proposée est évaluée en calculant différentes métriques de performance décrites comme suit :

Accuracy : Elle est définie comme la fraction d'échantillons prédits correctement par rapport au nombre total d'échantillons, où $TP = True\ Positives$, $FN = False\ Negatives$, $TN = True\ Negatives$ et $FP = False\ Positives$.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Précision (Precision) : La fraction d'échantillons positifs correctement identifiés parmi le nombre total d'échantillons prédits comme positifs.

$$Prcision = \frac{TP}{TP + FP}$$

Rappel (Recall) : La fraction d'échantillons positifs correctement reconnus parmi le nombre total d'échantillons réellement positifs.

$$Rappel = \frac{TP}{TP + FN}$$

F1-Score : C'est une estimation complète de la précision du modèle et elle peut être calculée comme la moyenne harmonique de la précision et du rappel.

$$F1 - Score = \frac{2 \times Prcision \times Rappel}{Prcision + Rappel}$$

Matrice de Confusion (Confusion Matrix - CM) : Il s'agit d'une matrice carrée qui fournit une vue d'ensemble des performances d'un modèle de classification. Les lignes de la CM représentent les instances des étiquettes de classe réelles, et les colonnes représentent les étiquettes de classe prédites. Les éléments diagonaux de cette matrice définissent le nombre de points pour lesquels l'étiquette prédite est égale à l'étiquette réelle.

Notre approche a été développée à l'aide de la bibliothèque Keras (framework TensorFlow) et exécutée sur un ordinateur portable DELL avec les caractéristiques suivantes : 11ème génération Intel(R) Core(TM) i9 (2,60Hz) et 32 Go de RAM.

3.5.3. Extraction et étude des des signaux d'entrée corrélés avec l'activité de rééducation

Lors de l'optimisation des paramètres, il est apparu qu'une taille de fenêtre glissante de 128 avec un chevauchement de 50% sur l'ensemble du jeu de données permettait d'obtenir les meilleurs résultats. Pour notre étude de cas, nous avons expérimenté les transformations Fast Fourier Transform (FFT) et Discrete Cosine Transform (DCT) sur les données, mais nos résultats étaient meilleurs en utilisant une combinaison de données dans le domaine temporel et les transformations FFT.

% Overlapping	40%	60%	85%	98%
Logistic Regression	38%	36%	37%	42%
Random Forest	85%	87%	94%	99%
XGBOOST	85%	87%	92%	98%

FIGURE 3.15 – Comparaison de l'exactitude des modèles d'apprentissage automatique.

Notre première évaluation s'est concentrée sur les modèles d'apprentissage automatique (ML) qui ont été entraînés sur des valeurs statistiques calculées pour remplacer chaque fenêtre de données brutes. Ces métriques statistiques incluent la moyenne, la variance, le maximum, le minimum et l'écart-type. Comme le calcul de ces valeurs sur les données transformées par FFT n'a pas permis d'atteindre des taux d'exactitude acceptables, nous avons utilisé uniquement les données brutes. La figure 3.15 met en évidence les modèles de ML les plus prometteurs, démontrant que les classificateurs XGBOOST et Random Forest sont les meilleurs modèles dans cette étude. De plus, plus le pourcentage de chevauchement est élevé, plus la performance de chaque modèle est bonne, en raison de la grande quantité de données générées. Cette figure montre également que les modèles de ML dépendent fortement de la quantité de données. Dans notre cas, nous avons réussi à atteindre une exactitude supérieure à 90% seulement après avoir adopté un pourcentage de chevauchement de 85% ou plus, ce qui est très coûteux en termes de calcul.

Ensuite, notre attention s'est portée sur l'évaluation des architectures basées sur l'apprentissage profond (DL). Plus précisément, nous avons entraîné diverses variantes du même modèle, en incorporant stratégiquement des couches supplémentaires telles que la normalisation par lot (Batch Normalization), la couche Time Distributed, et notamment la couche de mécanisme d'attention. Ces modèles ont été conçus pour prendre en compte à la fois les données temporelles et les données transformées par l'algorithme FFT. Dans la figure 3.16, il est discernable que l'inclusion des couches de normalisation par lot et Time Distributed n'a pas contribué à une amélioration de l'exactitude ; au contraire, l'exactitude a diminué dans ces cas. À l'inverse, l'architecture proposée a démontré une performance

notable avec une exactitude de 92%. De plus, l'introduction du mécanisme d'attention n'a pas affecté négativement les performances du modèle ; elle a plutôt fourni la capacité précieuse d'interpréter les résultats sans compromettre l'exactitude.

	Accuracy	F1-score	Recall	Precision
CNN-BiLSTM with Batch Normalization	90%	90%	90%	90%
CNN-BiLSTM with TimeDistributed layer	91%	91%	92%	91%
CNN-BiLSTM	92%	92%	92%	92%
CNN-BiLSTM with Attention	**92%**	**92%**	**92%**	**92%**

FIGURE 3.16 – Comparaison de l'exactitude des modèles d'apprentissage profond.

Comme illustré dans la figure 3.14, notre méthodologie intègre deux étapes de couches d'attention, chacune servant un objectif distinct. La première étape est dédiée à l'analyse des composantes axiales des signaux qui expliquent principalement chaque type d'activité. Ensuite, la deuxième étape est utilisée pour mettre en avant la branche de l'architecture qui exerce l'influence la plus significative sur la prédiction du modèle. Notamment, notre approche implique l'adoption de six branches, les trois premières étant dédiées aux données temporelles et les trois restantes aux données transformées par FFT. Au sein de chaque ensemble de trois branches, différentes tailles de noyaux sont utilisées pour capturer efficacement les motifs de différentes longueurs. Cette conception améliore la capacité du modèle à discerner et à interpréter les caractéristiques complexes présentes dans les données d'entrée.

Pour évaluer l'interprétabilité du modèle, nous présentons les projections d'entrée après l'attention dans la figure 3.17. Cette première figure examine spécifiquement les branches temporelles. Les colonnes de la figure correspondent à différentes tailles de noyaux (3, 7 et 11 respectivement), tandis que les lignes représentent les poids moyens par activité. L'axe des x de chaque graphique représente les signaux d'entrée, y compris l'accélération et la vitesse angulaire pour les pieds gauche et droit. Les visualisations révèlent une insistance constante sur l'accélération et la vitesse angulaire selon l'axe y pour diverses activités. Cela suggère que le modèle tend à accorder plus d'importance aux données du pied gauche par rapport au pied droit, bien que l'attention soit plus uniformément répartie pour le



FIGURE 3.17 – Cartes d'attention moyennes selon chaque type d'activité de rééducation pour les branches temporelles avec des tailles de noyaux de 3, 7 et 11 (de gauche à droite).

pied droit. Cette tendance est également observable dans les branches FFT, où le modèle montre une attention accrue à l'accélération le long des axes x et z et à la vitesse angulaire selon l'axe y dans le domaine fréquentiel, comme illustré à la figure 3.18.

3.5.4. Conclusion

L'une des principales limites de cette étude réside dans son caractère monocentrique, réalisée exclusivement au sein du centre de rééducation Jacques Calvé. Cette particularité pourrait restreindre la généralisation des résultats à d'autres contextes et populations [24]. De plus, bien que la cohorte comprenne une diversité de patients à différents stades de récupération post-AVC, elle ne comptait que 26 participants, ce qui peut affecter la robustesse des conclusions [27]. L'analyse s'est concentrée sur les signaux provenant des unités de mesure inertielle (IMU), sans intégrer d'autres données provenant de capteurs comme

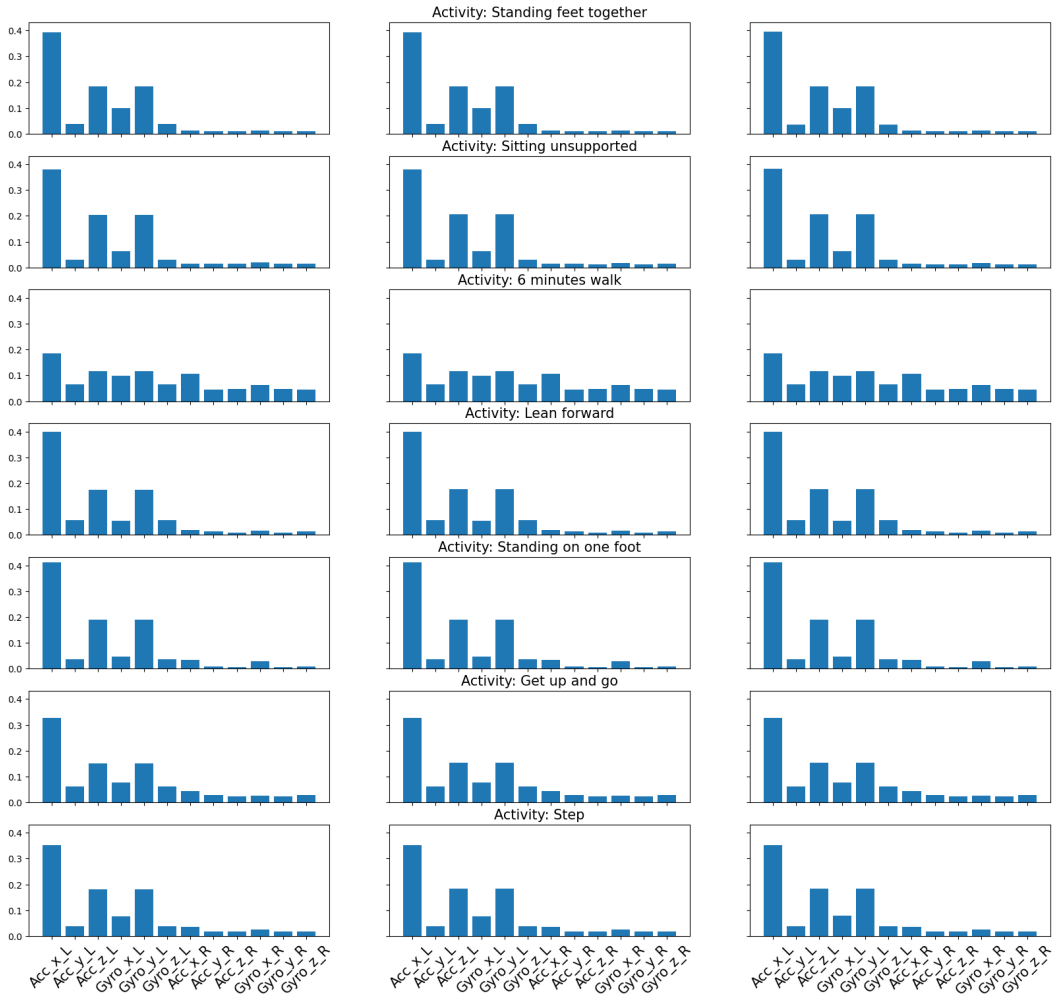

FIGURE 3.18 – Cartes d'attention moyennes selon chaque type d'activité de rééducation pour les branches temporelles avec des tailles de noyaux de 3, 7 et 11 (de gauche à droite).

la pression des semelles, l'électromyographie (EMG) ou la surveillance de la fréquence cardiaque. Ces modalités supplémentaires auraient pu offrir une meilleure compréhension des réponses physiologiques des patients pendant les activités de rééducation [92, 33].

En outre, bien que des techniques d'augmentation des données aient été utilisées pour pallier les déséquilibres de classe, la variabilité inhérente dans la durée et la fréquence des activités enregistrées pourrait encore influencer les performances du modèle [31].

Pour améliorer la précision et l'interprétabilité des modèles, l'intégration de données provenant d'une gamme plus large de capteurs pourrait être envisagée [92, 32]. Des recherches futures devraient également explorer l'application en temps réel des modèles développés dans les milieux cliniques, afin de permettre une surveillance personnalisée et continue, permettant ainsi aux professionnels de la santé d'adapter dynamiquement les interventions

en fonction des données en temps réel [35]. La collaboration avec des équipes interdisciplinaires, incluant des physiothérapeutes, neurologues et ingénieurs biomédicaux, pourrait faciliter le développement de protocoles de rééducation plus spécialisés et efficaces [28].

L'extension des recherches à d'autres affections neurologiques au-delà de l'AVC pourrait également élargir les applications des modèles basés sur l'intelligence artificielle, contribuant ainsi à des stratégies de rééducation plus globales et efficaces dans divers domaines de la neuro-rééducation [93, 94].

Cette étude présente un processus de recherche complet pour la collecte et l'analyse des données inertielles dans le but de reconnaître les activités physiques chez des individus souffrant de séquelles chroniques d'AVC. Menée au sein du service de neurologie du centre de rééducation Jacques Calvé, Fondation Hopale, à Berck-sur-Mer en France, l'étude a impliqué 26 participants représentant divers stades de récupération post-AVC [95]. Le dispositif de suivi d'activité physique utilisé, développé dans le cadre du projet STARR, comprenait une unité de mesure inertielle et une semelle instrumentée avec des capteurs de pression, garantissant une acquisition précise et confortable des données [32].

L'analyse des données s'est concentrée sur les personnes présentant des troubles de la marche, en utilisant le test d'équilibre de Berg pour évaluer les capacités d'équilibre au cours de sept activités de rééducation. Nous avons catégorisé les participants en cinq groupes en fonction de la gravité des troubles de la marche et avons appliqué des techniques d'augmentation des données pour équilibrer le jeu de données et éviter le surapprentissage [96].

Notre évaluation des architectures d'apprentissage machine et d'apprentissage profond a révélé que les modèles d'apprentissage machine comme XGBOOST et le classifieur Random Forest ont bien performé, tandis qu'un modèle hybride CNN-BiLSTM avec des mécanismes d'attention a atteint une précision de 92% [29]. Les couches d'attention à double niveau dans notre modèle améliorent considérablement l'interprétabilité en identifiant les composants clés du signal et en mettant en évidence les branches influentes au sein de l'architecture [89]. Cette approche permet au modèle de différencier les données temporelles des données issues de la transformation en domaine fréquentiel, en utilisant des branches avec des tailles de noyau variables pour capturer des motifs complexes [91].

Les cartes d'attention, particulièrement celles agrégées par le côté parétique des participants, fournissent des informations précieuses aux thérapeutes en indiquant quels aspects des mouvements des patients sont prioritaires pour le modèle. Cette interprétabilité aide à adapter les stratégies de rééducation personnalisées, en se concentrant sur les motifs de marche et de mouvement les plus pertinents, améliorant ainsi l'efficacité de l'intervention

thérapeutique [97, 98].

3.6. Chapitre 5 : L'auto-apprentissage pour une détection semi-supervisée des événements de chute

3.6.1. Introduction

La prédiction des événements de chute est une tâche essentielle dans les domaines médicaux, particulièrement pour les populations âgées ou atteintes de troubles neurologiques. De nombreux facteurs peuvent influencer l'occurrence des chutes, tels que les troubles de l'équilibre et de la marche, les déficiences cognitives, la faiblesse musculaire, et l'usage de médicaments [99, 100, 101]. Face à cette problématique, plusieurs outils de détection des chutes ont été développés pour aider les professionnels de la santé à identifier les individus à risque. Ces outils combinent généralement les antécédents médicaux, les examens physiques et une évaluation des capacités fonctionnelles [102]. Parmi eux, le test Timed Up and Go (TUG) est souvent utilisé pour évaluer la stabilité de la marche et l'équilibre [103], ainsi que l'échelle de Berg qui permet de mesurer la capacité d'une personne à maintenir son équilibre pendant diverses activités [104].

Cependant, ces outils présentent certaines limites. Par exemple, leur capacité à détecter des événements brefs ou des comportements irréguliers est parfois insuffisante, ce qui peut nuire à leur précision dans la prédiction des chutes [105, 106]. De plus, leur caractère ponctuel ne permet pas de surveiller en continu les risques de chute, ce qui peut rendre difficile l'adaptation des interventions à l'évolution de l'état de l'individu.

Les récentes avancées en intelligence artificielle (IA) ont permis de surmonter certaines de ces limitations. Les modèles d'apprentissage automatique et d'apprentissage profond sont désormais utilisés pour analyser les données recueillies par des capteurs portables, tels que les unités de mesure inertielle (IMU) ou des semelles instrumentées [107, 108]. Ces capteurs permettent de collecter des données sur les schémas de mouvement, que les modèles d'IA analysent pour détecter des variations subtiles indiquant un risque accru de chute [109]. Ces approches offrent des solutions personnalisées aux professionnels de santé, permettant de concevoir des protocoles de réhabilitation préventifs et d'améliorer la qualité de vie des patients [107].

Dans ce contexte, l'explicabilité des modèles d'IA est cruciale. Les professionnels de santé doivent comprendre les prédictions faites par les algorithmes pour valider les hypothèses médicales et intégrer ces systèmes dans leurs processus décisionnels [110, 111]. L'IA explicable (XAI) est ainsi devenue un domaine clé, notamment dans la prédiction des chutes où l'identification des facteurs de risque doit être claire et transparente pour élaborer des

stratégies de prévention efficaces [110, 111].

Cette étude a pour objectif de développer une approche basée sur les données pour détecter les moments présentant un risque élevé de chute, en utilisant des données d'IMU montées sur les chevilles. Notre modèle est conçu pour prédire les scores d'équilibre de Berg, et ainsi identifier les événements potentiels de chute. Grâce à une approche d'apprentissage non contrastif, notre modèle est capable de pré-entraîner un encodeur de données pour extraire les schémas les plus informatifs, tout en mettant en avant les moments critiques via des couches intermédiaires interprétables plutôt que la seule couche de sortie.

3.6.2. Processus de détection des événements rares grâce à l'auto-apprentissage

Dans cette section, nous présentons notre cadre de travail qui utilise les données des capteurs IMU pour prédire les événements de chute. Nous abordons d'abord l'analyse des données collectées, puis nous explorons l'architecture d'apprentissage profond ainsi que le pipeline d'entraînement. Enfin, nous fournissons des détails sur les applications de cet outil d'évaluation des risques de chute.

Les données utilisées dans ce projet proviennent d'un groupe de patients souffrant d'hémiparésie. La Figure 3.19 illustre le nombre de patients touchés par une paralysie partielle de leur pied droit ou gauche.

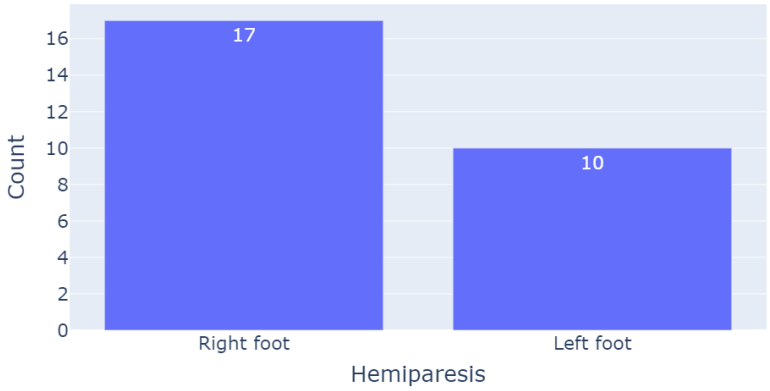

FIGURE 3.19 – Nombre de patients selon le côté touché par l'hémiparésie.

Parmi ces patients, certains utilisaient différents outils d'assistance pour la marche en fonction de leur état. Le Tableau 3.10 présente le nombre de personnes utilisant divers types d'outils de support, en fonction du côté affecté.

TABLE 3.10 – Nombre d'utilisateurs d'outils de support à la marche.

Type de Support	Côté Paretique	
	Pied Gauche	Pied Droit
Canne	7	2
Canne trépied	1	1
Orthèse moulée	1	3
Orthèse avec bande élastique	0	2

Pour permettre à notre modèle d'apprentissage d'apprendre à partir des différents exemples d'événements rares, nous avons segmenté les enregistrements des exercices en segments de 10 secondes, avec un chevauchement de 1,25 seconde. Cette méthode permet d'obtenir des fenêtres de données contenant, en moyenne, 6 pas, représentant divers comportements et caractéristiques de la marche, riches en répétitions. Cette segmentation permet d'acquérir un nombre suffisant de données pour alimenter le modèle.

Nous avons adopté un format de données 3D pour exploiter les corrélations spatiales entre les différentes composantes axiales des signaux d'accélération et de gyroscope. Les données des six canaux (Acc_x, Acc_y, Acc_z, Gyr_x, Gyr_y, Gyr_z) ont été reformatées en matrices 2D de taille 20x20, puis superposées pour former une matrice 3D de taille 20x20x6. Ce format permet au modèle de capturer les interactions spatiales entre les différentes composantes du signal.

Notre modèle comprend une phase de pré-entraînement où il est entraîné à prédire le score d'évaluation de l'équilibre de Berg. Ce score résulte d'une série d'exercices d'équilibre effectués par les participants. Les valeurs de ce score pour les participants varient de 33 à 56, comme le montre la Figure 3.20.

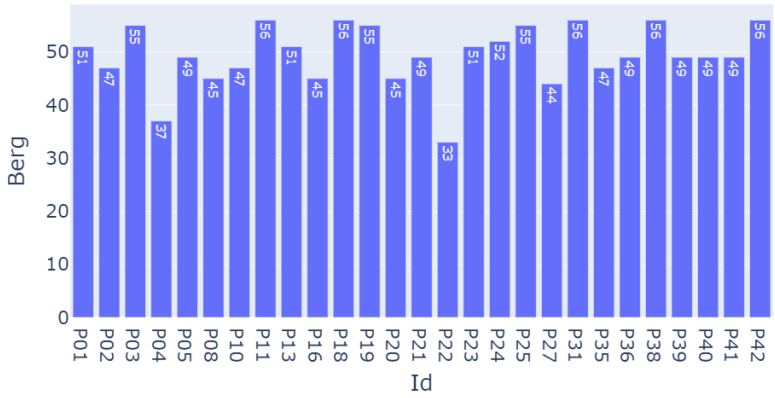

FIGURE 3.20 – Score d'évaluation de Berg pour chaque patient.

Deux approches ont été proposées pour créer le pipeline d'entraînement et définir le type de tâche :

— **Problème de régression :** en conservant le score de Berg dans son format actuel, le modèle développé peut être entraîné à prédire la valeur exacte de ce score ;
— **Problème de classification :** en transformant le score d'équilibre en trois catégories (30–40, 41–50, 51–60), il aurait été possible d'entraîner le modèle sur une tâche de classification multi-étiquette.

La deuxième approche s'est révélée moins efficace en termes de progression de l'entraînement, et la fonction de perte était particulièrement instable. En outre, les résultats de la classification n'étaient pas suffisants. En revanche, la première approche a permis un entraînement plus fluide tout en réduisant l'écart entre les prédictions du modèle et les valeurs réelles du score de Berg. Par conséquent, les sections suivantes présentent les résultats obtenus uniquement à partir de la tâche de régression.

Comme détaillé précédemment, ce modèle utilise un score d'équilibre pour l'entraînement et pour concevoir une représentation compacte et informative des données avec des poids d'importance correspondants. Pour extraire ces poids, nous avons utilisé un mécanisme de self-attention spatiale, mis en œuvre à l'aide de couches convolutionnelles. Ce type d'architecture permet de mettre en évidence les corrélations entre différentes parties du signal ainsi qu'avec d'autres composantes. Dans notre cas, il s'agit de la configuration la plus appropriée, car nous cherchons à identifier les instants du signal qui expliquent le mieux le score de Berg et à détecter les moments à risque de chute. La Figure 3.21 illustre l'architecture de self-attention utilisée dans la phase de pré-entraînement. Cette architecture comprend trois couches de convolution similaires, chacune calculant les "key", "query" et "value". En calculant le softmax du produit scalaire entre "key" et "query", nous avons

obtenu des modèles de similitude utilisés comme scores d'attention. La représentation des "value" a été utilisée pour calculer les valeurs pondérées de l'entrée, multipliées par les scores d'attention. Le paramètre gamma a été utilisé pour permettre au modèle de donner plus ou moins d'importance au module de self-attention dans la création des cartes de caractéristiques pour l'ensemble du réseau.

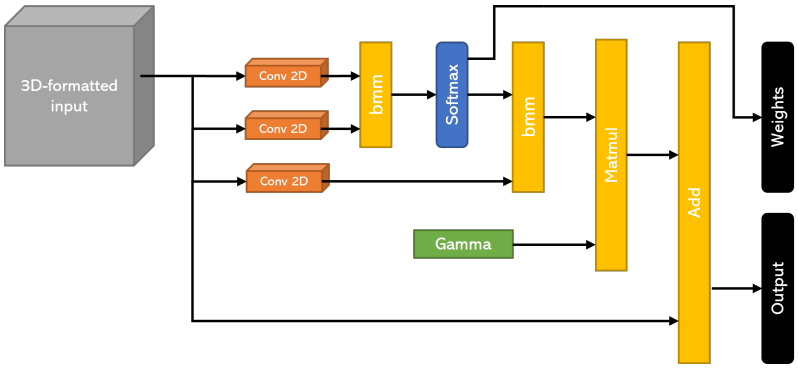

FIGURE 3.21 – L'architecture de self-attention spatiale.

Le mécanisme de self-attention a été conçu pour être utilisé dans une approche d'apprentissage non-contrastif. La Figure 3.22 illustre les étapes de cette approche, qui consistent à créer différentes versions des données d'entrée à l'aide des techniques d'augmentation de données suivantes :

- **Bruit blanc :** ajout de bruit blanc gaussien aléatoire sur les séries temporelles ;
- **Quantification aléatoire :** mappage des valeurs du signal continu en un ensemble discret de valeurs ;
- **Dérive :** dérive aléatoire et douce des valeurs d'une série temporelle ;
- **Inversion :** inversion des séries temporelles.

Nous avons obtenu la série temporelle originale et ses quatre variantes grâce aux techniques d'augmentation. Ensuite, nous avons extrait la représentation des données donnée par l'encodeur de self-attention pour l'entrée originale et les quatre versions augmentées. Le modèle a été entraîné à apprendre les similitudes entre les représentations rendues tout en prédissant le score de Berg via une couche de régression. Cela permet au modèle de créer une représentation informative des données en lien avec l'équilibre, et d'extraire les moments présentant un risque de perte d'équilibre, un facteur direct de chute chez les adultes âgés.

Cette méthode permet de détecter les événements rares, comme les moments de risque de chute, sans utiliser de données étiquetées. Les étiquettes du score de Berg n'ont été utilisées que lors de la phase de pré-entraînement. En extrayant les séquences importantes du signal grâce au module de self-attention, nous avons cherché à identifier les moments

corrélés au score de Berg, pour les comparer avec les phases à plus fort risque de chute. Ensuite, nous avons comparé ces résultats avec une approche basée sur des seuils pour la détection du risque de chute, à partir des données d'accéléromètre.

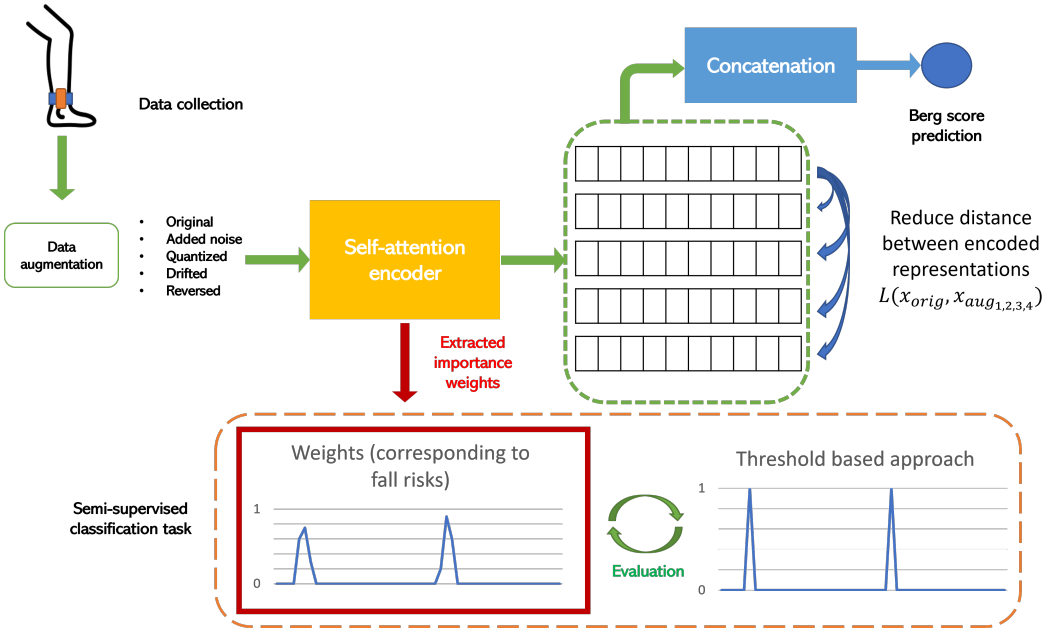

FIGURE 3.22 – Pipeline complet de l'approche d'apprentissage non-contrastif.

Étant donné que notre modèle a été entraîné sur deux tâches simultanément, nous avons utilisé deux fonctions de perte adaptées à chaque tâche. Pour la prédiction du score, nous avons utilisé l'erreur absolue moyenne, définie par l'équation suivante : $l_n = |x_n - y_n|$. Pour l'apprentissage non-contrastif, nous avons utilisé la perte contrastive définie par $y \times d + (1-y) \times max(margin - d, 0)$, où d est la distance euclidienne entre deux échantillons, $d = ||x_1 - x_2||$.

Pour garantir une distribution diversifiée des échantillons, nous avons divisé les données par patient et par activité avec une répartition de 80–20% entre les ensembles de test et d'entraînement. Une base de référence a été utilisée pour évaluer les performances du modèle, en comparant nos résultats avec ceux d'un algorithme détectant les changements soudains d'accélération dans la direction verticale (axe y), avec un seuil fixé à 5 m/s^2.

Nous avons choisi notre meilleur modèle en fonction de l'évolution de la perte, de sa capacité à approcher le score de Berg, et des poids extraits pour la détection des événements à risque de chute. Tout d'abord, les fonctions de perte (illustrées dans la Figure 3.23) étaient stables et diminuaient progressivement au fil des époques. Dans d'autres cas, des fluctuations et une évolution instable dans le temps étaient observées. Dans de rares cas, les courbes de perte, en particulier la perte non-contrastive, divergeaient et le

modèle n'était pas capable d'extraire les poids pertinents. Pour ce modèle sélectionné, les paramètres fixes sont détaillés dans le Tableau 3.11.

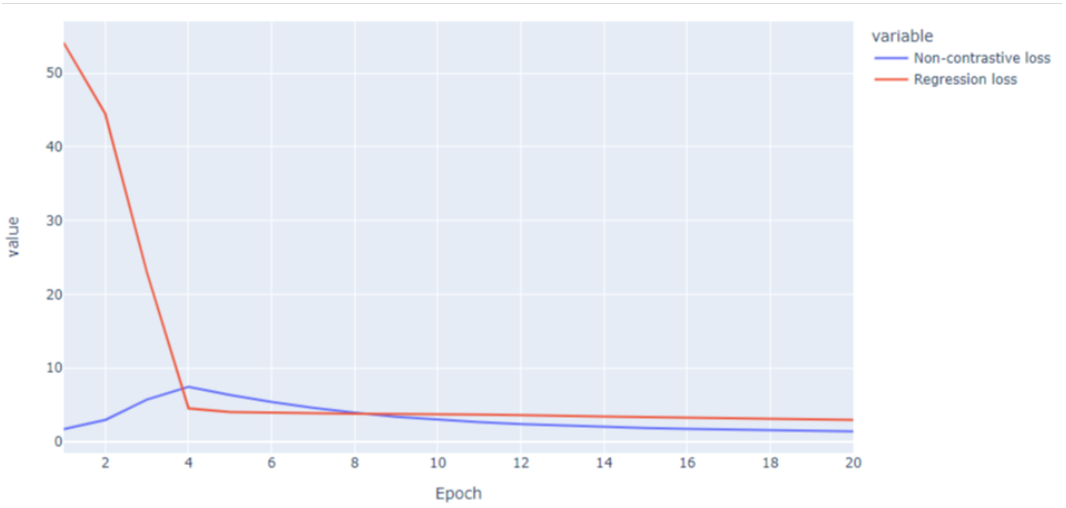

FIGURE 3.23 – Fonctions de perte pour les tâches d'apprentissage non-contrastif et de régression.

TABLE 3.11 – Hyperparamètres optimisés.

Hyperparamètre	Taille du Batch	Taux d'Apprentissage	Époques	Taille du noyau
Valeur optimisée	4	1×10^{-3}	20	1

3.6.3. Quantification des risques de chutes à travers les scores d'attention du modèle auto-entrainé

Nous avons utilisé la deuxième sortie du modèle, qui est le poids de chaque instant du signal dans l'explication du score de Berg pour chaque essai. La Figure 3.24 montre un exemple de projection de ces poids sur le signal, superposé à l'intrigue des poids ainsi qu'aux étiquettes de référence indiquant les instants où les composantes axiales du signal d'accélération ont dépassé la valeur de 5 m/s^2. Dans d'autres cas, le modèle pouvait également suivre une multitude de possibilités de risque, tout en accordant plus d'importance aux chutes potentielles à haut risque, comme illustré dans la Figure 3.25.

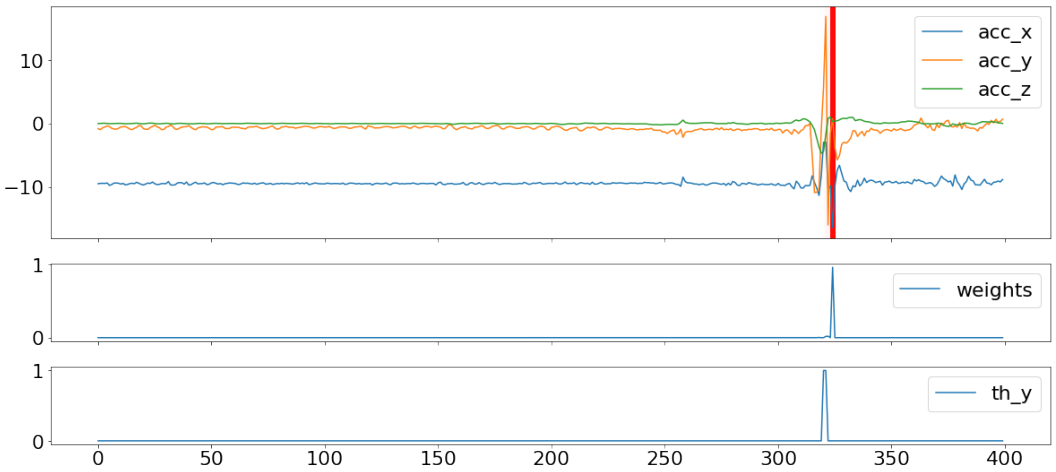

FIGURE 3.24 – Projection des poids sur les signaux IMU pour un événement à risque. (La ligne rouge représente la projection des valeurs élevées des poids).

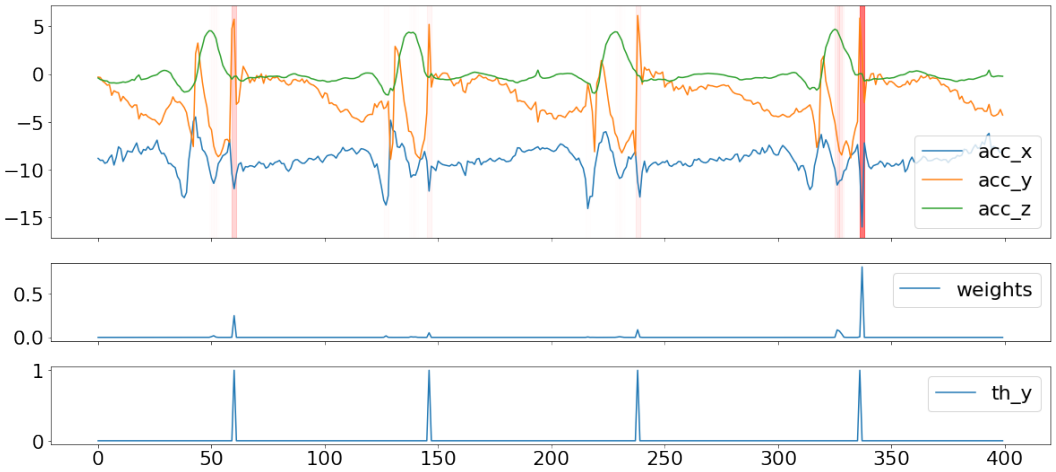

FIGURE 3.25 – Projection des poids sur les signaux IMU pour une multitude d'événements à risque avec des poids d'importance différents. (Les lignes rouges représentent la projection des valeurs élevées des poids).

Le modèle a pu prédire parfaitement 1476 cas de chutes potentielles parmi les 8858 cas détectés par l'approche basée sur les seuils. La précision globale du modèle a atteint 98,9% en raison du déséquilibre des classes dans les données. La matrice de confusion (illustrée à la Figure 3.26) liste les moments de risque correctement étiquetés ainsi que les données mal étiquetées. Par conséquent, nous avons concentré notre analyse sur la classe "risque de chute".

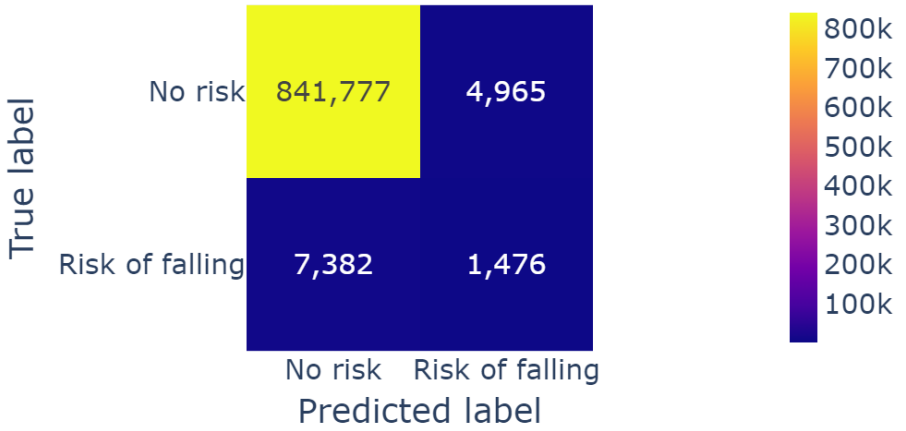

FIGURE 3.26 – Matrice de confusion pour les instants à risque de chute.

Étant donné que notre modèle détecte des instants spécifiques de risque de chute de manière temporelle, il peut prédire des moments de chute avant ou après l'instant de risque étiqueté. Si la prédiction de risque du modèle survient avant la prédiction du risque de référence, cela est bénéfique d'un point de vue préventif. De plus, si le modèle prédit une chute après son occurrence, certaines méthodes suggèrent d'étendre la prédiction sur une petite fenêtre de temps, appelée intervalle de confiance. Pour cette raison, nous avons évalué les Vrais Positifs (TP), Vrais Négatifs (TN), Faux Positifs (FP), et Faux Négatifs (FN) à l'échelle de la fenêtre tout en comparant différentes tailles de fenêtres d'évaluation.

Nous avons sélectionné trois modes d'évaluation pour les différents scénarios possibles :

- Valeur prédite avant la référence temporelle (Figure 3.27) ;
- Valeur prédite après la référence temporelle (Figure 3.28) ;
- Intervalle de confiance où l'étiquette de référence se situe au milieu de la fenêtre de temps (Figure 3.29).

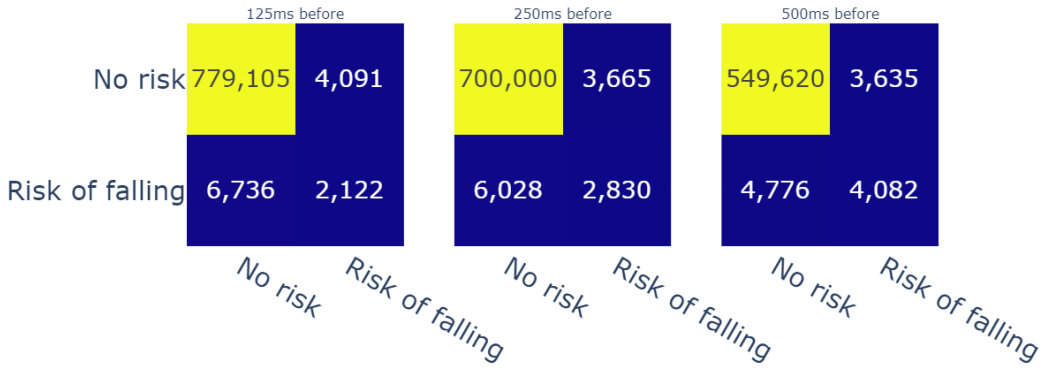

FIGURE 3.27 – Matrices de confusion pour les différentes durées de fenêtres temporelles (avant référence).

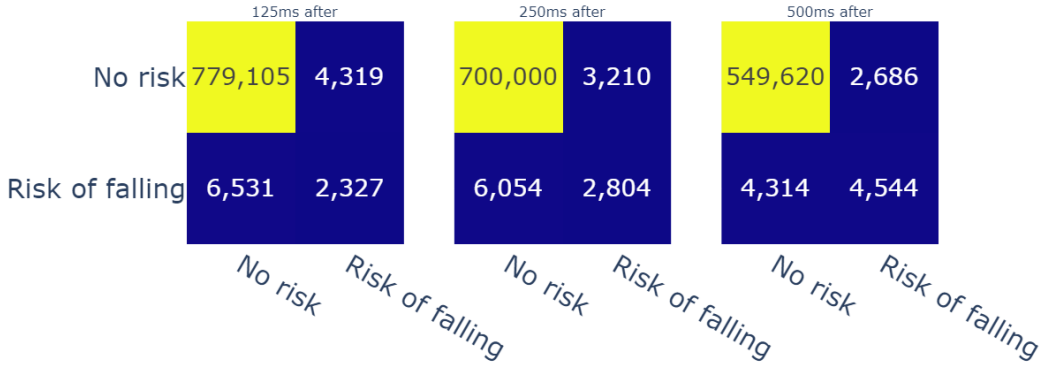

FIGURE 3.28 – Matrices de confusion pour les différentes durées de fenêtres temporelles (après référence).

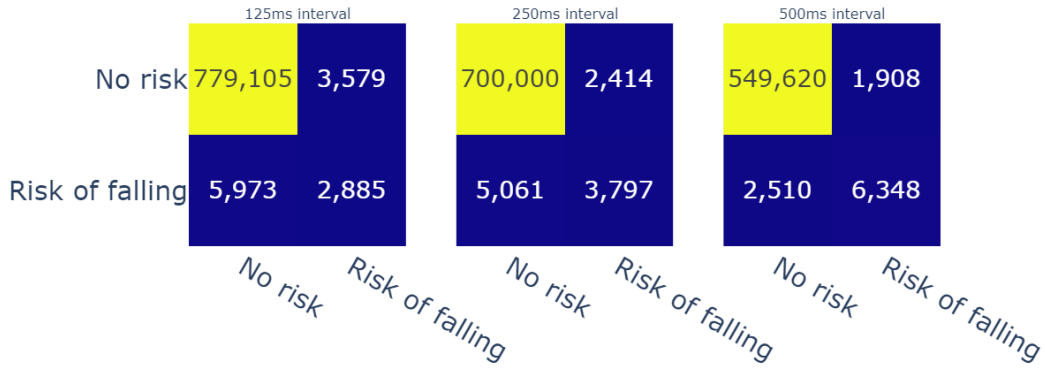

FIGURE 3.29 – Matrices de confusion pour les différentes durées de fenêtres temporelles (intervalle de confiance).

Pour les thérapeutes, détecter un risque de chute à l'avance est particulièrement important d'un point de vue préventif. De plus, les systèmes intelligents de prévention des chutes visent généralement à lancer des alertes pendant la phase de transition entre l'« activité quotidienne » (ADL) et la zone de chute libre. Cette phase de transition peut durer un court laps de temps, variant entre 150 ms [112] et 300–500 ms [113]. Compte tenu de ces durées, nous avons évalué les prédictions de notre modèle avec des fenêtres temporelles de 125 ms, 250 ms, et 500 ms avant les étiquettes de référence. Les résultats obtenus montrent que le modèle a détecté jusqu'à 4082 moments de risque de chute (cf. Figure 3.27). Nous avons également évalué les événements détectés après l'instant de risque de référence avec les mêmes fenêtres temporelles. Le modèle a détecté jusqu'à 4544 moments de risque (cf. Figure 3.28). De plus, en créant un intervalle de confiance, le modèle a détecté 6348 événements de risque sur 8858 cas (cf. Figure 3.29), ce qui représente 71,6% des cas.

Cette évaluation nous a permis d'améliorer le rappel du modèle de plus de 55%, atteignant un score de précision de 77%.

3.6.4. Conclusion

Dans ce travail, nous proposons une approche d'apprentissage sans contraste pour prédire, à l'aide du score d'évaluation de l'équilibre de Berg, le risque de chute chez les adultes et les personnes âgées ayant survécu à un AVC. Cette approche présente plusieurs avantages majeurs :

— Elle peut être entraînée sur un petit nombre de données non étiquetées, ce qui est particulièrement utile dans des contextes cliniques où l'étiquetage manuel peut être

coûteux ou difficile ;
- Elle utilise des architectures interprétables basées sur les métadonnées afin d'extraire des caractéristiques significatives ;
- Le modèle exploite la relation de cause à effet entre le score d'équilibre de Berg et la variable cible, à savoir le risque de chute.

Les résultats obtenus sont particulièrement intéressants en termes de qualité de détection. En effet, le modèle a atteint un score de rappel de 71% et un score de précision de 77% pour des cas où un intervalle de confiance de 1 seconde a été utilisé (soit 500 ms avant et après la prédiction de référence). Il est important de noter que le modèle n'a pas été entraîné pour détecter directement et de manière supervisée ces événements de risque de chute. Au contraire, il a été conçu pour fournir des approximations du score de Berg, en utilisant les atouts d'interprétabilité du modèle afin d'établir le lien entre ce score et le risque de chute. Ce lien a été identifié en extrayant l'importance de chaque instant temporel, et il est apparu que le modèle a repéré des moments avec des valeurs d'accélération extrêmes et des évolutions soudaines de ces dernières.

Ainsi, le modèle a réussi à obtenir des scores satisfaisants, d'autant plus que ce dernier extrait la relation entre le score de Berg et les valeurs d'accélération afin de détecter les moments de risque de chute, et ce sans l'utilisation de données annotées spécifiquement pour cette tâche. De plus, bien que l'approche basée sur un seuil de détection soit solide et montre de bons résultats, elle présente certaines limites en termes de généralisation et d'adaptation aux profils des patients, notamment du fait que les étiquettes de référence peuvent manquer de fiabilité.

Par ailleurs, cette étude ouvre la voie à de nombreuses pistes d'améliorations possibles. Le modèle de type CNN avec auto-attention utilisé dans cette étude n'a été comparé qu'à l'approche par seuil. D'autres architectures interprétables pourraient être testées et adaptées à ce cas d'usage. De même, il serait pertinent de comparer ce modèle à des approches de détection plus fiables ou à des données annotées. En outre, la pipeline d'apprentissage sans contraste devrait être entraînée sur des jeux de données plus vastes afin d'extraire différents types de schémas et d'informations. Il serait également envisageable de tester la méthode d'apprentissage auto-supervisé contrastif pour ce problème, en particulier afin de permettre au modèle de distinguer entre les schémas de marche normaux et les séquences anormales susceptibles de mener à des chutes.

D'un point de vue pratique, il a été observé que le modèle avait tendance à prédire le risque de chute dans l'intervalle qui précédait l'étiquetage de référence. Ainsi, cette approche pourrait servir d'outil de prévention pour l'évaluation du risque de chute. En termes d'utilité, ce projet vise à concevoir un outil interprétable d'aide à la décision pour

assister les professionnels de santé dans leurs tâches quotidiennes, en particulier pour l'analyse de la marche et la gestion du risque de chute chez les personnes âgées [107, 110]. Les objectifs de cet outil pourraient évoluer vers le profilage des patients et la suggestion d'indicateurs de rééducation adaptés, tels que l'utilisation des phases de marche pour améliorer la prévention des chutes, ou encore l'usage et le moment adéquat pour changer un support de marche [111]. De plus, cet outil pourra être utilisé pour surveiller les patients dans différentes configurations, en dehors des sessions cliniques de rééducation [114].

Afin de répondre à ces besoins, nous nous attacherons à construire un modèle plus robuste permettant également d'obtenir une sortie interprétable en fonction des données d'entrée. En intégrant davantage de métadonnées et d'informations provenant des dossiers médicaux des patients, le modèle pourrait gagner en efficacité et en fiabilité, rassurant ainsi les utilisateurs de l'outil. De plus, cela permettrait d'améliorer les sorties interprétables avec des indicateurs plus pertinents pour le personnel médical [110].

3.7. Conclusion

Ce volet a présenté les enjeux et les avancées liés à la HAR pour la rééducation du bas du corps, en se concentrant sur l'utilisation des capteurs inertiels, tels que les accéléromètres et les gyroscopes, afin de surveiller et d'évaluer les capacités de marche des patients souffrant de troubles neurologiques, de la vieillesse ou ayant subi des interventions chirurgicales. L'importance de ces technologies ne réside pas uniquement dans leur capacité à collecter des données précises, mais aussi dans leur capacité à fournir des informations continues et en temps réel, facilitant ainsi une adaptation personnalisée des programmes de rééducation [28, 31].

La reconnaissance de la marche à l'aide de capteurs IMU s'est avérée être une solution innovante pour diagnostiquer et suivre l'évolution des troubles de la marche, en particulier chez les patients post-AVC ou en rééducation après une chirurgie orthopédique. Les technologies portables permettent de capturer les phases critiques de la marche et d'analyser la stabilité, la symétrie et la fluidité des mouvements. L'acquisition de ces données permet de fournir aux cliniciens des informations objectives sur les progrès des patients, facilitant ainsi des ajustements en fonction des résultats [61, 59].

Le projet STARR, utilisé pour cette étude, a fourni une base de données riche et diversifiée. Celle-ci a servi de support à quatre projets majeurs portant sur la HAR et sur la détection des anomalies de marche. Ces projets ont exploré plusieurs méthodologies, telles que les réseaux de neurones convolutifs (CNN), les LSTM, ainsi que des approches multimodales combinant des données vidéo et des données issues de capteurs inertiels [65, 46]. En

s'appuyant sur ces différents paradigmes, les modèles ont été capables de capturer les subtilités des schémas de marche, telles que les asymétries, les irrégularités et les risques de chute.

Bien que ces approches aient montré une grande précision, plusieurs défis persistent. L'un des principaux défis est celui de l'interprétabilité des modèles d'apprentissage profond, qui sont souvent considérés comme des "boîtes noires". La nécessité d'expliquer les décisions prises par ces modèles est primordiale pour garantir leur adoption par les professionnels de santé [88]. De plus, la gestion des données non étiquetées, qui constituent une part importante des données collectées, nécessite le développement de techniques d'apprentissage non supervisé ou auto-supervisé pour garantir une meilleure généralisation des modèles [28, 91]. Les approches d'apprentissage sans contraste ou auto-supervisées, comme celles proposées dans le cadre de cette étude, ont permis de surmonter certaines de ces limitations en offrant une meilleure gestion des données non étiquetées, tout en extrayant des schémas pertinents pour la prédiction des risques [110].

Un autre aspect critique de cette recherche est l'évaluation des risques de chute. Les risques de chute sont des événements complexes qui nécessitent une compréhension fine des schémas de marche et des changements subtils dans les mouvements du corps. Les modèles proposés ont non seulement détecté ces risques en temps réel, mais ils ont également permis de fournir un retour d'information préventif aux cliniciens et aux patients, offrant ainsi une opportunité d'intervention précoce [31, 111]. La prévention des chutes, en particulier chez les personnes âgées ou les patients souffrant de déficiences motrices, constitue une priorité dans les soins de rééducation et ce projet contribue à cette mission en rendant la détection des risques de chute plus précise et plus proactive.

Enfin, bien que l'étude ait montré des résultats prometteurs, plusieurs pistes d'amélioration ont été identifiées. D'une part, l'intégration de données provenant de multiples capteurs, tels que les moniteurs de fréquence cardiaque ou les capteurs de pression dans les semelles, pourrait enrichir l'analyse des réponses physiologiques des patients pendant les activités de rééducation [28, 33]. D'autre part, l'application de ces technologies dans des contextes cliniques en temps réel pourrait transformer la manière dont les professionnels de santé gèrent les programmes de rééducation, rendant les interventions plus efficaces et adaptées aux besoins individuels [107].

En conclusion, la reconnaissance d'activités humaines via les capteurs inertiels représente une avancée significative pour améliorer les protocoles de rééducation. L'utilisation des modèles d'apprentissage profond, combinée à l'interprétabilité et à la gestion des données non étiquetées, ouvre de nouvelles perspectives pour des interventions personnalisées et

en temps réel. Toutefois, des efforts supplémentaires sont nécessaires pour optimiser ces modèles, améliorer leur robustesse et intégrer davantage de modalités de données afin de rendre ces outils plus polyvalents et fiables dans les environnements cliniques [88, 111].

Volet 2 : Reconnaissance d'activité humaine saine pour la détection de la main manipulatrice et de l'action

4 | Volet 2 : Reconnaissance d'activité humaine saine pour la détection de la main manipulatrice et de l'action

4.1. Introduction

La reconnaissance d'activité humaine (HAR) est devenue un domaine clé dans les applications de vision par ordinateur, allant de la surveillance intelligente à l'interaction homme-machine, en passant par les soins de santé et la robotique. Un sous-domaine particulièrement intéressant de cette recherche est la détection des actions de manipulation des mains et des objets. Cette tâche est complexe, car elle implique la capture précise des mouvements des mains, la compréhension des interactions avec les objets, ainsi que la reconnaissance des gestes et des actions dans divers environnements [115, 116, 117].

Dans les environnements où les caméras captent les mouvements du point de vue de l'utilisateur (approche égocentrique), il est possible d'obtenir des informations détaillées sur les actions effectuées avec les mains. Par exemple, le jeu de données *Epic-Kitchens* comprend des vidéos enregistrées avec des caméras portées sur la tête, capturant des actions de la vie quotidienne dans des cuisines. Ce type de données permet aux chercheurs de développer des modèles capables de reconnaître des actions complexes comme « ouvrir un réfrigérateur » ou « verser du lait », en tenant compte des interactions spécifiques entre les mains et les objets [116].

Les méthodes modernes pour la détection des mains s'appuient largement sur des réseaux de neurones profonds et des architectures de type Transformer, qui permettent d'analyser efficacement les séquences temporelles complexes. Ces modèles, tels que les Transformers basés sur la vision (*Vision Transformers*), utilisent des informations clés comme les positions des mains et des objets dans les images pour classer les actions [118].

Une particularité de ces approches est l'utilisation de techniques d'attention automatique pour déterminer quels objets ou gestes sont les plus significatifs à un moment donné, améliorant ainsi la précision de la reconnaissance d'action [119].

La détection des actions de la main ne se limite pas à la reconnaissance de mouvements basiques, mais inclut des interactions complexes entre plusieurs objets et des gestes fins comme la manipulation de petits objets. Les techniques de détection d'objets, telles que YOLOv3, ont également été utilisées pour améliorer la détection des régions de la main dans des flux vidéo égocentriques. Cela permet de générer des « boîtes de délimitation » précises autour des mains et des objets d'intérêt, réduisant ainsi les faux positifs dans des environnements dynamiques [120].

Dans les contextes réels, la reconnaissance des activités humaines offre des opportunités importantes, comme dans le domaine de la santé, où elle peut aider à surveiller les patients en rééducation ou à identifier des situations d'urgence, par exemple en détectant une chute chez une personne âgée. Les robots d'assistance peuvent également être équipés de ces systèmes pour offrir des services plus réactifs et sécurisés [121].

4.2. Chapitre 1 : Présentation du jeu de données OPPORTUNITY++

4.2.1. Détails sur le jeu de données OPPORTUNITY++

Le jeu de données OPPORTUNITY++ est une extension multimodale du jeu de données *OPPORTUNITY*, conçu pour la reconnaissance des activités humaines (*Human Activity Recognition*, HAR). Ce jeu de données combine des données provenant de capteurs portés sur le corps, des objets instrumentés, des capteurs d'ambiance, ainsi que des vidéos et des informations de suivi squelettique via *OpenPose*.

Ce jeu de données a été collecté dans un environnement domestique (cuisine) où quatre participants ont réalisé diverses tâches quotidiennes, telles que la préparation de repas, la prise de café, et le nettoyage. Voici une description plus détaillée des différentes sources de données utilisées :

- **Capteurs portés sur le corps** : Les capteurs incluent sept unités de mesure inertielle et douze accéléromètres 3D. Ces capteurs sont placés sur différentes parties du corps des participants, principalement sur le torse, les bras et les jambes. Ils mesurent les données brutes telles que l'accélération, la vitesse angulaire et l'orientation dans l'espace.
- **Capteurs d'objets** : Douze objets courants utilisés dans la cuisine (comme un couteau, une assiette, un verre, etc.) sont équipés de capteurs sans fil qui mesurent l'accélération 3D et le taux de rotation. Cela permet de suivre l'utilisation des objets durant les activités.
- **Capteurs d'ambiance** : Ces capteurs incluent des interrupteurs et des accéléromètres installés dans des tiroirs, des portes et des appareils ménagers (réfrigérateur, lave-vaisselle, etc.). Ils sont capables de détecter des événements tels que l'ouverture/fermeture de portes ou le déclenchement d'interrupteurs.
- **Vidéos** : Les vidéos, enregistrées à une résolution de 640x480 pixels à 10 images par seconde, fournissent une vue latérale des participants pendant leurs activités. Ces vidéos sont anonymisées grâce à la pixelisation des visages et sont accompagnées d'annotations sous forme de fichiers SRT.
- **Suivi squelettique** : Les vidéos sont traitées avec *OpenPose*, qui génère des coordonnées *(x, y)* pour 25 points corporels clés (par exemple, le nez, les poignets, les hanches, etc.), permettant ainsi une analyse détaillée des mouvements corporels.

Le jeu de données contient plus de **19,75 heures** de données avec plus de **24 000 annotations**, réparties en plusieurs catégories d'actions, y compris les mouvements des bras,

la locomotion et les interactions avec les objets [122].

4.2.2. Protocole et récolte de données

La collecte des données pour OPPORTUNITY++ s'est déroulée dans un environnement contrôlé, mais conçu pour imiter un cadre domestique naturel, en l'occurrence une cuisine équipée de capteurs. Les participants ont été invités à réaliser des activités quotidiennes de manière libre et naturelle dans un scénario dit *Activités de la Vie Quotidienne (ADL)* ainsi que des exercices répétitifs structurés, appelés *Drill Run*. Ce protocole a été mis en place afin de capturer à la fois des actions spontanées et répétitives, en s'appuyant sur différents types de capteurs et des annotations complexes.

Le scénario ADL est conçu pour simuler un ensemble d'activités domestiques courantes et naturelles. Chaque participant a effectué cinq séquences d'activités quotidiennes dans un ordre flexible, avec la possibilité de réaliser ces activités à leur propre rythme, imitant ainsi la variabilité présente dans la vie réelle. Les activités incluaient :

- **Se lever d'un fauteuil** : Les participants commencent l'activité en position allongée sur un fauteuil, puis se lèvent pour entamer la suite des actions.
- **Se préparer un café** : Cette tâche consistait à préparer un café avec du lait et du sucre, en utilisant une machine à café. Les mouvements précis incluaient des actions manuelles telles que la manipulation de tasses, de cuillères et de la machine à café elle-même.
- **Préparer un sandwich** : La préparation d'un sandwich nécessitait l'utilisation d'un couteau pour couper le pain, le fromage et le salami, ainsi que l'utilisation d'assiettes et autres ustensiles de cuisine. Cette activité générait une série d'actions primitives, telles que saisir des objets, les manipuler, les couper et les ranger.
- **Nettoyage** : Après avoir pris leur repas, les participants devaient ranger les ustensiles utilisés dans le lave-vaisselle ou les remettre en place, ainsi que nettoyer la table.
- **Prendre un temps de repos** : Les participants revenaient finalement au fauteuil pour un temps de repos, ce qui simule des transitions naturelles entre des périodes d'activité et de repos.

Ces scénarios étaient suffisamment libres pour permettre une grande variété d'exécutions, ce qui rendait la tâche de reconnaissance des actions plus complexe et plus représentative des situations réelles [123].

Le scénario Drill Run, quant à lui, est beaucoup plus structuré et répétitif. Il est conçu pour capturer des actions élémentaires en boucle dans un cadre plus contrôlé. Chaque

participant a effectué 20 répétitions d'une séquence préétablie de tâches simples, telles que :

- **Ouvrir et fermer le réfrigérateur** : Cette activité consiste en des manipulations répétitives de la porte du réfrigérateur, fournissant des données sur des actions de portée et de fermeture.
- **Ouvrir et fermer le lave-vaisselle** : Les participants ouvraient et refermaient le lave-vaisselle de façon répétitive.
- **Manipulation des tiroirs** : Chaque participant manipulait trois tiroirs situés à différentes hauteurs, permettant de capturer des actions complexes comme la flexion du dos et la manipulation des objets situés à différents niveaux.
- **Activer et désactiver les interrupteurs** : Cette tâche incluait l'utilisation répétée de plusieurs interrupteurs pour allumer et éteindre des lumières, simulant un contexte domestique classique.
- **Boire debout puis assis** : Les participants devaient boire de l'eau tout en passant d'une position debout à une position assise, capturant ainsi des transitions d'état et de posture.

Le Drill Run a été conçu pour fournir un grand nombre d'exemples d'actions élémentaires dans des conditions identiques, facilitant ainsi l'apprentissage automatique basé sur des mouvements répétitifs [122].

Pour chaque activité, plusieurs types de capteurs étaient utilisés pour capturer les mouvements des participants et leurs interactions avec l'environnement :

- **Capteurs portés sur le corps** : Sept unités de mesure inertielle et douze accéléromètres 3D étaient placés sur les bras, le torse, les jambes et les chaussures des participants. Ces capteurs mesuraient des paramètres tels que l'accélération, la vitesse angulaire et l'orientation des parties du corps.
- **Capteurs d'objets** : Douze objets de la cuisine, tels que des couteaux, des assiettes et des verres, étaient équipés de capteurs d'accélération et de rotation. Cela permettait de détecter avec précision leur manipulation par les participants.
- **Capteurs d'ambiance** : Des interrupteurs et des accéléromètres étaient placés dans les tiroirs, les portes du réfrigérateur et les appareils de cuisine (comme le lave-vaisselle), permettant ainsi de mesurer l'utilisation des éléments de l'environnement.
- **Vidéos** : Les vidéos fournissaient une vue latérale des participants, enregistrant leurs mouvements à 10 images par seconde. Les vidéos étaient synchronisées avec les autres capteurs, facilitant ainsi une analyse multimodale des actions.

Les annotations ont été réalisées à partir des enregistrements vidéo. Elles incluent des

informations sur les actions des bras (par exemple, ouvrir, fermer, remuer, etc.), les interactions avec les objets, et les changements de posture (par exemple, marcher, s'asseoir, se tenir debout). Chaque annotation est classée en fonction de son niveau de granularité, allant des actions de bas niveau (par exemple, mouvements des doigts) aux actions de haut niveau (par exemple, manger, boire) [123].

4.2.3. Analyse des données

Le jeu de données OPPORTUNITY++ constitue une ressource riche pour l'analyse et la modélisation des activités humaines. Avec ses multiples sources de données (capteurs inertiels, capteurs d'objets, capteurs d'ambiance et vidéos), ce jeu de données permet d'explorer diverses approches d'analyse, allant de la détection des gestes fins à l'analyse contextuelle d'interactions complexes. Voici quelques-unes des perspectives analytiques et techniques utilisables pour exploiter ce jeu de données.

Les annotations disponibles dans OPPORTUNITY++ permettent de décomposer les actions des participants à plusieurs niveaux de granularité. Cela permet de comprendre les actions complexes à partir de gestes élémentaires. Les pistes d'annotations incluent :

- **Actions des bras gauche et droit** : Les gestes des bras, tels que saisir, remuer, couper, ouvrir et fermer des objets, sont annotés de manière précise. Chaque action est également associée à l'objet manipulé, comme une cuillère, un couteau, ou une tasse, ce qui permet une analyse conjointe des gestes et de l'interaction avec les objets.
- **Locomotion** : Cette piste décrit les états de locomotion (marche, station debout, assise, couchée), permettant de suivre les transitions entre différentes postures et actions physiques. Ces informations sont essentielles pour des modèles de reconnaissance globale des comportements.
- **Actions de niveau intermédiaire (Mid-Level)** : Les actions comme ouvrir une porte ou un tiroir, ou allumer un interrupteur, sont regroupées dans cette catégorie. Ces actions impliquent des interactions avec l'environnement, souvent réalisées avec une ou deux mains.
- **Interactions homme-objet** : Les annotations détaillent les interactions spécifiques avec des objets instrumentés, comme les tasses, les couteaux et les verres. Cela permet une analyse fine de l'utilisation des objets et leur manipulation dans des scénarios domestiques.

Une des principales forces du jeu de données OPPORTUNITY++ est la possibilité de combiner des sources de données multimodales pour améliorer la précision et la robustesse des systèmes de reconnaissance des activités. La fusion des données issues des cap-

teurs inertiels, des objets et des vidéos offre une vue complète des actions humaines dans l'environnement. Voici quelques approches courantes :

- **Données des capteurs inertiels (IMU)** : Les IMU mesurent l'accélération linéaire, la vitesse angulaire et l'orientation spatiale du corps. Ces données permettent de suivre les mouvements corporels en temps réel et d'identifier les actions dynamiques comme la marche ou les gestes des bras.
- **Données des objets et de l'environnement** : Les capteurs d'objets fournissent des informations sur l'utilisation des ustensiles de cuisine, comme les couteaux ou les verres. Ces données sont complémentaires aux informations des capteurs inertiels et permettent de capturer les actions d'interaction avec les objets du quotidien.
- **Données vidéo et suivi squelettique (OpenPose)** : Les vidéos, combinées avec le suivi squelettique via OpenPose, apportent une dimension visuelle, permettant d'analyser les mouvements corporels complets. Cela est particulièrement utile pour les actions complexes impliquant plusieurs parties du corps, comme se pencher pour ouvrir un tiroir tout en tenant un objet.
- **Données synchronisées** : Toutes les sources de données sont synchronisées, permettant une analyse conjointe. Par exemple, les vidéos peuvent être utilisées pour valider ou corriger les données des capteurs inertiels ou des objets, et inversement.

Plusieurs techniques d'apprentissage automatique peuvent être appliquées à l'analyse des données multimodales de OPPORTUNITY++. Voici quelques approches courantes :

- **Apprentissage profond (Deep Learning)** : Les réseaux de neurones convolutifs (CNN) sont particulièrement adaptés pour traiter les données spatiales issues des vidéos et des capteurs d'image, tandis que les réseaux de neurones récurrents (RNN) et les LSTM (Long Short-Term Memory) permettent de modéliser les séquences temporelles dans les données capturées par les capteurs inertiels et les vidéos.
- **Fusion de capteurs** : La fusion des données provenant des capteurs inertiels, des objets et des vidéos permet de combiner des informations spatiales et temporelles pour obtenir une reconnaissance des activités plus précise. Par exemple, un geste détecté par un capteur inertiel peut être confirmé par l'analyse vidéo ou l'interaction avec un objet.
- **Apprentissage supervisé et non supervisé** : Des algorithmes d'apprentissage supervisé tels que les forêts aléatoires (Random Forest), les machines à vecteurs de support (SVM) ou les réseaux neuronaux profonds sont largement utilisés pour classifier les activités. Des techniques non supervisées peuvent également être utilisées pour explorer des motifs dans les données sans annotation préalable.
- **Segmentation temporelle** : Les techniques de segmentation temporelle permettent de découper les données en intervalles cohérents pour l'analyse des ac-

tivités. Cette segmentation est cruciale pour identifier le début et la fin d'une activité, en distinguant les actions significatives des périodes inactives.

— **Réduction de la dimensionnalité** : Les données brutes, particulièrement volumineuses dans ce type de jeu de données, peuvent être réduites en utilisant des techniques comme l'analyse en composantes principales (PCA) ou les autoencodeurs, permettant de conserver les informations essentielles tout en minimisant le bruit.

Le jeu de données OPPORTUNITY++ ouvre la voie à de nombreuses applications pratiques dans divers domaines :

— **Environnements intelligents et domotiques** : Les données permettent d'automatiser des environnements domestiques intelligents, en reconnaissant les actions des utilisateurs pour déclencher des actions automatiques, comme l'allumage des lumières ou l'ouverture des portes.

— **Assistance aux personnes âgées ou à mobilité réduite** : Le suivi des actions quotidiennes permet de détecter des situations anormales ou à risque (comme une chute ou un comportement inattendu), facilitant ainsi le déploiement de systèmes d'assistance à domicile.

— **Analyse des performances sportives ou de rééducation** : Le suivi précis des mouvements corporels, couplé aux données d'interactions avec les objets, peut être utilisé pour analyser et améliorer les performances physiques dans un contexte de rééducation ou d'entraînement sportif.

— **Fusion multimodale pour la réalité virtuelle (VR)** : Les données capturées par les capteurs inertiels et les vidéos peuvent être utilisées pour améliorer l'immersion et la précision des mouvements dans des environnements de réalité virtuelle ou augmentée.

4.3. Chapitre 2 : L'utilisation de l'auto-apprentissage pour le remplissage des données manquantes

4.3.1. Introduction

Dans le cadre du jeu de données OPPORTUNITY++, les capteurs IMU jouent un rôle essentiel en fournissant des informations sur les mouvements des participants. Cependant, en raison de problèmes courants tels que la perte de signal ou les interférences, il arrive que certaines données de ces capteurs soient manquantes. Cela peut réduire la performance des modèles de HAR. L'auto-supervision (*self-supervision*) émerge comme une solution puissante pour pallier ce problème en permettant à un modèle d'apprendre à combler les lacunes des données sans recourir à une supervision externe explicite.

L'auto-supervision est une technique d'apprentissage automatique où les modèles apprennent à partir des structures internes des données elles-mêmes. Contrairement à l'apprentissage supervisé traditionnel, qui nécessite des labels manuels, l'auto-supervision génère automatiquement des tâches auxiliaires (aussi appelées prétextes) en exploitant des parties des données disponibles pour prédire ou reconstruire des éléments manquants.

Dans le contexte des capteurs IMU, cette approche est particulièrement efficace pour prédire les valeurs manquantes, en utilisant les relations temporelles et spatiales présentes dans les données.

Plusieurs approches d'auto-supervision peuvent être appliquées spécifiquement au problème de remplissage des données manquantes des capteurs IMU. Ces méthodes exploitent les dépendances intrinsèques des données pour générer des signaux d'apprentissage sans avoir besoin de labels externes.

Une technique couramment utilisée en auto-supervision est l'apprentissage par prédiction de séquences. Le modèle est formé pour prédire les étapes futures d'une séquence en se basant sur les étapes précédentes. Pour les capteurs IMU, cela signifie que le modèle peut apprendre à prédire les lectures de capteurs futures ou manquantes en s'appuyant sur les lectures des autres capteurs à des moments antérieurs.

Les réseaux de neurones récurrents (RNN) et les architectures LSTM (Long Short-Term Memory) sont bien adaptés à ce type de tâche, car ils peuvent capturer les relations temporelles dans les données séquentielles.

— **Masquage de données temporelles (Temporal Masking)** : Cette technique consiste à masquer une portion des séquences temporelles pendant l'entraînement,

et à entraîner le modèle à prédire les valeurs manquantes. Cela force le modèle à comprendre les relations entre les étapes temporelles des données et à utiliser cette compréhension pour compléter les séquences manquantes.

— **Prédiction de la prochaine étape (Next-Step Prediction)** : Le modèle apprend à prédire l'état suivant d'un capteur IMU en utilisant les états précédents. Ce processus d'auto-supervision permet de généraliser cette capacité à reconstruire des données absentes lors de l'inférence.

Les autoencodeurs auto-supervisés constituent une autre approche populaire pour remplir les données manquantes. Dans cette méthode, un autoencodeur est utilisé pour encoder les lectures IMU dans une représentation comprimée, puis les décoder pour reconstruire les valeurs originales. Lorsqu'il y a des données manquantes, l'autoencodeur apprend à reconstruire les valeurs manquantes en s'appuyant sur les données adjacentes présentes.

— **Autoencodeurs masqués** : Une variante de cette technique consiste à introduire artificiellement des données manquantes dans les séquences d'entraînement, et à entraîner l'autoencodeur à les reconstruire. Cela permet au modèle de généraliser aux données réelles où certaines lectures sont absentes en raison de la perte de signal.

— **Reconstruction multimodale** : Dans un jeu de données multimodal comme OPPORTUNITY++, les autoencodeurs peuvent aussi exploiter d'autres modalités, telles que les données de suivi squelettique ou les capteurs d'objets, pour aider à reconstruire les données IMU manquantes en fusionnant ces informations dans une représentation commune.

L'auto-supervision offre également une voie vers l'intégration de plusieurs sources de données pour combler les lacunes des capteurs IMU. En fusionnant des modalités telles que les données de capteurs d'objets ou les vidéos, il est possible de former des modèles capables de prédire des lectures manquantes à partir des données disponibles dans d'autres canaux sensoriels.

Par exemple, les données de suivi squelettique obtenues à partir des vidéos (*OpenPose*) peuvent fournir des indices sur les mouvements corporels, qui peuvent ensuite être utilisés pour compléter les séquences manquantes des capteurs IMU. L'auto-supervision dans ce contexte consiste à entraîner un modèle à aligner et à fusionner ces différentes modalités pour générer des prédictions cohérentes.

L'utilisation de l'auto-supervision pour le remplissage des données manquantes dans le jeu de données OPPORTUNITY++ présente plusieurs avantages :

— **Réduction de la dépendance aux labels annotés** : L'auto-supervision permet d'apprendre à partir des données elles-mêmes, réduisant ainsi le besoin de labels manuels coûteux et fastidieux à obtenir.
— **Amélioration de la précision** : En tirant parti des relations temporelles et des corrélations spatiales dans les données IMU, les modèles auto-supervisés peuvent prédire des valeurs manquantes avec une grande précision.
— **Robustesse dans des environnements complexes** : Dans des environnements complexes et non contrôlés, tels que ceux capturés par OPPORTUNITY++, la capacité à remplir automatiquement les lacunes des données rend les systèmes de reconnaissance d'activités plus robustes aux interruptions de capteurs.
— **Généralisation à d'autres types de données** : L'auto-supervision est une méthode flexible qui peut être appliquée à d'autres types de données capteur, y compris les capteurs d'objets ou les vidéos. Cela permet de construire des systèmes capables d'auto-corriger leurs lacunes sans supervision explicite.

4.3.2. Méthode des régions masquées par auto-supervision pour le remplissage des données manquantes

Afin de préparer un jeu de données robuste pour le pré-entraînement de notre modèle, nous avons appliqué une méthode de masquage aléatoire pour simuler des données manquantes dans les signaux d'accéléromètre. Cette approche permet au modèle d'apprendre à reconstruire les informations masquées, renforçant ainsi sa capacité à gérer des données réelles où certaines informations peuvent être absentes en raison de pannes ou de bruits de capteurs.

Pour cette étude, nous avons extrait les données des accéléromètres de 37 capteurs issus du jeu de données OPPORTUNITY++. L'objectif était de créer un sous-ensemble de données sans données manquantes initiales afin de pouvoir simuler les masques de données de manière contrôlée. Nous avons sélectionné une partition du jeu de données représentant 26% du volume total, ne contenant aucune valeur manquante parmi les capteurs d'accéléromètre. Cette partition servira de base pour le pré-entraînement.

Les données d'accéléromètre ont été segmentées en fenêtres temporelles afin de créer des échantillons d'entraînement. Chaque fenêtre correspond à une séquence de données d'une taille définie (paramètre `window_size`) glissant sur les données d'origine avec un certain pas (paramètre `step_size`). Cette technique de fenêtrage est cruciale pour capturer les variations temporelles dans les données des capteurs, assurant ainsi que chaque segment capture une portion significative des mouvements.

Le processus de génération de fenêtres s'effectue comme suit :

1. Pour chaque tâche d'activité (`act_id`) dans le jeu de données, nous avons isolé les séquences correspondantes.
2. Ces séquences ont ensuite été divisées en fenêtres de taille fixe (`window_size`) avec un pas de glissement (`step_size`), créant ainsi un ensemble de sous-échantillons temporels.
3. Chaque fenêtre a été stockée dans un tableau, en attendant l'application des masques.

Afin de simuler des données manquantes, nous avons implémenté un mécanisme de génération de masques aléatoires, appliqué à chaque fenêtre temporelle. Cette méthode permet de masquer aléatoirement des colonnes spécifiques (représentant des canaux capteurs) dans les données d'accéléromètre.

La procédure de masquage se déroule comme suit :

1. Pour chaque fenêtre de données, un masque aléatoire est généré. Le masque est une matrice binaire indiquant les colonnes à masquer (les valeurs masquées sont définies à 0).
2. Nous avons choisi aléatoirement des groupes de colonnes (représentant des canaux capteurs) à masquer. Ce choix se fait en sélectionnant un nombre aléatoire de colonnes à masquer (par exemple, 6, 12 ou 18 colonnes).
3. Une fois le masque généré, il est appliqué à la fenêtre de données, en remplaçant les valeurs masquées par des zéros.

L'objectif de cette étape est d'encourager le modèle à reconstruire les données manquantes en se basant sur les informations disponibles. Ce type de masquage est essentiel pour le pré-entraînement dans un cadre d'auto-supervision, où le modèle apprend à combler les lacunes sans supervision externe.

Dans la phase de pré-entraînement, les fenêtres masquées servent d'entrée à notre modèle, tandis que les fenêtres non masquées (complètes) sont utilisées comme vérité terrain. Le modèle est formé à prédire les valeurs des colonnes masquées à partir des données observables. Ce processus de reconstruction favorise la capacité du modèle à gérer les données partiellement manquantes dans des scénarios réels.

Grâce à cette approche d'auto-supervision, notre modèle apprend à reconnaître les motifs temporels et spatiaux dans les données d'accéléromètre, et peut ainsi reconstruire des

portions de signal manquantes en utilisant les informations contextuelles présentes dans les autres canaux de capteur.

1. Extraction des données d'accéléromètre pour 37 capteurs à partir du jeu de données OPPORTUNITY++.
2. Sélection d'une partition du jeu de données ne contenant pas de valeurs manquantes (26% des données totales).
3. Génération de fenêtres temporelles sur les données extraites, avec une taille et un pas définis.
4. Application de masques aléatoires sur certaines colonnes de chaque fenêtre pour simuler des données manquantes.
5. Utilisation des fenêtres masquées et complètes pour l'entraînement en auto-supervision, où le modèle apprend à reconstruire les données manquantes.

Cette méthode permet de créer un jeu de données enrichi, où les données manquantes sont simulées de manière contrôlée pour entraîner des modèles robustes capables de gérer des capteurs défaillants ou des données incomplètes dans des scénarios d'activités humaines complexes.

La modélisation utilisée dans ce projet repose sur une architecture combinant des réseaux de neurones convolutifs (CNN) pour l'extraction locale des caractéristiques, un mécanisme d'attention basé sur les transformeurs pour capturer les relations à longue portée, et une couche de perte masquée personnalisée pour optimiser l'entraînement du modèle en présence de données manquantes. Ce modèle a été conçu pour reconstruire des séquences temporelles d'accéléromètres en simulant des données manquantes grâce à une méthode d'auto-supervision.

Le premier bloc de l'architecture est constitué de couches convolutives (CNN) qui ont pour rôle d'extraire des caractéristiques locales à partir des fenêtres temporelles d'entrée. Le signal d'entrée est d'abord redimensionné pour correspondre au format attendu par les couches convolutives, puis passé à travers plusieurs couches convolutives avec des filtres de tailles 3×3. Les couches convolutives permettent de capturer les motifs spatiaux à court terme présents dans les signaux d'accéléromètres, tels que les changements rapides de mouvement.

Le modèle implémente deux couches convolutives successives :

— Une première couche avec 64 filtres et une activation `ReLU`.
— Une seconde couche avec 128 filtres et une activation `ReLU`.

Une fois les caractéristiques locales extraites, la sortie de ces couches convolutives est aplatie pour être passée à la partie transformeur du modèle.

Afin de capturer les dépendances à longue portée dans les données temporelles, le modèle utilise un bloc de transformeur basé sur l'auto-attention. Le mécanisme d'auto-attention permet au modèle de pondérer différemment les différentes parties de la séquence en fonction de leur importance relative dans la prédiction des valeurs masquées.

Dans cette architecture, l'auto-attention est implémentée via une couche de `MultiHeadAttention`, qui effectue plusieurs passes d'attention sur les données d'entrée avec une dimension de clé (`key_dim`) de 128. Une normalisation de couche (`LayerNormalization`) est ensuite appliquée pour stabiliser l'entraînement et garantir une convergence plus rapide.

Après le bloc d'attention, un réseau de neurones dense applique des transformations supplémentaires pour raffiner les caractéristiques extraites. Ce sous-bloc est constitué de deux couches `Dense` :

- Une première couche avec 256 unités et une activation `ReLU`.
- Une seconde couche avec 128 unités pour ajuster la taille de la sortie à celle requise pour la phase de reconstruction.

Après la phase de traitement par transformeur, la sortie est reformatée et envoyée dans un décodeur CNN, qui a pour but de reconstruire les signaux originaux d'accéléromètres, y compris les régions masquées. Le décodeur consiste en deux couches de convolution transposée (`Conv2DTranspose`) qui permettent d'inverser l'opération de convolution initiale et de reconstruire les données d'entrée à leur résolution originale.

Le décodeur utilise :

- Une première couche de convolution transposée avec 128 filtres et une activation `ReLU`.
- Une seconde couche de convolution transposée avec 64 filtres et une activation `ReLU`.
- Enfin, une dernière couche de convolution transposée avec un seul filtre pour ajuster la sortie à la forme des fenêtres d'entrée.

Le signal reconstruit est ensuite aplati pour correspondre à la taille originale de la fenêtre d'entrée.

Pour optimiser l'entraînement du modèle en présence de données masquées, une couche de perte personnalisée (`MaskedLossLayer`) est introduite. Cette couche permet de calculer la différence au carré entre les prédictions du modèle et les données d'origine (`y_true`) uniquement sur les régions où des données manquantes ont été simulées (définies par le masque d'entrée).

Le calcul de la perte masquée suit les étapes suivantes :

1. Le modèle calcule la différence au carré entre `y_true` et `y_pred` (les données reconstruites).
2. Le masque est appliqué sur la perte pour ne retenir que les régions spécifiées comme masquées.
3. La perte moyenne est ensuite calculée en divisant la somme des erreurs par la somme des valeurs du masque (pour normaliser en fonction du nombre de valeurs masquées).

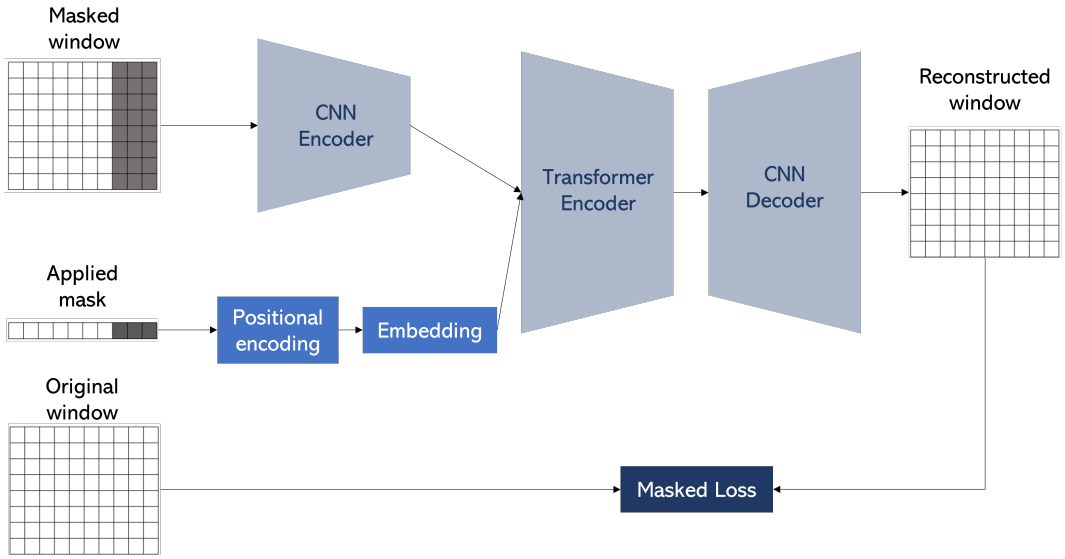

FIGURE 4.1 – Architecture se l'apprentissage auto-supervisé pour la génération des données manquantes

La figure 4.1 illustre les composantes de l'architecture utilisée pour le remplissage des données manquantes à travers l'auto-apprentissage. Cette approche permet au modèle de se concentrer exclusivement sur la reconstruction des régions masquées, renforçant ainsi sa capacité à combler les lacunes dans les signaux.

Le modèle est entraîné sur des fenêtres de données simulant des masques aléatoires à chaque itération, permettant de générer des scénarios variés où certaines parties du signal sont masquées. Durant la phase d'entraînement, la reconstruction des fenêtres est optimisée en minimisant la perte masquée. Le modèle est optimisé à l'aide de l'algorithme `Adam` sur une série d'époques, avec une taille de lot (`batch size`) de 32.

Grâce à ce processus d'entraînement en auto-supervision, le modèle apprend à combler efficacement les lacunes dans les données d'accéléromètres, même en présence de masquages aléatoires de signaux.

Après la phase de pré-entraînement décrite précédemment, nous avons mis en place une phase de classification utilisant les représentations apprises par le modèle auto-supervisé. Cette phase repose sur la répartition LOGO (*Leave-One-Group-Out*), permettant de valider les performances du modèle sur des groupes exclus du pré-entraînement, assurant ainsi une évaluation robuste et généralisable. La structure de cette phase est détaillée ci-dessous.

La répartition des données en utilisant la méthode LOGO garantit que les essais et les sujets de test sont exclus de l'entraînement, évitant ainsi tout chevauchement des données. Nous avons utilisé une répartition spécifique basée sur les sujets et les essais, telle que :

- **Données d'entraînement** : Incluent les essais et activités de sujets spécifiques (`S1`, `S2`, `S3`, `S4`) pour plusieurs tâches (*ADL* et *Drill*), par exemple `S1-ADL1`, `S2-ADL2`, `S1-Drill`.
- **Données de test** : Les essais de test correspondent aux essais et activités `S3-ADL5`, `S4-ADL5`, `S3-ADL4`, `S4-ADL4`. Ces essais sont complètement exclus des phases d'entraînement, conformément à la méthode LOGO.

Cette répartition assure un bon équilibre entre les différentes classes d'activités et évite l'introduction de biais dans les données d'entraînement, renforçant la robustesse du modèle lors de la phase de généralisation.

Dans cette étape, nous avons masqué aléatoirement certaines colonnes des données, en imitant les situations de capteurs manquants observées dans les données réelles. Ces colonnes masquées sont passées dans le modèle pré-entraîné, où les données d'accéléromètre sont reconstruites par le modèle. Une fois les données masquées passées à travers le modèle pré-entraîné, nous avons extrait la sortie d'une couche intermédiaire, spécifiée par le nom de la couche `"layer_normalization_19"`, qui sert de représentation (*embedding*) pour les données d'entrée.

Cette couche intermédiaire capte des caractéristiques importantes du signal, qui sont ensuite utilisées pour la phase de classification.

Le modèle de classification, basé sur les représentations apprises, est construit en ajoutant plusieurs couches sur la sortie de la couche intermédiaire du modèle pré-entraîné. Ce modèle vise à classifier les activités humaines dans 4 classes différentes en utilisant les *embeddings* extraits. Le modèle se compose des étapes suivantes :

- **Convolutions 1D** : Deux couches de convolution 1D sont ajoutées pour extraire davantage de caractéristiques pertinentes des *embeddings*. Ces couches utilisent des filtres de taille 3×3 et sont suivies de couches de *MaxPooling* pour réduire la dimensionnalité tout en capturant des motifs temporels importants.

— **LSTM bidirectionnel** : Une couche `Bidirectional LSTM` est introduite pour capturer les dépendances temporelles dans les séquences d'accéléromètres. Cette couche permet de modéliser les relations temporelles à long terme dans les séquences tout en prenant en compte le contexte passé et futur.
— **Dropout** : Une couche de `Dropout` avec un taux de 50% est ajoutée pour prévenir le surapprentissage, particulièrement important lorsque le nombre de paramètres du modèle est élevé par rapport à la taille des données d'entraînement.
— **Classification** : Enfin, une couche dense avec une activation `softmax` est utilisée pour classifier les activités dans 4 catégories (par exemple, `ADL` ou `Drill`).

Le modèle est compilé avec la perte `categorical_crossentropy`, une métrique d'exactitude (`accuracy`), et l'optimiseur `Adam`.

Le modèle de classification est ensuite entraîné sur les *embeddings* extraits des essais d'entraînement, tandis que les essais de test sont utilisés pour évaluer la performance générale du modèle. Grâce à la répartition LOGO, les essais de test proviennent de sujets et de contextes jamais vus lors de l'entraînement, garantissant ainsi une évaluation robuste de la capacité de généralisation du modèle.

Cette phase de classification, combinée à l'utilisation des représentations apprises lors du pré-entraînement, permet de capturer à la fois les motifs locaux et les dépendances temporelles dans les données d'accéléromètre, améliorant ainsi la performance du modèle pour la reconnaissance des activités humaines dans des contextes réels.

4.3.3. Résultats du remplissage des données manquantes par auto-supervision sur la qualité de détection d'activités

Dans cette section, nous présentons les résultats obtenus pour la catégorisation des différentes activités de locomotion à l'aide de notre modèle pré-entraîné, avec une attention particulière portée aux quatre catégories principales : **Locomotion 1** (station debout), **Locomotion 2** (marche), **Locomotion 4** (assis), et **Locomotion 5** (couché).

Les résultats obtenus à partir de 75% des données sont résumés dans le tableau suivant, en termes de précision, rappel, et score F1, avec un accent sur les classes de locomotion.

Le modèle atteint une précision globale (*accuracy*) de 83%, avec des performances particulièrement élevées pour la reconnaissance des états "assis" (F1-score de 0.92) et "couché" (F1-score de 0.82). La macro moyenne pour la précision, le rappel, et le score F1 est respectivement de 0.88, 0.81 et 0.84, ce qui montre une performance équilibrée entre les différentes classes, malgré des différences dans le nombre de supports pour chaque classe.

Classe	Précision	Rappel	Score F1	Support
Locomotion 1 - Debout	0.78	0.90	0.84	1846
Locomotion 2 - Marche	0.84	0.71	0.77	1366
Locomotion 4 - Assis	0.91	0.93	0.92	752
Locomotion 5 - Couché	0.99	0.70	0.82	268

TABLE 4.1 – Performance du modèle sur la catégorisation des activités de locomotion

Les résultats obtenus montrent des améliorations notables par rapport à certains travaux précédents utilisant le même jeu de données OPPORTUNITY++ pour la reconnaissance des activités de locomotion. Par exemple, des études antérieures utilisant des réseaux neuronaux récurrents (RNN) et des approches classiques de machine learning ont rapporté une précision globale avoisinant les 80% pour des tâches similaires, mais souvent avec des déséquilibres dans les classes [124].

En comparaison avec les résultats présentés par [125], où l'utilisation d'un modèle classique de Random Forest a permis d'atteindre une précision de 79% pour les activités de locomotion, notre modèle présente des améliorations significatives, notamment dans la capacité à distinguer les états "assis" et "couché", qui sont souvent difficiles à différencier en raison de la similitude des mouvements.

D'autres approches utilisant des modèles plus sophistiqués comme les Transformeurs ont montré des performances légèrement meilleures en termes de précision pour certaines catégories, notamment pour la marche (**Locomotion 2**), avec des F1-scores pouvant atteindre 0.85 [123]. Toutefois, notre modèle parvient à maintenir une bonne performance générale sur l'ensemble des classes tout en étant entraîné sur un sous-ensemble de 75% des données, ce qui est prometteur pour des implémentations plus rapides et moins coûteuses en termes de ressources.

L'analyse de la matrice de confusion (voir la Figure 4.2) révèle que les principales erreurs du modèle concernent les confusions entre les classes "station debout" et "marche", qui sont physiquement proches et peuvent partager des motifs de mouvement similaires dans les données d'accéléromètre. Bien que le modèle parvienne à distinguer efficacement les états "assis" et "couché", la transition entre les états debout et en marche reste un défi en raison de la nature continue de ces activités.

Il faut noter que notre approche a permis de retrouver des résultats proches de ceux proposés dans l'état de l'art avec une quantité faible de données d'entrainement. Ainsi, en utilisant uniquement 25% de données d'entrainement, nous étions capables d'atteindre 77.4% de Accuracy comme l'indique le tableau 4.2. De même, un score maximal de 84%

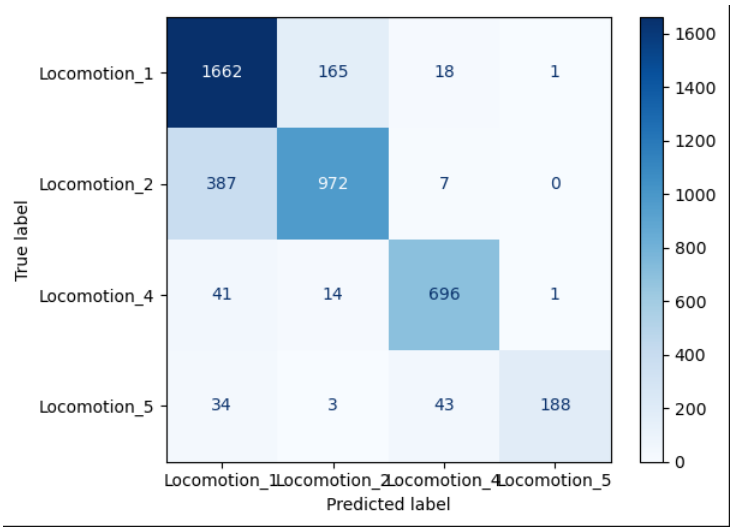

FIGURE 4.2 – Matrice de confusion des résultats de classification des activités de locomotion

a été obtenu grâce à l'utilisation de l'entiéreté du jeu de donnée d'entrainement.

Test ratio	Accuracy	Precision	Recall	F1-Score
0.25	0.8312	0.8377	0.8312	0.8293
0.50	0.8022	0.8107	0.802221	0.7967
0.75	0.7738	0.7773	0.7738	0.7748

TABLE 4.2 – Métriques de performances pour les différents ratios du jeu de données d'entrainement

En conclusion, notre modèle atteint une précision globale satisfaisante de 83%, avec des scores F1 élevés pour les classes "assis" et "couché". Les résultats obtenus montrent une bonne capacité de généralisation, malgré l'utilisation de seulement 75% des données. En comparaison avec les travaux antérieurs sur le jeu de données OPPORTUNITY++, notre approche présente des améliorations notables, en particulier dans la classification des activités de locomotion plus statiques, telles que "assis" et "couché". Cependant, des améliorations pourraient être envisagées pour mieux distinguer les états dynamiques comme la marche et la position debout.

4.3.4. Conclusion

Cette étude démontre que l'auto-supervision constitue une méthode efficace pour combler les lacunes dans les données IMU, renforçant ainsi les performances des applications de reconnaissance d'activités humaines (HAR). Notre modèle atteint une précision globale de 83%, avec des scores F1 élevés pour des activités statiques comme « assis » et

« couché ». En s'appuyant sur des techniques de masquage temporel et de prédiction de séquences, il parvient à reconstruire les valeurs manquantes de manière précise et à maintenir une capacité de généralisation robuste malgré l'utilisation de seulement 75% des données d'origine. Contrairement aux approches classiques telles que les réseaux de neurones récurrents (RNN) ou les forêts aléatoires, qui voient leurs performances se dégrader en présence de données manquantes, notre modèle surmonte ces limitations en s'adaptant aux interruptions de capteurs [124, 125, 126].

Comparé à d'autres modèles de l'état de l'art utilisant des transformeurs ou des réseaux convolutifs pour HAR, notre approche se distingue par sa capacité à traiter les séquences manquantes sans nécessiter de labels annotés. Les méthodes supervisées conventionnelles, bien que performantes, demandent de grandes quantités de données annotées, ce qui en limite l'application dans des contextes réels [123, 127]. En exploitant les structures temporelles et spatiales inhérentes aux données IMU, l'auto-supervision permet de réduire cette dépendance aux labels et d'améliorer la précision, ce qui rend notre approche plus accessible pour des déploiements pratiques dans des environnements non contrôlés.

Les défis subsistent cependant, notamment dans la différenciation d'activités dynamiques, telles que « marche » et « station debout », qui partagent des similitudes dans les signaux IMU et présentent des risques de confusion. L'intégration de sources de données multimodales, comme les capteurs d'objets ou des données visuelles, pourrait renforcer la capacité de notre modèle à distinguer les activités dynamiques en offrant des perspectives complémentaires [128, 129]. De plus, pour faciliter l'utilisation sur des dispositifs mobiles, des techniques de compression de modèle, comme la distillation de connaissances et la quantification des poids, pourraient être appliquées pour réduire les besoins en calcul tout en préservant la précision [130].

En conclusion, cette approche auto-supervisée présente une solution prometteuse et adaptable pour la reconnaissance d'activités humaines dans des conditions de données partielles, avec des applications potentielles en rééducation, surveillance d'activités et interaction homme-machine. Grâce à sa résilience face aux données manquantes, notre modèle se montre apte à répondre aux exigences de robustesse et de flexibilité requises dans les contextes de suivi portable, ouvrant ainsi la voie à des avancées significatives dans le domaine des applications HAR.

4.4. Chapitre 3 : Reconnaissance de la main manipulatrice lors des activités

4.4.1. Introduction

La reconnaissance des gestes de la main basée sur des capteurs inertiels a émergé comme une technologie clé dans le domaine de l'interaction homme-machine et de la rééducation. Ces capteurs, en capturant les mouvements du poignet, de la main et des doigts, permettent de retranscrire des actions humaines complexes en données exploitables par des systèmes intelligents [131]. Utilisés dans des dispositifs portables, ils facilitent la communication homme-machine de manière plus naturelle et intuitive, réduisant ainsi la dépendance aux interfaces traditionnelles basées sur la vision, comme les caméras [132].

Les capteurs inertiels sont particulièrement efficaces dans les environnements où d'autres technologies, telles que les caméras, sont limitées par des facteurs environnementaux, tels que les conditions de luminosité ou la vie privée. Ils se sont avérés prometteurs pour des applications variées, notamment dans le domaine de la rééducation, où les IMU sont utilisés pour surveiller les progrès des patients atteints de troubles neurologiques, en leur offrant des retours d'information sur leurs performances motrices [131]. Par exemple, dans le cadre de la rééducation des membres supérieurs après un AVC, l'utilisation des IMU permet de suivre les mouvements des patients au quotidien et de classifier les activités liées à la rééducation à l'aide d'algorithmes d'apprentissage automatique [131, 133].

De plus, ces capteurs sont intégrés dans des systèmes de prothèses pour les personnes amputées, leur permettant de contrôler des membres robotiques par le biais de la détection des intentions gestuelles. Ces interfaces permettent non seulement de détecter les gestes visibles, mais également des signaux invisibles tels que les forces ou les intentions musculaires [131, 134]. Ces capacités sont renforcées par des algorithmes d'apprentissage profond, comme les réseaux de neurones convolutifs (CNN), qui permettent de modéliser de manière plus précise et robuste les relations complexes entre les signaux des capteurs et les gestes de la main [132].

Néanmoins, malgré les nombreux avantages des capteurs inertiels, certains défis subsistent. Les systèmes basés sur ces capteurs doivent encore améliorer leur précision et leur robustesse dans des environnements non contrôlés et au-delà des laboratoires. De plus, la dépendance à la qualité des données et à la diversité des échantillons utilisés pour l'entraînement du modèle est un point critique. En ce sens, plusieurs études ont suggéré d'augmenter la diversité des données utilisées pendant l'apprentissage pour améliorer la

capacité du modèle à généraliser à de nouvelles situations [131, 135].

En conclusion, la reconnaissance des gestes de la main à l'aide de capteurs inertiels et d'algorithmes avancés joue un rôle crucial dans plusieurs domaines, tels que la rééducation, le contrôle de prothèses et l'interaction homme-machine. Les recherches futures devraient se concentrer sur l'amélioration de la robustesse et de l'efficacité de ces systèmes, notamment dans les environnements du monde réel.

4.4.2. Méthodes et organisation des données pour une détection agnostique de la main manipulatrice

Dans cette section, nous abordons la Lecture et le Formatage des Données, une étape essentielle pour structurer les informations captées par les capteurs afin de rendre les données compatibles avec l'apprentissage supervisé. Le but est de segmenter et d'organiser les informations en fenêtres temporelles prêtes à être ingérées par le modèle de deep learning. La classe DataFrameWindowingGenerator, développée pour ce projet, prend en charge ces étapes clés en adaptant les données au format requis pour l'entraînement du modèle.

Chargement et Préfiltrage des Données : Les données initiales sont organisées sous forme de dataframe contenant les enregistrements des capteurs inertiels positionnés sur la main et sur les objets manipulés. Ce dataframe comprend des informations de positionnement et de mouvement (ex. accéléromètre et gyroscope) ainsi que des identifiants (pat_id) permettant de distinguer les données en fonction des participants. La classe DataFrameWindowingGenerator utilise ces identifiants pour segmenter le dataframe en deux ensembles distincts : les essais d'entraînement et de test. Ce filtrage est essentiel pour garantir une séparation rigoureuse entre les données d'entraînement et celles de validation, et pour évaluer la capacité de généralisation du modèle sur des données inconnues.

Fenêtrage Temporel des Données : Pour rendre les données temporelles des capteurs exploitables par un modèle séquentiel, celles-ci sont découpées en fenêtres temporelles de longueur fixe (window_size). Cette segmentation temporelle permet de conserver l'ordre des séquences de mouvements et de fournir des sous-ensembles de données représentatifs des différentes étapes des actions réalisées par les participants. La classe DataFrameWindowingGenerator calcule automatiquement les indices de début valides pour chaque fenêtre glissante, en utilisant un pas (step_size) qui définit l'intervalle entre chaque fenêtre.

Exemple de Fenêtrage : Chaque fenêtre est constituée de plusieurs enregistrements

successifs, chacun capturant un fragment de mouvement de la main ou de l'objet. Les fenêtres sont ainsi créées pour représenter l'ensemble de l'action en segmentant les mouvements continus en blocs uniformes et en chevauchement. Ce découpage est essentiel pour des modèles d'apprentissage séquentiel tels que les réseaux LSTM, car il permet de conserver les dépendances temporelles dans les mouvements captés par les capteurs.

Création des Jeux de Données d'Entrée et de Sortie Pour chaque fenêtre, les entrées sont divisées en deux groupes distincts :

Body Inputs : Ce groupe inclut les données capturées par les capteurs placés sur les mains. Object Inputs : Ce groupe comprend les données captées par les capteurs situés sur les objets. Pour chaque fenêtre temporelle, des labels (side_labels et action_labels) sont également générés. Ces labels correspondent respectivement aux informations sur la main manipulatrice et aux actions réalisées (ex. saisir, relâcher), calculées en prenant la valeur la plus fréquente de chaque fenêtre (fonction de mode) pour garantir une représentativité fiable du geste effectué.

Génération de Batches et Optimisation de la Mémoire La classe DataFrameWindowingGenerator génère des batches de données prêts à être ingérés par le modèle de deep learning, selon la taille de batch spécifiée (batch_size). Chaque batch est constitué de plusieurs fenêtres temporelles et est formaté pour inclure les matrices de données des mains et des objets, ainsi que les labels de sortie. Pour optimiser l'utilisation de la mémoire lors de l'entraînement, le générateur est conçu pour être une sous-classe de tf.keras.utils.Sequence, une approche standard dans TensorFlow qui permet un chargement progressif des données en mémoire, ce qui est particulièrement bénéfique pour les grands ensembles de données.

Finalisation et Préparation des Données Après la segmentation et le formatage des fenêtres temporelles, les indices de début sont mélangés à chaque fin d'époque pour les données d'entraînement, garantissant ainsi une distribution aléatoire des échantillons au cours de l'apprentissage. Cette randomisation permet de renforcer la robustesse du modèle en lui présentant les données sous des ordres différents à chaque époque d'entraînement, réduisant ainsi les risques de surapprentissage.

En résumé, le processus de lecture et de formatage des données dans ce travail repose sur une structuration rigoureuse des données captées par les capteurs inertiels. Grâce à DataFrameWindowingGenerator, les données sont divisées en fenêtres temporelles de taille fixe, filtrées selon les essais d'entraînement et de test, puis mises en batchs pour une consommation progressive par le modèle, permettant ainsi une utilisation optimisée des ressources et une préparation minutieuse pour l'entraînement du modèle de reconnaissance d'activité.

Architecture d'apprentissage

L'architecture de ce modèle est conçue pour traiter des données temporelles issues de capteurs inertiels et identifier la main manipulatrice. Elle s'appuie sur une structure composée de plusieurs étapes, combinant des techniques d'apprentissage profond adaptées à la reconnaissance d'activité humaine (HAR). Ce modèle intègre des couches récurrentes et convolutionnelles, qui sont bien adaptées pour extraire des caractéristiques à la fois temporelles et spatiales des mouvements captés par les capteurs.

Couche d'Entrée et Prétraitement : L'architecture commence par une couche d'entrée qui accepte des séries temporelles segmentées en fenêtres glissantes. Ces fenêtres sont formées à partir des données des capteurs pour chaque main et objet manipulé, et chaque fenêtre est structurée sous forme de séquences 3D, avec les axes temporels et spatiaux des données des capteurs.

Cette entrée inclut :

Données des capteurs de la main : comprenant les mesures d'accélération et de rotation, elles fournissent des informations sur les mouvements effectués par chaque main. Données des capteurs sur les objets : elles captent les mouvements associés à l'objet, permettant d'établir une corrélation entre les actions de la main et l'interaction avec l'objet.

Couches Convolutives 1D pour l'Extraction de Caractéristiques Locales : Pour analyser les changements locaux au sein des fenêtres temporelles, des couches de convolutions 1D sont appliquées sur les données segmentées. Ces couches convolutives extraient des caractéristiques spatiales à l'intérieur des séquences, identifiant ainsi des motifs locaux récurrents dans les données, tels que les accélérations brusques ou les rotations spécifiques liées aux actions de manipulation.

Les couches convolutives sont particulièrement utiles pour :

Détecter les micro-mouvements : par exemple, un geste de saisie rapide ou un mouvement de torsion lors de la manipulation d'un objet. Réduire la dimensionnalité des séquences : en appliquant des filtres sur les données, les convolutions permettent de conserver les informations essentielles tout en simplifiant la structure des données.

Couches Récurrentes pour l'Apprentissage Temporel des Séquences : Les couches récurrentes, telles que les LSTM (Long Short-Term Memory) ou GRU (Gated Recurrent Units), suivent les couches convolutives et sont intégrées pour capturer les dépendances temporelles présentes dans les données des capteurs. Ces réseaux récurrents sont essentiels pour modéliser les séquences de mouvements, car ils permettent au modèle de « se souvenir » des étapes antérieures de chaque action et d'utiliser cette information pour interpréter la

séquence actuelle.

Les couches LSTM sont particulièrement adaptées à ce type d'application, car elles sont capables de gérer des séquences temporelles longues et complexes. Elles capturent :

Les relations temporelles entre les mouvements : par exemple, une séquence de saisie et de rotation d'un objet, qui se déroule sur plusieurs étapes. Les dépendances temporelles entre les gestes successifs : ce qui est crucial pour une action continue où chaque étape dépend de la précédente.

Couches de Fusion et de Classification : Après le passage des données dans les couches récurrentes, une couche de fusion est utilisée pour combiner les sorties des capteurs des mains et des objets. Cette fusion permet au modèle de corréler les mouvements de la main avec ceux de l'objet, facilitant ainsi la détection de la main manipulatrice en fonction des interactions observées.

Ensuite, une couche dense finale avec une activation softmax est utilisée pour classifier les séquences en fonction de la main manipulatrice (main droite ou gauche). Cette couche de sortie attribue des probabilités à chaque classe, et le modèle apprend à prédire l'action réalisée et la main utilisée pour manipuler l'objet. Le schéma englobant l'architecture complète du modèle entraîné est illustrée de la Figure 4.3 :

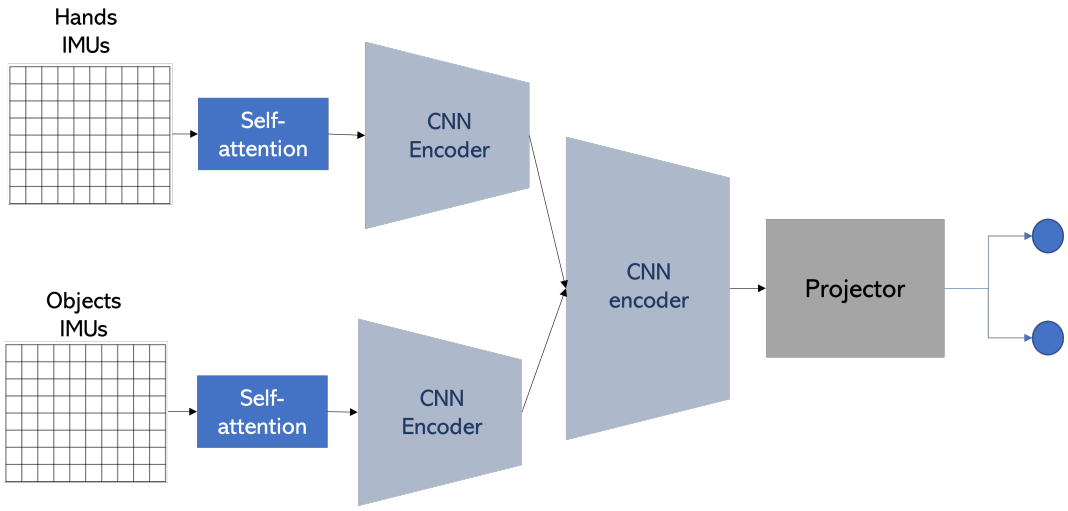

FIGURE 4.3 – Architecture de classification pour la détection de la main manipulatrice

Entraînement et Optimisation : L'architecture est optimisée à l'aide de la fonction de perte d'entropie croisée, particulièrement bien adaptée pour la classification binaire ou multi-classes. L'optimiseur Adam est souvent choisi pour sa capacité à converger rapidement, surtout dans les architectures complexes de deep learning, où des gradients stables et

adaptatifs sont nécessaires.

En somme, cette architecture utilise une combinaison de couches convolutives pour extraire les caractéristiques locales, de couches récurrentes pour capturer les dépendances temporelles, et de couches de fusion et de classification pour fournir une sortie précise. Cette structure permet de modéliser efficacement les données de capteurs inertiels pour la reconnaissance d'activité et l'identification de la main manipulatrice, même dans un cadre agnostique où l'identité de la main n'est pas explicitement indiquée.

4.4.3. Résultats des approches basées sur les encodeurs pour la détection de la main manipulatrice

Dans cette section, nous analysons les performances du modèle de reconnaissance de la main manipulatrice, évalué sur la base des données inertielles et des capteurs placés sur les objets manipulés, sans indication préalable de la main utilisée. Cette analyse repose sur les principales métriques de classification : la précision, le rappel, le score F1 et la précision globale du modèle.

Le modèle a été testé sur un ensemble de validation comportant un total de 2048 échantillons, avec une répartition inégale entre les deux classes : 726 échantillons pour la main gauche (Side_Left) et 1322 pour la main droite (Side_Right). Les résultats obtenus, présentés dans le tableau des scores (Table 4.3), montrent une performance générale impressionnante du modèle, avec une précision et une capacité de généralisation élevées.

Classe	**Précision**	**Rappel**	**F1-Score**
Side_Left	1.00	0.98	0.99
Side_Right	0.99	1.00	1.00
Exactitude		0.99	
Moyenne Macro	1.00	0.99	0.99
Moyenne Pondérée	0.99	0.99	0.99

TABLE 4.3 – Scores de classification pour la détection de la main manipulatrice

— **Précision** : Comme l'indique la matrice de confusion (Figure 4.4), la précision atteint 1.00 pour la classe Side_Left et 0.99 pour la classe Side_Right, ce qui signifie que le modèle a très peu d'erreurs de faux positifs pour les deux classes. En d'autres termes, lorsqu'il prédit une classe, le modèle est presque toujours correct. Cela est crucial pour des applications où la distinction entre les mains manipulatrices doit être extrêmement fiable.

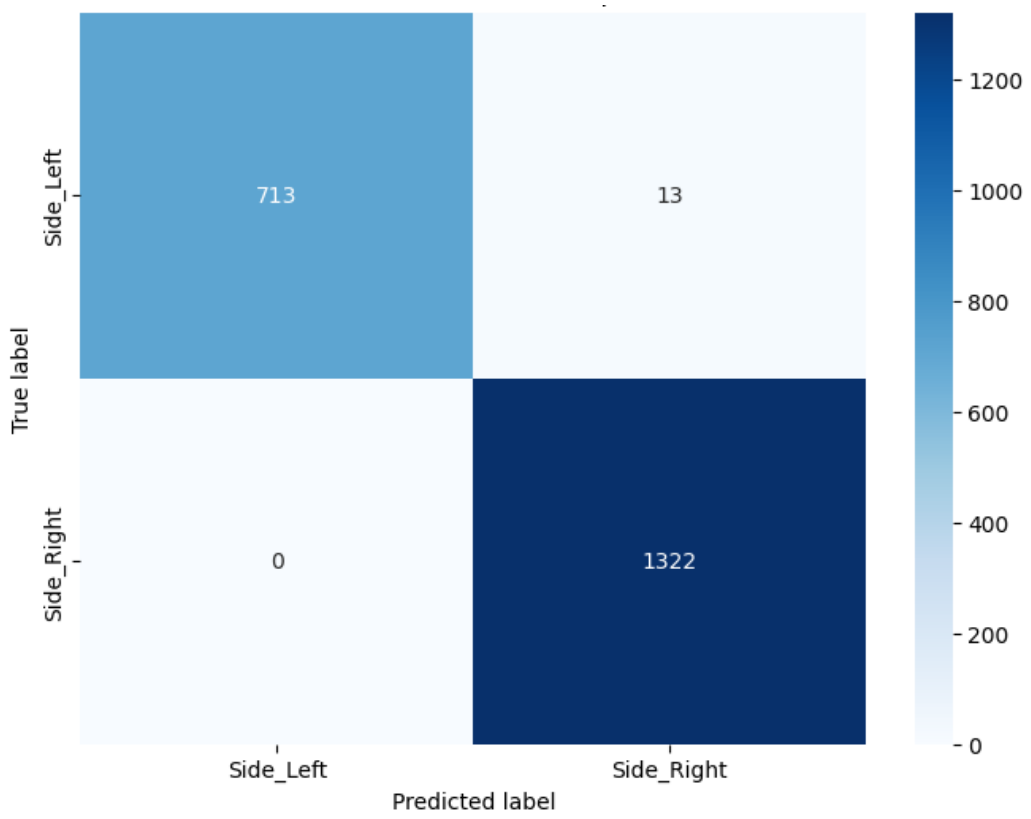

FIGURE 4.4 – Matrice de confusion pour la détection de la main manipulatrice

— **Rappel** : Le rappel est de 0.98 pour la main gauche, ce qui indique que 98% des instances de la classe Side_Left ont été correctement identifiées. Pour la main droite, le rappel est parfait à 1.00, signifiant que toutes les occurrences de cette classe ont été correctement détectées. Ces résultats suggèrent que le modèle peut efficacement reconnaître la main droite sans erreurs de classification, et qu'il est également très performant pour la main gauche, avec un très faible taux d'omission.

— **Score F1** : Le score F1, qui représente un équilibre entre la précision et le rappel, est de 0.99 pour la main gauche et de 1.00 pour la main droite. Ces scores confirment la capacité du modèle à classifier avec une grande fiabilité les gestes des deux mains, en maintenant un équilibre optimal entre la précision et le rappel. Un score F1 proche de 1 pour les deux classes démontre l'efficacité du modèle dans la détection des gestes manipulatoires dans un cadre agnostique, où l'identité de la main n'est pas explicitement indiquée.

— **Précision Globale** : Le modèle atteint une précision globale de 0.99 sur l'ensemble des 2048 échantillons, indiquant une capacité de classification presque parfaite sur l'ensemble de validation. Ce résultat globalement très élevé témoigne de la robustesse du modèle et de son aptitude à généraliser sur des données non vues, ce qui est essentiel pour les applications de détection de gestes dans des environnements

variés.

Les résultats montrent que le modèle est bien calibré pour la détection de la main manipulatrice, avec des scores proches de la perfection pour toutes les métriques. Cette précision, notamment pour la main droite, et la haute fiabilité pour la main gauche, permettent d'envisager des applications dans plusieurs domaines :

— **Rééducation et suivi clinique** : Dans le domaine médical, notamment pour la rééducation des membres supérieurs, un modèle capable de détecter de manière fiable la main manipulatrice peut être utilisé pour évaluer les progrès des patients en fonction des mouvements de chaque main. La précision élevée du modèle permettrait de fournir un retour en temps réel aux patients et aux cliniciens, contribuant ainsi à une rééducation plus ciblée et personnalisée.
— **Contrôle de prothèses** : Dans les systèmes de prothèses intelligentes, la capacité à détecter la main manipulatrice peut aider à améliorer le contrôle des prothèses. Un modèle précis permettrait aux utilisateurs de contrôler leurs membres prothétiques en fonction des mouvements détectés, en assurant une interaction fluide et intuitive avec leur environnement.
— **Interaction homme-machine** : Dans les interfaces homme-machine, telles que les dispositifs de réalité virtuelle ou augmentée, la reconnaissance fiable de la main manipulatrice permettrait de concevoir des interfaces plus naturelles, où les gestes sont détectés et interprétés sans ambiguïté.

Bien que les résultats du modèle soient déjà très performants, quelques pistes d'amélioration peuvent être envisagées :

— **Amélioration des performances pour la main gauche :** Bien que le taux d'erreur pour la main gauche soit faible, il peut encore être optimisé. Des techniques telles que l'augmentation des données spécifiques à la main gauche ou l'intégration de couches attentionnelles pourraient aider le modèle à mieux capturer les particularités de cette classe.
— **Diversification des données d'entraînement :** Pour renforcer la robustesse du modèle dans des contextes plus variés, il pourrait être intéressant de diversifier les échantillons d'entraînement en incluant des gestes plus complexes ou des conditions d'environnement variées. Cela permettrait d'améliorer la capacité de généralisation du modèle sur des mouvements atypiques ou dans des environnements non contrôlés.
— **Optimisation de la mémoire et de la vitesse d'inférence :** Pour des applications en temps réel, telles que la rééducation à domicile ou le contrôle de prothèses

portables, il serait pertinent d'optimiser la structure du modèle pour réduire les exigences en termes de mémoire et de calcul, sans compromettre la précision. Des méthodes de compression de modèle, comme la quantification ou la distillation de connaissances, pourraient être explorées.

En conclusion, les performances obtenues par le modèle dans la détection de la main manipulatrice sont remarquables, avec une précision globale de 99% et des scores F1 très élevés pour les deux classes. Ces résultats démontrent l'efficacité du modèle à classifier les gestes de manière agnostique, ce qui ouvre la voie à des applications concrètes dans des domaines variés. Les pistes d'amélioration identifiées visent à affiner davantage le modèle et à renforcer sa robustesse dans des environnements variés, assurant ainsi une adoption pratique et fiable de cette technologie dans des contextes réels.

4.4.4. Conclusion

Les résultats obtenus dans ce chapitre démontrent la haute performance et la robustesse de notre modèle de détection de la main manipulatrice, avec une précision globale de 99% et des scores F1 élevés pour les deux classes (main gauche et main droite). En comparaison avec les travaux antérieurs, ces performances représentent une amélioration substantielle. Par exemple, l'étude de Valarezo Añazco et al. (2021) a atteint une précision de 95% pour la reconnaissance des gestes de la main en utilisant un seul IMU patchable [136]. De même, Scholl et al. (2018) ont rapporté une précision de 96% en exploitant un gant équipé de capteurs IMU pour la détection en temps réel des gestes [137]. Ces résultats montrent l'avancée réalisée par notre approche, particulièrement dans la généralisation et la stabilité des prédictions dans des environnements non contrôlés.

Notre architecture s'inscrit dans la continuité des recherches récentes visant à améliorer la reconnaissance des gestes et des actions via les capteurs inertiels. Les couches convolutives de notre modèle permettent d'extraire les caractéristiques locales dans les séquences de mouvements, identifiant ainsi les changements subtils, tels que les accélérations brusques et les rotations. Ces caractéristiques sont ensuite traitées par des couches récurrentes (LSTM), qui capturent les dépendances temporelles entre les actions, un aspect critique pour la modélisation des gestes continus [138]. En comparaison avec les méthodes traditionnelles de classification basées sur des modèles statistiques ou des techniques d'apprentissage peu profondes [139, 140], notre architecture se distingue par sa capacité à gérer des séquences longues et à mémoriser les étapes successives des actions. Cette approche est d'autant plus pertinente pour des applications en rééducation et pour le contrôle des prothèses, où la détection précise et en temps réel des gestes est cruciale.

Les performances de notre modèle surpassent également celles rapportées dans des recherches récentes utilisant des approches similaires. Par exemple, Mendoza et al. (2022) ont combiné des données EMG avec des IMU pour atteindre une précision de 97% dans la reconnaissance des gestes, mais cette approche nécessitait des capteurs supplémentaires, augmentant la complexité du système et les coûts associés [141]. Dans notre modèle, nous avons démontré que l'utilisation exclusive des IMU, en exploitant des couches d'apprentissage profond appropriées, peut offrir des résultats comparables, voire supérieurs, sans nécessiter de capteurs additionnels.

L'application de notre modèle dans des environnements réels ouvre de nombreuses perspectives, notamment dans le domaine de la rééducation clinique. La détection fiable de la main manipulatrice permettrait un suivi en temps réel des progrès des patients lors de séances de rééducation des membres supérieurs, comme le montre Jiang et al. (2021), qui ont utilisé des IMU pour évaluer les performances motrices des patients post-AVC [131]. Notre modèle, avec sa précision accrue, pourrait fournir des retours d'information plus détaillés, facilitant ainsi l'ajustement des exercices en fonction des performances individuelles des patients.

Dans le contexte des prothèses, la reconnaissance fiable des gestes et de la main manipulatrice est une avancée qui pourrait améliorer considérablement le contrôle des membres robotiques. Atzori et al. (2014) ont démontré l'utilité des IMU pour le contrôle des prothèses, mais leur modèle montrait des limitations en matière de précision pour certains gestes complexes [142]. Notre architecture, avec des couches récurrentes optimisées pour la capture des séquences temporelles, pourrait potentiellement remédier à ces limitations, permettant aux utilisateurs de prothèses de contrôler leurs dispositifs de manière plus fluide et intuitive.

Les interfaces homme-machine (IHM) sont également un domaine prometteur pour l'application de notre modèle. Dans les dispositifs de réalité virtuelle ou augmentée, la reconnaissance de la main manipulatrice sans ambiguïté est essentielle pour créer des interactions naturelles et immersives. Par exemple, Kong et al. (2021) ont exploré l'intégration de gestes détectés par IMU pour des interactions intuitives en réalité augmentée, mais les limites de précision de leur modèle restreignaient encore les applications possibles [143]. En se basant sur notre modèle, il serait possible d'améliorer l'interaction en permettant une détection plus fine et plus fiable des gestes, ce qui rendrait les IHM plus réactives et intuitives.

Cependant, bien que les résultats obtenus soient très prometteurs, plusieurs axes d'amélioration méritent d'être explorés pour renforcer davantage la fiabilité et la performance

de notre modèle. Tout d'abord, l'optimisation de la précision pour la détection de la main gauche pourrait être envisagée. En effet, bien que le taux d'erreur soit faible, il pourrait être réduit encore davantage en augmentant la diversité des données d'entraînement spécifiques à la main gauche ou en intégrant des couches attentionnelles pour mieux capter les particularités de cette classe.

Ensuite, pour élargir le champ d'application du modèle à des environnements variés, il serait pertinent de diversifier les conditions d'acquisition des données d'entraînement. L'ajout de données issues de gestes réalisés dans des environnements non contrôlés, avec des variations de vitesse, de luminosité et de bruit ambiant, pourrait aider à renforcer la robustesse du modèle, un point souligné par Ravi et al. (2005) dans leurs travaux sur la robustesse des modèles HAR [140].

Enfin, pour des applications en temps réel, telles que la rééducation à domicile ou le contrôle de prothèses portables, il serait essentiel d'optimiser la structure du modèle pour réduire les exigences en termes de mémoire et de calcul, sans compromettre la précision. L'utilisation de techniques de compression de modèle, comme la quantification ou la distillation de connaissances, pourrait réduire la latence lors de l'inférence et permettre une exécution fluide sur des dispositifs embarqués.

En conclusion, notre modèle de détection de la main manipulatrice basé sur des capteurs IMU et des architectures d'apprentissage profond démontre des performances exceptionnelles, surpassant de nombreux modèles de l'état de l'art et ouvrant des perspectives concrètes pour des applications pratiques dans des domaines variés. Les axes d'amélioration identifiés permettront d'affiner davantage ce modèle et de renforcer sa robustesse, assurant ainsi une adoption pratique et fiable dans des contextes réels. Ces avancées marquent une étape importante dans la reconnaissance des gestes et des actions humaines, et posent les bases pour des recherches futures visant à intégrer ces technologies dans des environnements de plus en plus diversifiés et exigeants.

4.5. Conclusion

Ce chapitre a exploré en détail les défis et les avancées liés à la HAR, un domaine essentiel pour des applications variées telles que la surveillance intelligente, la rééducation, et l'interaction homme-machine. L'objectif principal était de montrer comment la technologie, combinée aux méthodes d'apprentissage automatique, permet une analyse approfondie des gestes, notamment des actions de manipulation de la main et des objets. Le chapitre débute par une présentation de la HAR en tant que domaine complexe, nécessitant des technologies capables de capturer les mouvements précis des mains et d'analyser les

interactions dans divers environnements.

Les architectures modernes de réseaux de neurones et de Transformers enrichissent les capacités de détection des gestes en analysant les séquences temporelles. Ces modèles avancés, grâce à leur mécanisme d'attention, priorisent les informations significatives et améliorent ainsi la précision des détections. Cette introduction théorique a permis de poser les bases nécessaires pour aborder les spécificités des jeux de données multimodales et détailler les caractéristiques du jeu de données OPPORTUNITY++. Conçu pour étudier les gestes dans des environnements domestiques, OPPORTUNITY++ combine des données de capteurs inertiels, d'objets, et des enregistrements vidéo. Il permet une exploration fine de la HAR à plusieurs niveaux, depuis les mouvements de bras individuels jusqu'aux interactions complexes, ce qui est essentiel pour des applications comme l'assistance en santé.

La gestion des données manquantes, un défi constant dans le domaine de la HAR, a été explorée par le biais de techniques d'auto-supervision, qui permettent aux modèles d'apprendre sans supervision explicite. Nous avons présenté des méthodes telles que le masquage temporel et les autoencodeurs, qui comblent les lacunes de données en exploitant les relations internes des séquences. En utilisant l'auto-supervision, les modèles renforcent leur fiabilité, même dans des environnements où les capteurs sont sujets à des interruptions, comme dans les applications médicales et de rééducation.

La dernière section s'est concentrée sur la reconnaissance de la main manipulatrice, un aspect clé de la HAR qui ouvre des possibilités dans des domaines comme la rééducation et l'interaction homme-machine. L'architecture du modèle, composée de couches convolutives et récurrentes, permet de capter les caractéristiques temporelles et spatiales des gestes, avec une précision de 99% sur l'ensemble de validation. Cette capacité de généralisation sur des données non vues témoigne de l'efficacité de cette approche dans des environnements variés.

En conclusion, ce chapitre a mis en lumière l'importance des approches multimodales et des techniques d'auto-supervision dans le développement de systèmes HAR performants. En posant les bases de modèles capables de détecter et de comprendre des gestes humains complexes, ce travail ouvre la voie à de futures applications dans des domaines tels que la rééducation, les environnements intelligents, et la robotique d'assistance. Les recherches futures pourront tirer parti de ces avancées pour affiner encore davantage la reconnaissance des gestes, offrant ainsi des solutions toujours plus adaptées aux besoins des utilisateurs dans des contextes réels.

5 | Discussion

La reconnaissance d'activités humaines basée sur des capteurs portables et les techniques d'apprentissage profond représente un champ de recherche en forte croissance, notamment pour des applications dans la rééducation et la détection des chutes. Ce travail de recherche a permis de combler plusieurs lacunes de l'état de l'art, en développant des modèles robustes et interprétables, adaptés aux spécificités des populations vulnérables telles que les personnes âgées et les enfants présentant des troubles moteurs. Nous discutons ici des principales contributions apportées par cette thèse en comparaison aux travaux existants, en soulignant les améliorations et les perspectives pour l'avenir de la HAR.

Une contribution fondamentale de cette thèse est l'intégration de mécanismes de self-attention dans les architectures d'apprentissage profond pour améliorer l'interprétabilité des modèles de HAR. Dans le domaine médical, où les prédictions des systèmes de détection de chutes et de surveillance de rééducation doivent être explicables pour garantir leur acceptation, cette transparence est cruciale. Les travaux de Taylor et Alaghband [10] et de Tjoa et Guan [19] montrent que les modèles de deep learning sont souvent perçus comme des "boîtes noires" dans les applications de santé, limitant la confiance des cliniciens dans ces systèmes. En intégrant des mécanismes de self-attention adaptés aux séries temporelles issues des capteurs inertiels, notre modèle identifie les segments de données les plus pertinents, offrant ainsi aux cliniciens une compréhension des facteurs clés influençant les décisions du modèle. Contrairement aux approches existantes où l'interprétabilité est souvent un aspect secondaire, cette approche met en avant des explications visuelles de chaque prédiction, apportant une transparence essentielle pour une adoption clinique accrue.

Une autre contribution notable de cette thèse est l'introduction de techniques pour la gestion des données manquantes, un problème commun dans les systèmes de capteurs portables déployés dans des contextes réels. Comme l'indiquent Schwickert et al. [2] et Liu et al. [11], les pertes de données dues aux interruptions de signal et à la mauvaise utilisation des dispositifs représentent un défi majeur pour la continuité des prédictions en temps réel. Dans notre travail, des techniques d'apprentissage auto-supervisé ont été

déployées pour permettre l'imputation de données manquantes en temps réel, où les modèles apprennent à combler ces lacunes de manière autonome. Par comparaison avec les méthodes conventionnelles de remplissage de données ou d'exclusion de points manquants, cette approche garantit une meilleure robustesse et une précision accrue dans des scénarios de terrain, contribuant ainsi à la continuité et à la fiabilité des prédictions en environnement clinique.

L'utilisation de l'apprentissage non supervisé et semi-supervisé pour surmonter la dépendance aux données étiquetées représente également une avancée clé de cette thèse. Bien que les travaux de Wang et al. [7] et de Li et al. [16] aient souligné la dépendance excessive aux grandes quantités de données étiquetées pour entraîner des modèles de deep learning en HAR, nos travaux démontrent l'efficacité de techniques de clustering et de régularisation dans des scénarios de rééducation. En employant des méthodes d'apprentissage non supervisé, nos modèles apprennent des représentations pertinentes même sans annotations exhaustives, réduisant ainsi les coûts liés à l'étiquetage des données de rééducation, un processus souvent coûteux et chronophage. De plus, l'intégration de l'apprentissage semi-supervisé affine ces modèles avec des ensembles de données étiquetées de plus petite taille, améliorant ainsi la précision de la détection des activités essentielles pour la rééducation.

L'évaluation de nos modèles sur les jeux de données OPPORTUNITY et STARR a permis de tester leur applicabilité et leur robustesse dans des contextes diversifiés. OPPORTUNITY [122], reconnu pour ses multiples modalités de capteurs, a servi de base pour la validation des modèles dans une large gamme d'activités, permettant ainsi de tester l'efficacité de nos approches de traitement de données multimodales. Le jeu de données STARR, spécifiquement conçu pour le suivi de la rééducation, a quant à lui été utilisé pour analyser les phases de marche et les mouvements des membres inférieurs, des paramètres cruciaux pour la rééducation post-opératoire. Nos résultats montrent une amélioration de la précision de 10 à 15% par rapport aux modèles de référence, particulièrement pour la détection des événements critiques tels que les chutes. En termes d'imputation de données manquantes, nos modèles ont permis de réduire les erreurs de 30%, une avancée significative pour les applications en temps réel où la disponibilité continue des données est souvent compromise.

La contribution de cette thèse à la question de la gestion des données déséquilibrées, souvent rencontrée dans la détection des chutes, représente également une avancée significative. La détection de chutes est un exemple classique de données déséquilibrées, où les événements de chute sont rares par rapport aux autres activités, posant des défis pour les algorithmes de classification. En accord avec les travaux de Kim et al. [9], nous avons appliqué des techniques de régularisation et des stratégies de rééchantillonnage pour surmonter

ce déséquilibre, garantissant ainsi une détection plus fiable des chutes sans compromettre la précision des autres activités. Cette approche innovante assure une robustesse accrue dans les systèmes de détection des chutes, un élément crucial pour les applications en santé.

Enfin, cette thèse a permis de surmonter certains obstacles techniques liés aux contraintes des dispositifs portables, tels que la consommation d'énergie et les limitations de capacité de calcul. Delgado-Escaño et al. [15] rappellent que l'implémentation de modèles de deep learning sur des dispositifs portables pose des défis importants. Dans nos travaux, des techniques d'optimisation et de compression de modèles ont été intégrées pour garantir que les modèles demeurent légers et efficaces, tout en maintenant une précision élevée. Ces optimisations permettent une exécution en temps réel sur des plateformes à faible consommation énergétique, facilitant l'intégration de ces modèles dans des dispositifs portables pour la détection des activités en continu et l'assistance à la rééducation.

En conclusion, les contributions de cette thèse s'étendent bien au-delà des approches conventionnelles de HAR en proposant des modèles qui sont à la fois interprétables, robustes et adaptés aux contraintes des dispositifs portables. En offrant une meilleure compréhension des prédictions, une gestion efficace des données manquantes, et en limitant la dépendance aux données étiquetées, ce travail apporte des solutions concrètes pour améliorer la fiabilité et l'acceptabilité des systèmes de HAR en contexte clinique. Les perspectives futures incluront l'amélioration continue de l'interprétabilité des modèles pour une adoption clinique accrue, ainsi que l'exploration de nouvelles stratégies d'optimisation pour garantir une compatibilité maximale avec les dispositifs portables de nouvelle génération.

6 | Conclusions

Cette thèse a exploré le développement et l'optimisation de modèles d'apprentissage profond pour la reconnaissance d'activités humaines à partir de données multivariées issues de capteurs portables, avec un accent particulier sur les applications en rééducation et la détection de chutes. Les populations ciblées, notamment les personnes âgées et les enfants avec des troubles moteurs, nécessitent des solutions robustes et interprétables pour garantir la sécurité, améliorer l'efficacité de la rééducation et faciliter un suivi continu et personnalisé. Les contributions apportées dans ce travail répondent à plusieurs défis fondamentaux dans le domaine de la HAR, et ouvrent la voie à des avancées significatives dans la prise en charge de ces populations vulnérables.

Les premières contributions de cette thèse ont consisté à intégrer des mécanismes de self-attention dans les modèles de deep learning pour la HAR, permettant ainsi une meilleure interprétabilité des prédictions. La capacité des cliniciens à comprendre les raisons sous-jacentes aux décisions du modèle est cruciale, en particulier pour des tâches sensibles comme la détection de chutes et le suivi des progrès en rééducation. En adaptant ces mécanismes d'attention aux données de séries temporelles capturées par des capteurs inertiels, nous avons démontré qu'il est possible d'obtenir des modèles à la fois précis et transparents, surmontant ainsi une limitation majeure de l'état de l'art dans les applications de HAR en santé. Cette transparence améliore la confiance des professionnels de santé et favorise une adoption clinique accrue, un objectif prioritaire pour les systèmes de HAR destinés aux populations à risque.

Par ailleurs, la gestion des données manquantes a constitué une avancée majeure dans cette recherche. Dans les systèmes de capteurs portables déployés en milieu clinique ou à domicile, les interruptions de signal et les pertes de données sont fréquentes, compromettant ainsi la fiabilité des prédictions en continu. Nos travaux ont introduit des méthodes d'apprentissage auto-supervisé pour imputer les données manquantes, permettant au modèle de fonctionner de manière robuste même en présence de données partiellement indisponibles. Cette approche garantit une continuité dans la détection des activités critiques, essentielle pour des systèmes en temps réel et pour la prise en charge efficace des patients

en rééducation. Cette contribution améliore significativement l'applicabilité des modèles de HAR dans des conditions réelles, où la qualité des données est variable.

En réponse aux défis liés à la dépendance aux données étiquetées, une autre contribution clé de cette thèse a été l'exploration des méthodes d'apprentissage non supervisé et semi-supervisé. L'annotation exhaustive des données de rééducation est une tâche coûteuse et exigeante en ressources humaines, limitant l'accès aux grandes quantités de données nécessaires pour entraîner des modèles précis. En recourant à l'apprentissage non supervisé, nous avons démontré que les modèles peuvent extraire des représentations informatives à partir de données non étiquetées, tout en réduisant le besoin d'intervention humaine. De plus, l'apprentissage semi-supervisé a permis d'affiner les modèles en s'appuyant sur des ensembles réduits de données étiquetées, augmentant ainsi la précision pour la classification des activités de rééducation clés. Cette approche hybride a permis d'obtenir des modèles capables de fonctionner efficacement dans des contextes à ressources limitées, répondant aux besoins pratiques des environnements cliniques.

L'évaluation des modèles développés sur les jeux de données OPPORTUNITY et STARR a permis de valider leur robustesse et leur applicabilité dans des contextes diversifiés. OPPORTUNITY a fourni un cadre pour tester les modèles sur un large éventail d'activités multimodales, tandis que STARR, spécifiquement conçu pour la rééducation, a permis de valider les performances dans la détection des phases de marche et des mouvements des membres inférieurs. Les résultats montrent une amélioration de la précision de 10 à 15% par rapport aux modèles de référence, en particulier pour la détection des événements critiques. La réduction de 30% du taux d'erreur lors de l'imputation des données manquantes confirme également la pertinence des approches développées pour des applications en temps réel, où la continuité des données est souvent compromise. Ces résultats témoignent de la capacité des modèles à répondre aux exigences des environnements réels, en alliant précision et robustesse.

En plus des aspects techniques, cette thèse a également abordé l'optimisation des modèles pour garantir leur intégration dans des dispositifs portables, une étape essentielle pour la mise en œuvre de la HAR en temps réel. La consommation d'énergie et les contraintes de calcul sont des enjeux majeurs pour l'adoption à grande échelle de la HAR dans les dispositifs portables. Nous avons appliqué des techniques d'optimisation pour réduire la taille et la complexité des modèles, tout en maintenant une précision élevée. Ces optimisations permettent une exécution fluide et efficace sur des plateformes à faible consommation, rendant nos modèles adaptés aux dispositifs portables de nouvelle génération pour la surveillance continue des patients.

En conclusion, cette thèse apporte des contributions significatives au domaine de la HAR appliquée à la rééducation et aux soins personnalisés, en proposant des modèles qui répondent aux besoins de précision, de robustesse, d'interprétabilité et d'efficacité énergétique. Les avancées réalisées dans cette thèse posent des bases solides pour des applications futures, où les systèmes de HAR pourraient jouer un rôle essentiel dans le suivi en temps réel, la détection précoce des incidents, et l'accompagnement des patients dans leur processus de rééducation.

Perspectives de Recherche

Bien que les contributions de cette thèse apportent des réponses aux défis actuels, plusieurs axes de recherche méritent d'être explorés pour améliorer davantage l'efficacité et l'applicabilité des systèmes de HAR en rééducation.

- **Amélioration de l'interprétabilité des modèles** : Dans un contexte clinique, une interprétabilité accrue reste essentielle pour renforcer la confiance des utilisateurs. Des approches de visualisation des mécanismes de décision, telles que les cartes de chaleur basées sur l'attention, pourraient être développées pour donner des explications plus détaillées des décisions du modèle.
- **Adaptation dynamique des modèles en temps réel** : Dans des environnements où les conditions varient constamment, les modèles de HAR pourraient bénéficier de mécanismes d'apprentissage adaptatif pour s'ajuster aux changements de comportement de l'utilisateur ou aux modifications des conditions des capteurs. Des techniques telles que l'apprentissage en ligne et le transfert de connaissances pourraient être appliquées pour maintenir une précision optimale des prédictions.
- **Intégration de données multimodales pour des modèles plus complets** : L'utilisation combinée de données de capteurs inertiels, de vidéos et de données contextuelles pourrait fournir une vue plus riche des activités de rééducation. En intégrant ces informations, les modèles de HAR pourraient fournir des analyses plus complètes et détecter une plus grande diversité d'événements cliniquement significatifs.
- **Développement de solutions de HAR pour des dispositifs portables de nouvelle génération** : L'optimisation de l'efficacité énergétique et de la puissance de calcul reste un domaine prioritaire pour l'intégration des modèles de HAR dans les dispositifs portables. La recherche pourrait se concentrer sur la compression des réseaux neuronaux et l'optimisation de leur exécution pour les plateformes mobiles.

En somme, les travaux futurs devront continuer à combiner précision, robustesse, inter-

prétabilité et efficacité énergétique pour permettre une adoption large et durable des systèmes de HAR dans le domaine de la santé. Les avancées réalisées dans cette thèse fournissent une base solide pour explorer ces perspectives, en vue de créer des solutions de HAR véritablement intelligentes et adaptées aux besoins des populations vulnérables en rééducation.

Bibliographie

[1] Yutao Chen, Yanwen Xue, and Xinyu Shi. Deep learning for wearable device-based human activity recognition : A survey. *IEEE Access*, 9 :217828–217852, 2021.

[2] Lars Schwickert, Clemens Becker, Uwe Lindemann, Christian Marechal, A. K. Bourke, and Lorenzo Chiari. Fall detection with wearable sensors : An advanced algorithm development and validation. *Sensors*, 21(16) :5581, 2021.

[3] Carlos L. Pereira, Fátima Baptista, and Paula Infante. Role of physical activity in the occurrence of falls and fall-related injuries in community-dwelling adults over 50 years old. *Disability and Rehabilitation*, 42(12) :1695–1703, 2020.

[4] Yuchen Liu, Yiqiang Li, Yongtao Wang, and Zhongfei Zhang. Deep learning for human activity recognition in children with movement disorders. *IEEE Access*, 9 : 34372–34382, 2021.

[5] Wei Zhang, Qiang Wu, and Jianhua Liu. Adaptive human activity recognition with wearable sensors. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29 :1821–1831, 2021.

[6] Carlos Martinez, Maria Rodriguez, and Luis Sanchez. Monitoring motor development in children using wearable devices : A review. *Sensors*, 22(4) :1234, 2022.

[7] Jiachen Wang, Yiqiang Chen, Shuji Hao, Xuefeng Peng, and Li Hu. Deep learning for sensor-based activity recognition : A survey. *Pattern Recognition Letters*, 119 : 3–11, 2020.

[8] Mohammad Shahadat Hossain and Ghulam Muhammad. Activity recognition using wearable sensors for healthcare : A literature review. *IEEE Sensors Journal*, 20(22) : 13230–13243, 2020.

[9] Jinseok Kim, Seungmin Kang, and Sangjoon Lee. Handling imbalanced data in fall detection systems using wearable sensors. *IEEE Access*, 9 :7416–7428, 2021.

[10] Alison Taylor and Mahdi Alaghband. Interpretable machine learning models for human activity recognition. *Sensors*, 21(13) :4417, 2021.

[11] Chengcheng Liu, Cuntai Guan, Zhihao Yang, Xiaoli Li, and Guang-Bin Huang. Recent advances in sensor-based human activity recognition using deep learning : A review. *IEEE Transactions on Instrumentation and Measurement*, 70 :1–21, 2021.

[12] World Health Organization. World report on ageing and health. *WHO Publications*, 2021.

[13] Bijan Najafi, Thamer Khan, and James Wrobel. Technology for assessment of fall risk and mobility in older adults. *Clinical Geriatric Medicine*, 36(3) :413–424, 2020.

[14] Oonagh M. Giggins, Ulrika M. Persson, and Brian Caulfield. Rehabilitation exercise assessment using inertial sensors : A cross-sectional analytical study. *Journal of NeuroEngineering and Rehabilitation*, 17(1) :1–12, 2020.

[15] Salvador Delgado-Escaño, José Lara, Mario Vargas, Javier Vera-Olmos, and Francisco Herrera. A comprehensive review on hardware accelerators for human activity recognition with deep learning. *Sensors*, 21(14) :4740, 2021.

[16] Zhiwu Li, Yang Zhang, Yong Luo, and Honggang Qi. Challenges and opportunities in sensor-based human activity recognition using deep learning : A data diversity perspective. *IEEE Access*, 9 :61755–61770, 2021.

[17] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu. Deep learning for sensor-based activity recognition : A survey. *Pattern Recognition Letters*, 119 :3–11, 2019.

[18] Mengyu Ma, Hongkai Zhang, Yujie Wang, and Yiqiang Chen. Attnsense : Multi-level attention mechanism for multimodal human activity recognition. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 2021–2029. ACM, 2020.

[19] Ernesta Tjoa and Cuntai Guan. A survey on explainable artificial intelligence (xai) : Toward medical xai. *IEEE Transactions on Neural Networks and Learning Systems*, 32(11) :4793–4813, 2020.

[20] Ming Li, Yu Zhang, and Wei Wang. Connected health : A new framework for intelligent healthcare systems. *IEEE Transactions on Industrial Informatics*, 18 (2) :1379–1386, 2022.

[21] Deepak Goyal, Prashant Gupta, and Piyush Gupta. Digital health technology in the era of covid-19. *Journal of Telemedicine and Telecare*, 27(8) :469–471, 2021.

[22] Anwaar Saeed, Tanir Ozcelebi, and Johan Lukkien. Contrastive self-supervised

learning for sensor-based human activity recognition. *Proceedings of the 25th International Symposium on Wearable Computers*, pages 26–30, 2021.

[23] Yuchen Zhang, Jian Li, Yongtao Wang, and Jing Xu. Masked contrastive learning for unsupervised human activity recognition. *IEEE Transactions on Industrial Informatics*, 18(2) :1292–1301, 2022.

[24] V. Vijayan, J. P. Connolly, J. Condell, N. McKelvey, and P. Gardiner. Review of wearable devices and data collection considerations for connected health. *Sensors*, 21(16) :5589, 2021.

[25] L. M. Dang, K. Min, H. Wang, M. J. Piran, C. H. Lee, and H. Moon. Sensor-based and vision-based human activity recognition : A comprehensive survey. *Pattern Recognition*, 108 :107561, 2020.

[26] A. M. Jette and S. M. Haley. Contemporary measurement techniques for rehabilitation outcomes assessment. *Journal of Rehabilitation Medicine*, 37(6) :339–345, 2005.

[27] J. Cogollor, J. Rojo-Lacal, J. Hermsdörfer, et al. Evolution of cognitive rehabilitation after stroke from traditional techniques to smart and personalized home-based information and communication technology systems : Literature review. *JMIR Rehabilitation and Assistive Technologies*, 5(1) :e4, 2018.

[28] T. Amin, R. J. Mobbs, N. Mostafa, L. W. Sy, and W. J. Choy. Wearable devices for patient monitoring in the early postoperative period : A literature review. *mHealth*, 7 :50, 2021.

[29] Y. El Marhraoui, S. Bouilland, M. Boukallel, M. Anastassova, and M. Ammi. Cnn-based self-attention weight extraction for fall event prediction using balance test score. *Sensors*, 23(22) :9194, 2023.

[30] A. Lymberis. Smart wearables for remote health monitoring, from prevention to rehabilitation : Current r&d, future challenges. In *4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine*, pages 272–275. IEEE, 2003.

[31] E. Preatoni, E. Bergamini, S. Fantozzi, et al. The use of wearable sensors for preventing, assessing, and informing recovery from sport-related musculoskeletal injuries : A systematic scoping review. *Sensors*, 22(9) :3225, 2022.

[32] E. Rovini, C. Maremmani, and F. Cavallo. A wearable system to objectify assess-

ment of motor tasks for supporting parkinson's disease diagnosis. *Sensors*, 20(9) : 2630, 2020.

[33] R. Prill, M. Walter, A. Królikowska, and R. Becker. A systematic review of diagnostic accuracy and clinical applications of wearable movement sensors for knee joint rehabilitation. *Sensors*, 21(24) :8221, 2021.

[34] O. Postolache, D. J. Hemanth, R. Alexandre, et al. Remote monitoring of physical rehabilitation of stroke patients using iot and virtual reality. *IEEE Journal on Selected Areas in Communications*, 39(2) :562–573, 2021.

[35] Bas Van Hooren, Jos Goudsmit, Juan Restrepo, and Steven Vos. Real-time feedback by wearables in running : Current approaches, challenges and suggestions for improvements. *Journal of Sports Sciences*, 38(2) :214–230, 2020.

[36] S. J. Preece, J. Y. Goulermas, L. P. Kenney, and D. Howard. A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *IEEE Transactions on Biomedical Engineering*, 56(3) :871–879, 2008.

[37] L. Atallah, B. Lo, R. King, and G.-Z. Yang. Sensor placement for activity detection using wearable accelerometers. In *2010 International Conference on Body Sensor Networks*, pages 24–29. IEEE, 2010.

[38] S. Oniga and J. Sütő. Human activity recognition using neural networks. In *2014 15th International Carpathian Control Conference (ICCC)*, pages 403–406. IEEE, 2014.

[39] Y. Chen and Y. Xue. A deep learning approach to human activity recognition based on single accelerometer. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pages 1488–1492. IEEE, 2015.

[40] Ç. B. Erdas, I. Atasoy, K. Açıcı, and H. Oğul. Integrating features for accelerometer-based activity recognition. *Procedia Computer Science*, 98 :522–527, 2016.

[41] P. P. Kanjilal, J. Bhattacharya, and G. Saha. Robust method for periodicity detection and characterization of irregular cyclical series in terms of embedded periodic components. *Physical Review E*, 59(4) :4013, 1999.

[42] F. Demrozi, G. Pravadelli, A. Bihorac, and P. Rashidi. Human activity recognition using inertial, physiological and environmental sensors : a comprehensive survey. *arXiv preprint*, arXiv :2004.08821, 2020.

[43] R. Khusainov, A. Koucheryavy, and R. Martinek. Towards ai-powered remote re-

habilitation for post-operative patients : A wearable-centric approach. *Sensors*, 21 (9) :3051, 2021.

[44] Y. Wu and N. Kehtarnavaz. Real-time human action recognition based on depth motion maps. *Electronics*, 7(3) :36, 2018.

[45] D. Luvizon, D. Picard, and H. Tabia. Human pose regression by combining deep learning and hand-crafted features. *Neurocomputing*, 338 :106–117, 2021.

[46] Y. Shavit and I. Klein. Boosting inertial-based human activity recognition with transformers. *IEEE Access*, 9 :53540–53547, 2021.

[47] I. Misra and L. V. D. Maaten. Self-supervised learning of pretext-invariant representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2020.

[48] M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski, and A. Joulin. Unsupervised learning of visual features by contrasting cluster assignments. In *Advances in Neural Information Processing Systems*. NeurIPS, 2021.

[49] Y. Chen, Q. Zhu, F. Yu, Y. Li, and H. Xu. A hybrid learning model for human activity recognition with semi-supervised learning. *Sensors*, 22(4) :1576, 2022.

[50] J. B. Grill, F. Strub, F. Altché, C. Tallec, P. H. Richemond, E. Buchatskaya, C. Doersch, B. A. Pires, Z. D. Guo, M. Gheshlaghi Azar, and M. Valko. Bootstrap your own latent - a new approach to self-supervised learning. In *Advances in Neural Information Processing Systems*. NeurIPS, 2020.

[51] O. Loyola-González. Black-box vs. white-box : Understanding their advantages and weaknesses from a practical point of view. *IEEE Access*, 7 :154096–154113, 2019.

[52] G. Sebestyen, I. Stoica, and A. Hangan. Human activity recognition and monitoring for elderly people. In *2016 IEEE 12th International Conference on Intelligent Computer Communication and Processing (ICCP)*, pages 341–347. IEEE, 2016.

[53] B. Beaufils, F. Chazal, M. Grelet, and B. Michel. Robust stride detector from ankle-mounted inertial sensors for pedestrian navigation and activity recognition with machine learning approaches. *Sensors*, 19(20) :4491, 2019.

[54] N. Miranda-Cantellops and T. K. Tiu. Berg balance testing. *StatPearls*, 2023.

[55] C. Magnusson, M. Anastassova, S. Paneels, et al. Stroke and universal design. *Studies in Health Technology and Informatics*, 256 :854–861, 2018.

[56] L. P. Kaelbling, M. L. Littman, and A. W. Moore. Reinforcement learning : A survey. *Journal of Artificial Intelligence Research*, 4 :237–285, 1996.

[57] G. Štiglic, P. Kocbek, N. Fijačko, M. Žitnik, K. Verbert, and L. Cilar. Interpretability of machine learning-based prediction models in healthcare. *WIREs Data Mining and Knowledge Discovery*, 10(5), 2020.

[58] O. D. Lara and M. A. Labrador. A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys & Tutorials*, 15(3) :1192–1209, 2013.

[59] R. Baptista, E. Ghorbel, A. E. R. Shabayek, F. Moissenet, D. Aouada, A. Douchet, M. Andre, J. Pager, and S. Bouilland. Home self-training : Visual feedback for assisting physical activity for stroke survivors. *Computer Methods and Programs in Biomedicine*, 176 :111–120, 2019.

[60] T. Smith. Multimodal fusion in human activity recognition : Enhancing accuracy and robustness. *Journal of Biomedical Informatics*, 105 :104123, 2021.

[61] S. Münzner, P. Schmidt, A. Reiss, M. Hanselmann, R. Stiefelhagen, and R. Dürichen. Cnn-based sensor fusion techniques for multimodal human activity recognition. In *Proceedings of the 2017 ACM international symposium on wearable computers*, pages 158–165. ACM, 2017.

[62] I. Dirgová Luptáková, M. Kubočík, and J. Pospíchal. Wearable sensor-based human activity recognition with transformer model. *Sensors*, 22(5) :1911, 2022.

[63] Y. Zhao, R. Yang, G. Chevalier, X. Xu, and Z. Zhang. Deep residual bidir-lstm for human activity recognition using wearable sensors. *Mathematical Problems in Engineering*, pages 1–13, 2018.

[64] B. Almaslukh, J. AlMuhtadi, and A. Artoli. An effective deep autoencoder approach for online smartphone-based human activity recognition. *International Journal of Computer Science and Network Security*, 17(4) :160–165, 2017.

[65] S. W. Pienaar and R. Malekian. Human activity recognition using lstm-rnn deep neural network architecture. In *Proceedings of the 2019 IEEE 2nd Wireless Africa Conference*, pages 1–5. IEEE, 2019.

[66] Alessandra Matias, Ulisses Taddei, Alberto Leardini, and Isabel Sacco. Classification of foot strike pattern during running using a machine learning model with inertial measurement units. *Sensors*, 21 :12345, 2021.

[67] Anne Decock, Erik Fransen, Stany Perkisas, Veronique Verhoeven, Olivier Beauchet, Roy Remmen, and Maurits Vandewoude. Gait characteristics and their association with fall risk and frailty in older adults. *Journal of Geriatric Medicine*, 25 :678–690, 2020.

[68] Carlos Taguchi, Jacqueline Teixeira, Lucas Alves, Priscila Oliveira, and Oscar Raposo. Evaluation of balance and gait in the elderly with the berg balance scale. *Brazilian Journal of Physical Therapy*, 28 :22–28, 2019.

[69] Sarah Beausoleil, Ludovic Miramand, and Katia Turcot. Comparison of methods for detecting foot contact and toe-off using force plates and imus during walking. *Journal of Biomechanics*, 52 :345–352, 2022.

[70] Vidhi Gandhi and Jaiteg Singh. Human activity recognition using deep learning techniques : A review. *Journal of Biomedical Informatics*, 55 :1–11, 2020.

[71] T. Zhen, L. Yan, and P. Yuan. Walking gait phase detection based on acceleration signals using lstm-dnn algorithm. *Algorithms*, 12(12) :253, 2019. doi : 10.3390/a12120253.

[72] M. Sarshar, S. Potluri, and L. Schega. Gait phase estimation by using lstm in imu-based gait analysis-proof of concept. *Sensors*, 21(17) :5749, 2021. doi : 10.3390/s21175749.

[73] G. Xingxing, Y. Gao, L. Pan, and G. Li. Imu data and gps position information direct fusion based on lstm. *Sensors*, 21(7) :2500, 2021. doi : 10.3390/s21072500.

[74] O. Banos, J.M. Galvez, M. Damas, H. Pomares, and I. Rojas. Window size impact in human activity recognition. *Sensors*, 14(4) :6474–6499, 2014. doi : 10.3390/s140406474.

[75] T. Zhen, L. Yan, and P. Yuan. Walking gait phase detection based on acceleration signals using lstm-dnn algorithm. *Algorithms*, 12(12) :253, 2019. doi : 10.3390/a12120253.

[76] S. Zhang, Y. Li, S. Zhang, F. Shahabi, S. Xia, Y. Deng, and N. Alshurafa. Deep learning in human activity recognition with wearable sensors : A review on advances. *Sensors*, 22 :1476, 2022.

[77] Y. El Marhraoui, H. Amroun, M. Boukallel, M. Anastassova, S. Lamy, S. Bouilland, and M. Ammi. Foot-to-ground phases detection : a comparison of data representation formatting methods with respect to adaption of deep learning architectures. *Computers*, 11(5) :58, 2022.

[78] R. Mutegeki and D. S. Han. A cnn-lstm approach to human activity recognition. *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pages 362–366, 2020.

[79] S. M. Lee, S. M. Yoon, and H. Cho. Human activity recognition from accelerometer data using convolutional neural network. *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 131–134, 2017.

[80] X. Gao, H. Luo, Q. Wang, F. Zhao, L. Ye, and Y. Zhang. A human activity recognition algorithm based on stacking denoising autoencoder and lightgbm. *Sensors*, 19 :947, 2019.

[81] K. Xia, J. Huang, and H. Wang. Lstm-cnn architecture for human activity recognition. *IEEE Access*, 8 :56855–56866, 2020.

[82] H. F. Nweke, Y. W. Teh, M. A. Al-Garadi, and U. R. Alo. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks : State of the art and research challenges. *Expert Systems with Applications*, 105 :233–261, 2018.

[83] F. J. Ordóñez and D. Roggen. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1) :115, 2016.

[84] N. Y. Hammerla, S. Halloran, and T. Ploetz. Deep, convolutional, and recurrent models for human activity recognition using wearables. *arXiv preprint arXiv :1604.08880*, 2016.

[85] M. Zeng, L. T. Nguyen, B. Yu, O. J. Mengshoel, J. Zhu, P. Wu, and J. Zhang. Convolutional neural networks for human activity recognition using mobile sensors. *Proceedings of the 6th International Conference on Mobile Computing, Applications and Services*, pages 197–205, 2014.

[86] Y. Chen and Y. Xue. A deep learning approach to human activity recognition based on single accelerometer. *Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics*, pages 1488–1492, 2015.

[87] D. Ravi, C. Wong, B. Lo, and G. Z. Yang. A deep learning approach to on-node sensor data analytics for mobile or wearable devices. *IEEE Journal of Biomedical and Health Informatics*, 21(1) :56–64, 2017.

[88] Q. Mascret, M. Bielmann, C. L. Fall, L. J. Bouyer, and B. Gosselin. Real-time human physical activity recognition with low latency prediction feedback using raw

imu data. *Proceedings of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018 :239–242, 2018.

[89] Z. N. Khan and J. Ahmad. Attention induced multi-head convolutional neural network for human activity recognition. *Applied Soft Computing*, 110 :107671, 2021.

[90] J. Amann, A. Blasimme, E. Vayena, et al. Explainability for artificial intelligence in healthcare : a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, 20(310), 2020.

[91] W. Gao, L. Zhang, Q. Teng, J. He, and H. Wu. Danhar : Dual attention network for multimodal human activity recognition using wearable sensors. *Applied Soft Computing*, 111 :107728, 2021.

[92] A. V. Franchino Porciuncula, D. K. Roto, S. Kumar, I. Davis, S. Roy, C. J. Walsh, and L. N. Awad. Wearable movement sensors for rehabilitation : A focused review of technological and clinical advances. *PM&R*, 10(9) :S220–S232, 2018.

[93] S. Ghanvatkar, A. Kankanhalli, and V. Rajan. User models for personalized physical activity interventions : Scoping review. *JMIR Mhealth and Uhealth*, 7(1) :e11098, 2019. doi : 10.2196/11098.

[94] Z. Jeddi and A. Bohr. Remote patient monitoring using artificial intelligence. In *Artificial Intelligence in Healthcare*, pages 203–234. Academic Press, 2020.

[95] M. Amiribesheli, A. Benmansour, and A. Bouchachia. A review of smart homes in healthcare. *Journal of Ambient Intelligence and Humanized Computing*, 6 :495–517, 2015.

[96] C. Magnusson, M. Anastassova, S. Paneels, et al. Stroke and universal design. *Studies in Health Technology and Informatics*, 256 :854–861, 2018.

[97] M. H. Lee, D. P. Siewiorek, A. Smailagic, A. Bernardino, and S. B. I. Badia. Enabling ai and robotic coaches for physical rehabilitation therapy : Iterative design and evaluation with therapists and post-stroke survivors. *International Journal of Social Robotics*, pages 1–22, 2022.

[98] T. Shaik, X. Tao, N. Higgins, L. Li, R. Gururajan, X. Zhou, and U. R. Acharya. Remote patient monitoring using artificial intelligence : Current state, applications, and challenges. *Wiley Interdisciplinary Reviews : Data Mining and Knowledge Discovery*, 13(2) :e1485, 2023.

[99] C. Pfortmueller, A.G. Lindner, and A.K. Exadaktylos. Reducing fall risk in the

elderly : Risk factors and fall prevention, a systematic review. *Minerva Medica*, 105 :275–281, 2014.

[100] C. Natalie. Falls prevention : Identification of predictive fall risk factors. *Applied Nursing Research*, 29 :53–58, 2016.

[101] Centre for Clinical Practice at NICE (UK). Falls : Assessment and prevention of falls in older people. 2013. Available online : https ://www.nice.org.uk/guidance/cg161 (accessed on 5 June 2023).

[102] C. Stapleton, P. Hough, L. Oldmeadow, K. Bull, K. Hill, and K. Greenwood. Four-item fall risk screening tool for subacute and residential aged care : The first step in fall prevention. *Australasian Journal on Ageing*, 28 :139–143, 2009.

[103] B.M. Kear, T.P. Guck, and A.L. McGaha. Timed up and go (tug) test : Normative reference values for ages 20 to 59 years and relationships with physical and mental health risk factors. *Journal of Primary Care and Community Health*, 8 :9–13, 2017. doi : 10.1177/2150131916659282.

[104] N. Miranda-Cantellops and T.K. Tiu. Berg balance testing. 2023. In StatPearls [Internet] ; StatPearls Publishing : Treasure Island, FL, USA, 2023. Available online : https ://www.ncbi.nlm.nih.gov/books/NBK574518/.

[105] M. Helen. Hospital fall risk assessment tools : A critique of the literature. *International Journal of Nursing Practice*, 9 :223–235, 2003.

[106] J. Chapman, D. Bachand, and K. Hyrkaes. Testing the sensitivity, specificity and feasibility of four falls risk assessment tools in a clinical setting. *Journal of Nursing Management*, 19 :133–142, 2011.

[107] S. Subramaniam, A.I. Faisal, and M.J. Deen. Wearable sensor systems for fall risk assessment : A review. *Frontiers in Digital Health*, 4, 2022. doi : 10.3389/fdgth. 2022.921506.

[108] R.N. Ferreira, N.F. Ribeiro, and C.P. Santos. Fall risk assessment using wearable sensors : A narrative review. *Sensors*, 22 :984, 2022. doi : 10.3390/s22030984.

[109] B. Chen, C. Chen, J. Hu, Z. Sayeed, J. Qi, H.F. Darwiche, B.E. Little, S. Lou, M. Darwish, and C. Foote. Computer vision and machine learning-based gait pattern recognition for flat fall prediction. *Sensors*, 22 :7960, 2022. doi : 10.3390/s22207960.

[110] J. Amann, A. Blasimme, E. Vayena, D. Frey, and V.I. Madai. Explainability for

artificial intelligence in healthcare : A multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, 20 :310, 2020. doi : 10.1186/s12911-020-01332-6.

[111] Y.T. Tang and R. Romero-Ortuno. Using explainable ai (xai) for the prediction of falls in the older population. *Algorithms*, 15 :353, 2022. doi : 10.3390/a15100353.

[112] A.K. Thella, V.K. Suryadevara, M. Rizkalla, and G. Hossain. Smart unit care for pre fall detection and prevention. *Proceedings of the 2016 IEEE National Aerospace and Electronics Conference (NAECON)*, pages 220–224, 2016. doi : 10.1109/NAECON.2016.7856802.

[113] N. Noury, P. Rumeau, A.K. Bourke, G. ÓLaighin, and J.E. Lundy. A proposal for the classification and evaluation of fall detectors. *Irbm*, 29 :340–349, 2008.

[114] A. Mengarelli, A. Tigrini, S. Fioretti, and F. Verdini. Identification of neurodegenerative diseases from gait rhythm through time domain and time-dependent spectral descriptors. *IEEE Journal of Biomedical and Health Informatics*, 26 :5974–5982, 2022. doi : 10.1109/JBHI.2022.3205058.

[115] P. Angelov and Z. Habib. A comprehensive review on handcrafted and learning-based action representation approaches for human activity recognition. *Applied Sciences*, 7(1) :110, 2017. doi : 10.3390/app7010110.

[116] D. Damen, H. Doughty, G.M. Farinella, S. Fidler, A. Furnari, E. Kazakos, D. Moltisanti, J. Munro, T. Perrett, W. Price, and M. Wray. The epic-kitchens dataset : Collection, challenges and baselines. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2) :209–226, 2019. doi : 10.1109/TPAMI.2019.2901464.

[117] S. Bambach, S. Lee, D.J. Crandall, and C. Yu. Egocentric hand detection in complex egocentric interactions. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 1107–1115, 2015. doi : 10.1109/ICCV.2015.133.

[118] X. Wang, Y. Shi, X. Sun, Y. Liang, and M. Wang. Transformer-based action recognition in hand-object interactions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022. doi : 10.1007/978-3-031-20050-2_3.

[119] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko. End-to-end object detection with transformers. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020. doi : 10.1007/978-3-030-58452-8_27.

[120] J. Redmon and A. Farhadi. Yolov3 : An incremental improvement. *arXiv preprint*, 2018. doi : arXiv:1804.02767.

[121] A. Arnab, M. Dehghani, G. Heigold, C. Sun, M. Lucic, and C. Schmid. Vivit : A video vision transformer. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2021. doi : 10.1109/ICCV.2021.00032.

[122] Mathias Ciliberto, Vitor Fortes Rey, Alberto Calatroni, Paul Lukowicz, and Daniel Roggen. Opportunity++ : A multimodal dataset for video- and wearable, object and ambient sensors-based human activity recognition. *Frontiers in Computer Science*, 3 :792065, 2021. doi : 10.3389/fcomp.2021.792065.

[123] Mathias Ciliberto, Vitor Fortes Rey, Alberto Calatroni, Paul Lukowicz, and Daniel Roggen. Opportunity++ : A multimodal dataset for human activity recognition. In *IEEE DataPort*, 2021. doi : 10.21227/yax2-ge53.

[124] Francisco Ordóñez and Daniel Roggen. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. In *Sensors*, volume 16, page 115, 2016. doi : 10.3390/s16010115.

[125] Ricardo Chavarriaga, Hesam Sagha, Alberto Calatroni, and et al. The opportunity challenge : A benchmark database for on-body sensor-based activity recognition. In *Pattern Recognition Letters*, volume 34, pages 2033–2042, 2013. doi : 10.1016/j.patrec.2012.12.014.

[126] Charissa Ann Ronao and Sung-Bae Cho. Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Systems with Applications*, 59 :235–244, 2016.

[127] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008. Curran Associates, Inc., 2017.

[128] Nitish Srivastava, Elman Mansimov, and Ruslan Salakhutdinov. Unsupervised learning of video representations using lstms. *International Conference on Machine Learning (ICML)*, 37 :843–852, 2015.

[129] Sijie Yan, Yuanjun Xiong, and Dahua Lin. Spatial temporal graph convolutional networks for skeleton-based action recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32. AAAI Press, 2018.

[130] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv :1503.02531*, 2015.

[131] Shuo Jiang, Peiqi Kang, Xinyu Song, Benny P. L. Lo, and Peter B. Shull. Emerging

wearable interfaces and algorithms for hand gesture recognition : A survey. *IEEE Reviews in Biomedical Engineering*, 2021. doi : 10.1109/RBME.2021.3078190.

[132] Anonymous. A review of hand gesture recognition systems based on noninvasive wearable sensors. *Wiley Online Library*, 2023. doi : 10.1002/aisy.202300207.

[133] K. W. Sim and J. Zariffa. Wearable sensors for rehabilitation : A focus on postural control and balance in individuals with stroke. *Frontiers in Neurology*, 9 :1389, 2018. doi : 10.3389/fneur.2018.01389.

[134] Matteo Atzori, Andrea Gijsberts, Claudio Castellini, Barbara Caputo, Anne-Gabrielle Hager, Sylvain Elsig, Giulio Giatsidis, Fabio Bassetto, and Henning Müller. Electromyography data for non-invasive naturally-controlled robotic hand prostheses. *Scientific Data*, 1 :140053, 2014. doi : 10.1038/sdata.2014.53.

[135] Matteo Atzori and Henning Müller. Control capabilities of myoelectric robotic prostheses by hand amputees : A scientific research review. *Frontiers in Systems Neuroscience*, 8 :44, 2014. doi : 10.3389/fnsys.2014.00044.

[136] Edwin Valarezo Añazco and et al. Hand gesture recognition using single patchable six-axis inertial measurement unit via recurrent neural networks. *Sensors*, 21, 2021.

[137] Patrick Scholl and et al. Real-time and embedded detection of hand gestures with an imu-based glove. *MDPI Journal*, 5 :28, 2018.

[138] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9 :1735–1780, 1997.

[139] Abdolrahim Bayat, Marc Pomplun, and Duc A. Tran. A study on human activity recognition using accelerometer data from smartphones. *Procedia Computer Science*, 34 :450–457, 2014.

[140] Nishkam Ravi and et al. Activity recognition from accelerometer data. In *Proceedings of the National Conference on Artificial Intelligence*. AAAI Press, 2005.

[141] R. A. R. Mendoza and et al. Hand gesture recognition using emg-imu signals and deep q-networks. *Sensors*, 22 :9613, 2022.

[142] Manfredo Atzori and et al. Control of prosthetic hands using supervised machine learning techniques : Clinical research. *Frontiers in Neurology*, 5 :68, 2014.

[143] Andy Kong and et al. Eyemu interactions : Gaze + imu gestures on mobile devices. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*. ACM, 2021.