

UNIVERSITÉ DE MONTRÉAL

ALGORITHMES DE DÉTECTION DES MALADIES NEURODÉGÉNÉRATIVES À
PARTIR DE LA DÉMARCHE D'UN INDIVIDU

IMANNE EL MAÂCHI
DÉPARTEMENT DE GÉNIE INFORMATIQUE ET GÉNIE LOGICIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE INFORMATIQUE)
DÉCEMBRE 2018

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

ALGORITHMES DE DÉTECTION DES MALADIES NEURODÉGÉNÉRATIVES À
PARTIR DE LA DÉMARCHE D'UN INDIVIDU

présenté par : EL MAÂCHI Imanne

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. DAGENAIS Michel, Ph. D., président

M. BILODEAU Guillaume-Alexandre, Ph. D., membre et directeur de recherche

M. BOUACHIR Wassim, Ph. D., membre et codirecteur de recherche

Mme ZOUAQ Amal, Ph. D., membre

DÉDICACE

*Pour ma chère maman,
la meilleure maman du monde...*

REMERCIEMENTS

Je tiens à remercier profondément mon directeur de recherche Professeur Guillaume-Alexandre Bildoeau pour son encadrement, son précieux temps et ses conseils qui m'ont permis de vivre une expérience de recherche très enrichissante. Je dois également mes sincères remerciements à mon co-directeur Professeur Wassim Bouachir pour ses corrections et ses judicieux conseils tout au long du projet. Je tiens à remercier mes collègues du laboratoire avec qui j'ai partagé cette expérience autant sur le plan social qu'intellectuel. Finalement, je vous remercie chers membres du jury d'avoir pris le temps d'examiner mon travail de recherche.

Mes amis ont été d'un grand support pour moi tout au long du projet, et je les remercie tous. Je tiens à remercier particulièrement mes chères amies Oumaima et Sarah pour leur soutien et leur chaleureuse écoute. Ma famille a été ma première source de motivation et j'en suis grandement reconnaissante. Je remercie particulièrement ma mère Souad pour sa confiance en moi et son encouragement constant. Finalement, je remercie mon cher mari Anas qui a toujours été présent pour moi. Ton amour, ta patience et ta générosité embellissent chaque instant de ma vie.

RÉSUMÉ

Le diagnostic des maladies neurodégénératives est un défi en médecine. Il repose principalement sur l'interprétation des symptômes par les médecins. Il existe donc un besoin important d'outils automatiques pouvant assister les médecins dans leurs prises de décision. Dans ce contexte, l'objectif de ce travail est de développer des algorithmes de détection des maladies neurodégénératives à partir des données de la démarche des patients.

Dans un premier temps, nous avons développé un algorithme de détection de la maladie du Parkinson. En entrée, l'algorithme utilise les forces de réaction verticale du sol (*vertical ground reaction force-VGRF*) enregistrées par plusieurs capteurs placés sous les pieds. La première composante de notre algorithme est constituée de 18 réseaux neuronaux convolutifs 1D (1D-Convnets) parallèles, traitant chacun un signal de VGRF. Chacun de ces 1D-Convnets extrait un vecteur de caractéristiques propre au signal traité. Ensuite, tous ces vecteurs de caractéristiques sont concaténés et envoyés à un réseau pleinement connecté qui les intègre et donne en sortie la classification finale. L'algorithme a été comparé à d'autres méthodes récentes dans la littérature et a démontré une amélioration de la précision de classification.

Dans un deuxième temps, nous avons utilisé les données spatio-temporelles de la démarche pour développer quatre détecteurs de maladies neurodégénératives : détecteur de Parkinson, de Huntington, de sclérose latérale amyotrophique et un détecteur de l'ensemble de ces maladies neurodégénératives combinées. Puisque la base de données était de taille réduite, nous avons utilisé les algorithmes d'apprentissage machine classiques. Pour ce faire, nous avons extrait les caractéristiques de l'amplitude et de la dynamique des fluctuations des données spatio-temporelles. Ensuite, nous avons entraîné différents classificateurs à classifier ces vecteurs. Les meilleurs résultats ont été obtenus avec une machine à support de vecteurs.

Les résultats obtenus confirment la performance des algorithmes développés. Ceux-ci pourraient être utilisés en milieu clinique dans le but d'effectuer des tests précoces pour identifier les patients qui peuvent être atteints de maladies neurodégénératives.

ABSTRACT

Neurodegenerative disease diagnosis is still a very challenging problem in medicine. It relies mainly on physician expertise and interpretation of patient's physical symptoms. Therefore, there is a great need for automatic tools that can assist physicians in their decision making. In this context, the objective of this research work is to develop algorithms for the detection of neurodegenerative diseases.

First, we developed a detection algorithm for Parkinson's disease. As input, the algorithm uses vertical ground reaction forces (VGRF) recorded from several sensors placed under the subjects' feet. The first component of our algorithm consists of 18 parallel 1D convolution neural networks (1D-Convnets), each processing a VGRF signal. Each of these convolution networks extracts a feature vector of the processed signal. Then, all these feature vectors are concatenated and sent to a fully connected network that integrates them and outputs the final classification. The algorithm has been compared to recent state-of-the-art methods and has shown an improved classification accuracy.

Second, we used spatiotemporal data of gait to develop four detectors of neurodegenerative diseases: a detector of Parkinson, Huntington, of amyotrophic lateral sclerosis and a detector of all these neurodegenerative diseases combined. Since the database was smaller, we explored classic machine learning algorithms. We first extracted fluctuation magnitudes and dynamic features of the spatiotemporal data. Then we trained different classifiers to classify these vectors. The best results were obtained with a support vector machine.

The obtained results confirm the performance of our algorithms. Our method could be used clinically to screen patients in order to identify those who may potentially have neurodegenerative diseases.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vi
TABLE DES MATIÈRES	vii
LISTE DES TABLEAUX	x
LISTE DES FIGURES	xii
LISTE DES SIGLES ET ABRÉVIATIONS	xiv
CHAPITRE 1 INTRODUCTION	1
1.1 Mise en contexte	1
1.1.1 Maladie de Parkinson	1
1.1.2 Huntington	2
1.1.3 Sclérose latérale amyotrophique (SLA)	2
1.2 Éléments de la problématique	3
1.2.1 Les données	3
1.2.2 Extraction des caractéristiques	3
1.2.3 Classification	3
1.3 Objectifs de recherche	4
1.4 Contribution	4
1.5 Plan du mémoire	4
CHAPITRE 2 REVUE DE LITTÉRATURE	5
2.1 Caractérisation de la démarche	5
2.2 Caractéristiques de démarche	6
2.2.1 Démarche des patients atteints de la maladie de Parkinson	6
2.2.2 Démarche caractéristique de la maladie de Huntington	8
2.2.3 Démarche caractéristique de la SLA	8
2.2.4 Démarche du groupe contrôle	8

2.3	Travaux antérieurs de classification de démarche	8
2.4	Travaux antérieurs de classification de signaux biomédicaux	11
2.5	Méthodes d'apprentissage machine classique	15
2.5.1	Machine à vecteurs de support (SVM)	15
2.5.2	Arbre de décision	16
2.5.3	Forêt d'arbres décisionnels	17
2.6	Méthodes d'apprentissage profond	18
2.7	Discussion	20
2.7.1	Extraction de caractéristiques	20
2.7.2	Méthodes de classification	21
CHAPITRE 3 DÉMARCHE DE L'ENSEMBLE DU TRAVAIL DE RECHERCHE .		23
3.1	Classification Parkinson-Contrôle	23
3.2	Classification des maladies neurodégénératives	24
CHAPITRE 4 ARTICLE 1 : A DEEP 1D-CONVNET FOR PARKINSON DISEASE DETECTION FROM GAIT		26
4.1	Abstract	26
4.2	Introduction	26
4.3	Related work and background	27
4.3.1	Previous gait classification methods	27
4.3.2	Background concepts	28
4.4	Methods	29
4.4.1	Method overview	29
4.4.2	Parallel 1D-Convnet branches	29
4.4.3	Fully connected bloc	30
4.5	Experiments	31
4.5.1	Database	31
4.5.2	Evaluation metrics	32
4.5.3	Training	32
4.5.4	Models used for comparaison	33
4.6	Results & Discussion	34
4.6.1	Validation and test results	34
4.6.2	Ablation study: VGRF signal selection	36
4.7	Conclusion	36
4.8	Aknowledgement	37

CHAPITRE 5	DÉTECTION DES MALADIES NEURODÉGÉNÉRATIVES	38
5.1	Données et prétraitement	38
5.2	Extraction des caractéristiques	38
5.3	Classification	40
5.4	Résultats et discussion	42
CHAPITRE 6	AUTRES PISTES EXPLORÉES	45
6.1	Étude du Parkinson	45
6.1.1	Réseaux de neurones récurrents	45
6.1.2	ResNet	45
6.1.3	Augmentation des données	46
6.2	Classification des maladies neurodégénératives	46
6.2.1	Apprentissage profond	46
6.2.2	Multiclassification	47
CHAPITRE 7	DISCUSSION GÉNÉRALE	48
7.1	Atteinte des objectifs	48
7.2	Approches utilisées	48
7.3	Méthodes d'évaluation	49
CHAPITRE 8	CONCLUSION	50
RÉFÉRENCES	52

LISTE DES TABLEAUX

Tableau 2.1	Travaux antérieurs de classification de démarche	12
Table 4.1	Layer descriptions: the first part of the DNN is composed of 18 1D-Convnet, and the second part is a fully connected network.	33
Table 4.2	Validation results: the best results are in bold and the second best results are in italic. (n(P): # of Parkinson subjects, n(N): # of control subjects, TP: # of true positive, FP: # of false positive)	35
Table 4.3	Test results: the best results are in bold and the second best results are in italic. (n(P): # of Parkinson subjects, n(N): # of control subjects, TP: # of true positive, FP: # of false positive)	35
Table 4.4	Impact of individual time series on the final score. In each row, our DNN was retrained without (w/o) two symmetric inputs Right (R) and Left (L) VGRF signals. Test set results are reported here. Results corresponding to the most significant features are in bold. (n(P): # of Parkinson subjects, n(N): # of control subjects, TP: # of true positive, FP: # of false positive)	36
Tableau 5.1	Hyperparamètres des classificateurs (Park : Parkinson, Hunt : Huntington, MND : Maladies neurodégénératives (Hunt, Park et SLA), ctrl : contrôle.	41
Tableau 5.2	SLA vs contrôle : Les meilleurs résultats (en gras) ont été obtenus avec la SVM (VF/FP : Vrais/Faux positifs, Sp : Spécificité, Se : Sensibilité, Ex : Exactitude, n(P) : # sujets atteints de SLA, n(N) : # sujets contrôle).	42
Tableau 5.3	Parkinson vs contrôle : Les meilleurs résultats sont en gras, nos meilleurs résultats ont été obtenus avec l'arbre décisionnel (VF/FP : Vrais/Faux positifs, Sp : Spécificité, Se : Sensibilité, Ex : Exactitude, n(P) : # sujets atteints de Parkinson, n(N) : # sujets contrôle).	42
Tableau 5.4	Huntington vs Contrôle : Nos meilleurs résultats (en gras) ont été obtenus avec la SVM (VF/FP : Vrais/Faux positifs, Sp : Spécificité, Se : Sensibilité, Ex : Exactitude, n(P) : # sujets atteints de Huntington, n(N) : # sujets contrôle).	42

Tableau 5.5	Maladies neurodégénératives (Parkinson, Huntington, SLA) vs Contrôle : Les meilleurs résultats (en gras) ont été obtenus avec la SVM (VF/FP : Vrais/Faux positifs, Sp : Spécificité, Se : Sensibilité, Ex : Exactitude, n(P) : # sujets atteints de maladies neurodégénératives, n(N) : # sujets contrôle).	43
-------------	---	----

LISTE DES FIGURES

Figure 2.1	Le cycle de marche : Chaque pas est composé d'une phase d'appui (<i>stance phase</i>) et d'une phase de balancement (<i>swing phase</i>) ©IEEE [29].	5
Figure 2.2	Le groupe contrôle (a) a une démarche talon-orteil, alors que les patients atteints de la maladie de Parkinson (b) ont une démarche de type pied-plat ©IEEE [31].	7
Figure 2.3	Algorithme de détection des crises d'épilepsie : Chaque signal EEG est traité séparément jusqu'à la binarisation de la probabilité de détection. La décision finale est prise par un OU logique sur l'ensemble des signaux EEG ©IEEE [42].	13
Figure 2.4	Post-traitement pour la détection des crises d'épilepsie ©IEEE [42] .	13
Figure 2.5	Classification d'ECG ©IEEE [2].	14
Figure 2.6	Réseau de neurones convolutif de Acharya et al. ©IEEE [2].	14
Figure 2.7	Comparaison d'hyperplans avec différents noyaux : A) noyau linéaire, B) noyau RBF C) noyau polynomial [1].	15
Figure 2.8	Effet du coefficient de régularisation C : A) petites valeurs de C; B) grandes valeurs de C.	16
Figure 2.9	Architecture d'un neurone artificiel	18
Figure 2.10	<i>dropout</i> : a) modèle original b) modèle pendant l'entraînement lors de l'application du <i>dropout</i> ©IEEE [40].	19
Figure 3.1	VGRF totale du pied gauche d'un patient atteint de la maladie de Parkinson (orange) et d'un sujet sain (bleu).	24
Figure 3.2	Exemple de séquence d'intervalle de pas (en msec) pour (de haut à en bas) des patients atteints de la maladie de Parkinson, Huntington, SLA et des sujets sains. Les graphiques représentent l'intervalle de temps de pas de chaque cycle de marche ©IEEE [32].	25
Figure 4.1	The architecture of the DNN: The first part is composed of 18 parallel 1D-Convnets, each processing a VGRF signal (F) coming from foot sensors. The 1D-Convnets are followed by fully connected layers that produce the final classification	30
Figure 4.2	1D-Convnet: Each 1D-Convnet block is composed of 4 convolutional layers, 2 max-pooling layers and 1 fully-connected layer.	31

Figure 5.1 Exemple de l'arbre décisionnel après entraînement pour la classification du Parkinson. L'entraînement est fait avec 30 échantillons. L'échantillon restant est utilisé pour le test (sujet contrôle dans ce cas-ci). . . 43

LISTE DES SIGLES ET ABRÉVIATIONS

SLA	Sclérose latérale amyotrophique
ApEn	Entropie approximative
AVC	Accident vasculaire cérébral
ReLu	<i>Rectified linear unit</i>
tanh	Tangeante hyperbolique
SVM	Machine à vecteurs de support
KNN	K-plus proches voisins
LBP-1D	motifs binaires locaux 1D
ECG	électrocardiogramme
EEG	électroencéphalogramme
RGB-D	<i>Red Greed Blue - Depth</i>
HMM	Modèle Markov caché (<i>Hidden markov model</i>)
Convnet	Réseau convolutif
LSTM	Long-short term memory
RBF	Fonction de base radiale
CV	Coefficient de Variation
STD_diff	Écart-type sur le signal différentiel
DFA	Analyse de fluctuation dérivée
LBP	<i>Local binary pattern</i>
VGRF	Force de réaction verticale au sol (<i>Vertical ground reaction force</i>)

CHAPITRE 1 INTRODUCTION

Le système nerveux est le chef d'orchestre de l'ensemble des organes. Lorsqu'on décide de poser une action, un signal électrique se transmet du cerveau vers les nerfs moteurs à travers la moelle épinière. Les nerfs moteurs transmettent ensuite le signal aux muscles pour exécuter l'action désirée. Les mouvements sont ainsi coordonnés par le réseau de plusieurs neurones. Ces derniers sont très précieux, car leur rôle est central et ils ne peuvent être régénérés. En effet, leur dégénération, si elle a lieu, est définitive. La dégénération des neurones est une des causes de plusieurs maladies neurodégénératives, telles que la maladie de Parkinson, la sclérose latérale amyotrophique (SLA) et le Huntington. Afin d'améliorer la qualité de vie des patients, une prise en charge et un diagnostic précoce sont primordiaux. L'expertise des médecins joue un rôle important pour la détection des signes précoces. Cependant, il y a un manque d'outils automatiques pouvant assister les médecins dans leur prise de décision. Dans ce contexte, le présent mémoire porte sur le développement de nouvelles méthodes de diagnostic des maladies neurodégénératives à partir de la démarche. Plus particulièrement, nous nous intéressons à la maladie de Parkinson, Huntington et à la SLA.

1.1 Mise en contexte

Le développement d'outils de diagnostic requiert d'abord une compréhension des maladies traitées. La présente section est consacrée à une présentation de chacune des maladies qui seront traitées dans ce mémoire.

1.1.1 Maladie de Parkinson

La maladie de Parkinson est la deuxième maladie neurodégénérative la plus fréquente après l'Alzheimer [9]. Les origines de la maladie ne sont pas encore bien précises. Des études suggèrent des causes génétiques [37], alors que d'autres des causes environnementales [3]. Toutefois, la maladie de Parkinson résulte d'une dégénération des neurones dopaminergiques du cerveau qui entraîne un dérèglement du système dopaminergique [28]. Ce déséquilibre est la principale cause des symptômes moteurs connus de la maladie de Parkinson. Le seul examen permettant de confirmer avec certitude le diagnostic de la maladie de Parkinson est un examen histologique du cerveau. Bien évidemment, ceci ne peut pas être appliqué sur les patients vivants. Ainsi, les médecins appliquent une série d'examens cliniques pour déduire si le patient est atteint de la maladie de Parkinson [23]. Premièrement, on évalue

l'hypertonie musculaire, qui est présente chez 85% des patients. Celle-ci correspond à une rigidité des membres et une résistance à initier des mouvements, elle est une des causes de la forme "dos penché" des patients. Deuxièmement, on évalue le tremblement des membres, qui est un symptôme présent chez 65% des patients. Toutefois, le symptôme le plus important est la bradykinésie, celle-ci se traduit par une lenteur marquée dans les mouvements, elle est particulièrement évidente dans les mouvements périodiques comme la marche. Le médecin procède généralement aussi à un examen d'imagerie par résonance magnétique afin d'éliminer d'autres maladies candidates qui peuvent avoir des symptômes similaires.

1.1.2 Huntington

La maladie de Huntington est un trouble rare qui touche environ 5 à 7 personnes sur 100 000 dans la population caucasienne [44]. Cette dernière est une maladie génétique causée par une forme défectueuse du gène codant pour la protéine Huntingtine [7], qui est importante pour la régularisation des fonctions cellulaires. L'évolution de la maladie diffère entre les individus, mais le développement des symptômes commence généralement vers l'âge de 30-40 ans. Le Huntington cause des troubles moteurs et cognitifs. Les troubles moteurs consistent en des mouvements irréguliers, désorganisés et spontanés. Pour le diagnostic de la maladie, le médecin évalue la présence des symptômes du Huntington. Pour confirmer le diagnostic [44], le médecin procède à un examen génétique afin de détecter la présence du gène causant la maladie.

1.1.3 Sclérose latérale amyotrophique (SLA)

La SLA touche environ 1 personne sur 100 000. Les causes de la maladie peuvent être génétiques ou environnementales [25]. La SLA est causée par la dégénération des neurones moteurs. Ceci résulte en une diminution de la masse musculaire [49]. Les premiers symptômes de la maladie sont très subtils, le patient peut ressentir de la faiblesse musculaire dans certains membres. Plus la maladie progresse, plus l'ensemble des mouvements devient limité par la faiblesse musculaire. Comme pour la maladie de Parkinson, il n'y a pas de test qui peut confirmer le diagnostic de la SLA. Les médecins procèdent ainsi à une analyse des symptômes moteurs du patient. Il est également commun de procéder à un électromyogramme. Il s'agit d'une analyse des signaux électriques dans les muscles [8].

1.2 Éléments de la problématique

Puisque la perturbation de la marche est un symptôme commun à ces maladies, nous allons étudier celle-ci. Le développement d'un détecteur de la démarche est un problème complexe. Les principaux éléments de cette problématique reposent sur la nature des données, l'extraction des caractéristiques ainsi que la classification.

1.2.1 Les données

L'analyse des données médicales peut s'avérer difficile pour plusieurs raisons. Premièrement, les bases de données sont généralement petites, ce qui peut limiter l'entraînement d'un algorithme de classification. Ceci est particulièrement important avec les algorithmes d'apprentissage profond, dont la performance dépend grandement de la taille de la base de données. Deuxièmement, les données de la démarche peuvent être biaisées par d'autres facteurs tels que l'âge et les facteurs intrinsèques des patients. Ainsi, il devient difficile de trouver des caractéristiques discriminatoires entre les classes, tout en étant communes à l'intérieur de chacun des groupes. Troisièmement, les données de la démarche présentent du bruit qui peut rendre la classification plus difficile.

1.2.2 Extraction des caractéristiques

Choisir et concevoir les caractéristiques appropriées pour décrire chacun des groupes est une tâche complexe. Dans la littérature, plusieurs études ont présenté des caractéristiques prédéfinies extraites de manière manuelle à partir des données. Toutefois, il y a tellement de différences interindividuelles, qu'il est difficile de trouver des caractéristiques qui sont robustes face à la diversité physiologique de la démarche.

1.2.3 Classification

Le choix du classificateur qui prédira le diagnostic final est une étape majeure dans le développement de l'algorithme. D'abord, nous devons choisir le type de classificateur : classificateur basé sur l'apprentissage machine classique ou sur l'apprentissage profond. Ceci dépendra principalement de la taille de la base de données et de la complexité des échantillons. Ensuite, la régularisation du classificateur est une étape qui demande de la minutie, car il faut trouver le compromis entre l'optimisation de sa capacité, tout en évitant le surentraînement. Ceci se fait à travers la recherche des hyperparamètres optimaux.

1.3 Objectifs de recherche

L'objectif principal de ce projet est de développer des algorithmes intelligents de diagnostic des maladies neurodégénératives à partir de l'analyse de la démarche. Plus spécifiquement, nous visons le développement de deux méthodes :

- Algorithme de détection de la maladie de Parkinson à partir de la démarche.
- Algorithmes de détection des trois maladies neurodégénératives (Parkinson, Huntington et SLA) à partir de la démarche.

1.4 Contribution

Les principales contributions de ce travail s'inscrivent dans le développement d'outils de diagnostic des maladies neurodégénératives à partir de la démarche. Plus précisément, nos contributions sont :

- Développement d'un algorithme de détection de la maladie de Parkinson basé sur l'apprentissage profond. Cette méthode a été présentée dans un article de journal soumis au chapitre 3.
- Identification des caractéristiques qui contribuent le plus à la détection de la maladie de Parkinson.
- Développement d'algorithmes de détection des trois maladies neurodégénératives (Parkinson, Huntington, SLA).
- Évaluation et comparaison avec les méthodes existantes de détection de la maladie de Parkinson dans les mêmes conditions expérimentales.

1.5 Plan du mémoire

Les prochains chapitres sont organisés ainsi : le chapitre 2 présente la revue de littérature qui a inspiré ce travail. Le chapitre 3 est une introduction aux méthodes utilisées. Le chapitre 4 est constitué de l'article de journal où nous présentons notre algorithme de détection de la maladie de Parkinson. Le chapitre 5 présente les méthodes développées pour détecter l'ensemble des maladies neurodégénératives étudiées à partir de la démarche. Le chapitre 6 est un survol sur les autres pistes explorées pendant la recherche des algorithmes de détection. Le chapitre 7 est une discussion générale des limites des méthodes présentées. Finalement, nous allons conclure l'ensemble de ce travail au chapitre 8.

CHAPITRE 2 REVUE DE LITTÉRATURE

Dans ce chapitre, nous allons présenter une revue de la littérature qui a inspiré ce travail de recherche. Dans un premier temps, nous présentons les principales caractéristiques de démarche et les outils de mesure de celle-ci. Ensuite, nous allons décrire la démarche de chacune des maladies étudiées par rapport à la démarche normale des sujets sains (groupe contrôle). Par la suite, nous présenterons les travaux antérieurs de la classification de la démarche et des signaux biomédicaux. Nous discuterons ensuite des classificateurs d'apprentissage machine classiques et des méthodes d'apprentissage profond. Finalement, nous allons clore le chapitre avec une discussion critique de l'état de l'art.

2.1 Caractérisation de la démarche

La principale unité pour caractériser la démarche est le cycle de marche. Un cycle comprend deux pas : un pas droit et un pas gauche. Chaque pas comprend une phase d'appui et une phase de balancement. La phase d'appui commence lorsque le talon touche le sol et se conclut lorsque le pied quitte le sol par le contact des orteils. Ensuite, la phase de balancement commence lorsque le pied quitte le sol et se termine lorsque le pied atterrit sur celui-ci. Au milieu du cycle de marche, on retrouve le moment du double support où les deux pieds touchent le sol simultanément. Cet intervalle correspond au chevauchement des phases d'appui des deux pieds (voir figure 2.1). L'ensemble de ces mesures correspond aux caractéristiques spatio-temporelles de la marche.

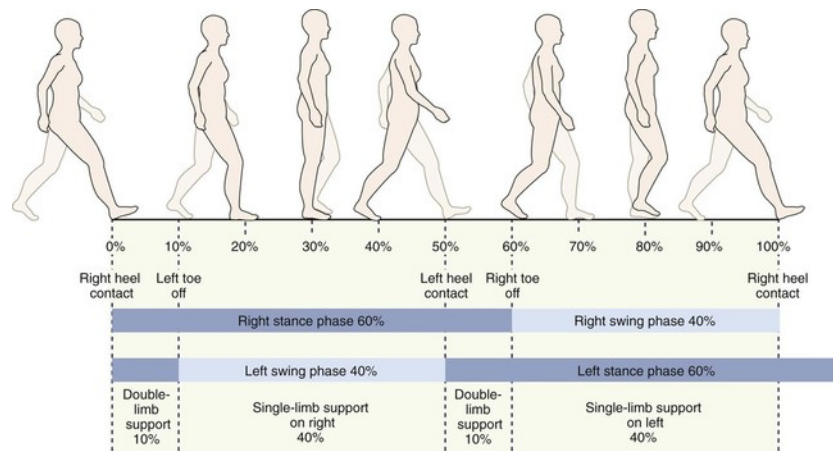


Figure 2.1 Le cycle de marche : Chaque pas est composé d'une phase d'appui (*stance phase*) et d'une phase de balancement (*swing phase*) ©IEEE [29].

Les principaux outils permettant de suivre le cycle de marche sont les capteurs portables et les caméras RGB-D (*Red Green Blue-Depth*).

1. **Capteurs portables** : Les capteurs portables sont placés sur les membres du sujet. Ils permettent d'enregistrer les données cinématiques et mécaniques. Par exemple, on peut avoir des capteurs qui mesurent les accélérations des articulations de la hanche, du genou et de la cheville. Il existe également des capteurs de force qu'on place sous les pieds pour enregistrer la force d'appui des pieds (force de réaction verticale au sol-VGRF) pendant que le sujet marche . La variation périodique de ces mesures permet de suivre le cycle de marche.
2. **Caméra RGB-D** : La caméra kinect a été utilisée plus récemment pour caractériser la démarche. La caméra a l'avantage de ne pas nécessiter des capteurs ou des marqueurs sur le sujet [14]. Cet outil permet l'analyse du cycle de marche grâce aux déplacements des jointures.

On retrouve dans la banque de données Physionet [16] deux bases de données sur des maladies neurodégénératives enregistrées à l'aide des capteurs de force. La première base de données comprend la démarche de patients atteints de la maladie de Parkinson et des sujets sains. La deuxième base de données regroupe des sujets atteints de la maladie de Parkinson, du Huntington, de la sclérose latérale amyotrophique ainsi que des sujets sains. Pour suivre le cycle de marche, des capteurs ont été placés sous chaque pied. Plusieurs études qui seront présentées dans les sections suivantes ont utilisé ces bases de données pour décrire et classifier les sujets atteints de maladies neurodégénératives.

Dans la prochaine section, nous allons décrire le cycle de marche de chacun des groupes étudiés.

2.2 Caractéristiques de démarche

Les maladies neurodégénératives étudiées affectent le système locomoteur de différentes manières. Ce sont ces différences qui permettront de développer des classificateurs pour identifier le type de maladies. Dans cette section, les démarches caractéristiques de la maladie de Parkinson, de la SLA, du Huntington ainsi que du groupe contrôle seront présentées.

2.2.1 Démarche des patients atteints de la maladie de Parkinson

La démarche d'un patient atteint de la maladie de Parkinson est affectée par l'akinésie et la bradykinésie. L'akinésie correspond à une lenteur prononcée à initier les mouvements complexes. Par exemple, le patient peut avoir de la difficulté à initier sa démarche. La bra-

dykinésie est une lenteur marquée dans les mouvements. Ainsi, la diminution de vitesse dans la démarche est répandue chez les patients atteints de la maladie de Parkinson [39].

Perumal et al. [31] ont analysé la démarche parkinsonnienne en comparant les patients avec un groupe témoin. Ils ont observé que les patients atteints de la maladie de Parkinson ont un cycle de marche plus lent, une phase de balancement plus courte, et une phase d'appui plus longue. De plus, la force exercée par le talon est plus faible chez les patients atteints de la maladie de Parkinson, ce qui indiquerait une plus faible stabilité. En outre, en analysant les forces en fonction du cycle de marche (figure 2.2), on remarque que les sujets du groupe contrôle ont deux maximums locaux pendant le cycle de marche. Ces derniers correspondent au contact du talon avec le sol, suivi par le contact des orteils avec le sol. Cependant, chez les patients atteints de la maladie de Parkinson, on remarque un seul maximum au début du cycle de marche. Ceci indique que le pied est plat lors du contact avec le sol.

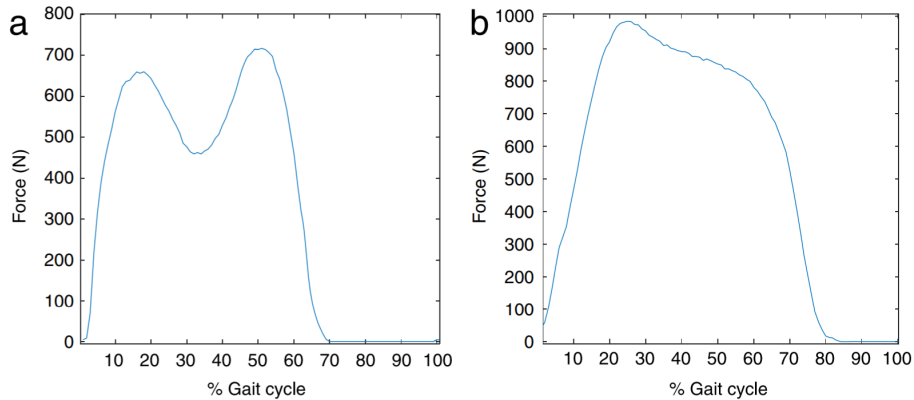


Figure 2.2 Le groupe contrôle (a) a une démarche talon-orteil, alors que les patients atteints de la maladie de Parkinson (b) ont une démarche de type pied-plat ©IEEE [31].

La démarche des patients atteints de la maladie de Parkinson est beaucoup plus désorganisée. Wu et al. [46] ont utilisé des mesures d'entropie afin de différencier les patients atteints de la maladie de Parkinson de ceux du groupe contrôle. Leurs résultats démontrent que les individus atteints de la maladie de Parkinson ont une entropie plus grande dans leur démarche, ce qui indiquerait une diminution de la stabilité.

De plus, puisque la démarche est un mouvement périodique, Sarbaz et al. [38] ont analysé le spectre fréquentiel des données spatio-temporelles. Le spectre de la phase de balancement des sujets atteints de la maladie de Parkinson a des amplitudes beaucoup plus grandes que celles du groupe contrôle. Les fréquences dominantes des intervalles de pas du groupe contrôle se situent dans les basses fréquences, alors que celles du groupe du Parkinson sont plutôt

dispersées sur le spectre.

2.2.2 Démarche caractéristique de la maladie de Huntington

La maladie de Huntington se distingue des autres maladies par une plus grande instabilité et des fluctuations beaucoup plus importantes entre les cycles de marche. La dynamique des fluctuations est tellement importante qu'il n'y a pas de modèle pour définir les données spatio-temporelles : les fluctuations dynamiques sont presque complètement aléatoires [18].

2.2.3 Démarche caractéristique de la SLA

Les patients atteints de la SLA ont également une démarche anormale. Tout comme la démarche parkinsonienne, leur démarche est moins stable et plus désorganisée dans le temps. Ceci est particulièrement traduit par les variations successives de l'intervalle de temps des pas. Toutefois, cette instabilité est moins prononcée que celle des patients atteints de la maladie de Huntington. On note également une diminution de la vitesse à travers des cycles de marche beaucoup plus lents [18].

2.2.4 Démarche du groupe contrôle

La démarche du groupe contrôle est beaucoup plus stable que celles des sujets atteints de maladies neurodégénératives. Toutefois, on note une différence inter-individus. En effet, Hausdorff et al. [20] ont remarqué que le système locomoteur de chaque individu semble avoir une mémoire qui caractérise la variabilité entre les cycles de marche.

2.3 Travaux antérieurs de classification de démarche

Plusieurs méthodes de classification ont été développées dans les travaux antérieurs. La majorité des algorithmes se divisent en deux étapes principales :

1. **Construction du vecteur de caractéristiques** : Extraction des caractéristiques de démarche à l'aide d'outils statistiques ou de traitement de signal.
2. **Classification** : Entraînement d'un classificateur pour apprendre à déterminer la classe des sujets à l'aide du vecteur de caractéristiques construit à l'étape précédente.

D'une part, il y a les recherches qui ont utilisé directement les données spatio-temporelles de la démarche, comme celles fournies par Physionet. Par exemple, Xia et al. [47] ont classifié les maladies neurodégénératives en utilisant des classificateurs binaires. Comme entrée, ils ont utilisé les séquences des intervalles des pas, de la phase d'appui et le double support.

Pour chaque caractéristique, la moyenne, l'écart-type, le maximum, le minimum, l'asymétrie et le coefficient d'acuité ont été extraits. Au niveau de la classification, ils ont comparé des algorithmes d'apprentissage machine classiques tels que la machine à vecteurs de support (SVM), la forêt d'arbres décisionnels, le perceptron et les K-plus proches voisins (KNN). L'algorithme SVM a montré la meilleure performance (exactitude entre 96% et 100%).

Comme cité précédemment, Sarbaz et al. [38] ont utilisé le spectre fréquentiel des données spatio-temporelles. Ils ont choisi les intervalles de temps des pas droit et gauche et la phase de balancement du pas gauche. Ils ont omis la phase d'appui, car elle est similaire à l'intervalle des pas. Ils ont choisi le pied droit et gauche pour les intervalles de pas, car la maladie de Parkinson provoque une asymétrie dans la démarche. Cette asymétrie est plus visible dans l'intervalle des pas que dans celui de la phase de balancement. Après avoir obtenu les spectres fréquentiels de chaque signal. Chaque spectre a été divisé en 10 bandes fréquentielles sur lesquelles l'aire sous la courbe a été calculée. Un réseau de neurones à trois couches a ensuite été utilisé pour classifier les maladies. Ils ont obtenu une exactitude de 93.75%.

Comme mentionné précédemment, Wu et al. [46] ont utilisé des mesures d'entropie. Les données spatio-temporelles des intervalles de temps des pas, des phases de balancement et des phases d'appui constituaient les entrées de leur algorithme. Dans un premier temps, ils ont filtré les données ayant une différence de plus de trois écarts-types avec la moyenne à l'aide d'un filtre médian. Dans un deuxième temps, ils ont utilisé les mesures d'entropie suivantes :

1. Entropie approximative (ApEn) : l'ApEn permet de mesurer les irrégularités dans les séquences temporelles en comparant des sous-segments du signal original. Plus le ApEn est grand, plus le signal est irrégulier [33].
2. L'entropie symbolique normalisée : Pour obtenir cette mesure d'entropie, on procède en trois étapes.
 - (a) Binarisation du signal temporel à l'aide d'un seuillage.
 - (b) Construction de segments de longueur L en regroupant L signaux consécutifs.
 - (c) Construction d'un histogramme de la distribution des segments formés à l'étape précédente.
 - (d) Calcul de l'entropie à partir de l'histogramme.
3. Nombre de changements de la direction du signal : le changement de la direction du signal est un changement du signe de la dérivée du signal. Afin d'être robuste face au bruit, on compte uniquement les changements ayant une différence absolue supérieure à un seuil défini. Cet indicateur est beaucoup plus grand chez les patients atteints de la maladie de Parkinson parce qu'ils ont une démarche moins stable.

Pour la classification, les auteurs ont entraîné un SVM à distinguer entre une démarche normale et une démarche parkinsonienne en utilisant ces trois mesures d'entropie. Ils ont obtenu une exactitude de 84.48%

Dans un travail subséquent, Joshi et al. [24] ont introduit les ondelettes pour diagnostiquer la maladie de Parkinson. Les ondelettes permettent de décomposer de manière fréquentielle le signal tout en restant dans le domaine temporel. En utilisant les données spatio-temporelles de la démarche normale et parkinsonienne, ils ont construit un vecteur de caractéristiques avec les coefficients de la convolution du signal avec les ondelettes. Ensuite, ils ont entraîné un classificateur SVM à discriminer entre les deux démarches. Ils ont obtenu une exactitude de 90.32%

D'autre part, des études ont utilisé directement les forces d'appui des pieds avec le sol comme entrée à leur algorithme. Par exemple, Ertugrul et al. [11] ont appliqué les motifs binaires locaux (LBP-1D) sur les mesures de forces fournies par la base de données de Physionet. Ils ont ensuite construit un histogramme représentant la distribution des motifs 1D-LBP. Par la suite, ils ont extrait des mesures statistiques (temporelles et fréquentielles) de l'histogramme. Celles-ci ont été classifiées avec différents algorithmes, tels que le réseaux de Bayes, le perceptron, et la forêt d'arbres décisionnels. La meilleure exactitude obtenue est de 88.88% avec le perceptron.

Mannini et al. [27] ont proposé une approche différente pour classifier la démarche des sujets âgés, des patients atteints de la maladie de Huntington, et des patients ayant subi un accident vasculaire cérébral (AVC). Les données de la démarche ont été obtenues grâce à six capteurs d'accélération situés dans différentes parties du corps. Le signal temporel de chaque capteur a été analysé selon la manière suivante :

1. Modélisation de la démarche de chaque classe à l'aide d'un modèle de Markov caché (*Hidden Markov model* HMM).
2. Construction d'un vecteur de caractéristiques pour chaque sujet en utilisant la vraisemblance logarithmique du signal avec chacun des modèles HMM construits à l'étape précédente. On ajoute ensuite à ces vraisemblances logarithmiques obtenues des caractéristiques temporelles (moyenne, écart-type, etc.) et fréquentielles du signal (fréquence dominante, puissance, etc.).
3. Entraînement d'un classificateur SVM à classifier les vecteurs de caractéristiques selon la classe du sujet.

Finalement, la décision est prise à l'aide d'un principe de vote de la prédiction de tous les signaux des capteurs. Cette méthode leur a permis d'obtenir une exactitude de 90.5%.

Wang et al. [45] ont adopté un algorithme similaire pour classifier les démarches des patients atteints de la maladie de Parkinson, de Huntington et des sujets sains. Les entrées de leur algorithme sont les séquences d'intervalle des phases de balancement, d'appui et de double support. Dans un premier lieu, ils ont entraîné le classificateur KNN à modéliser chacune des maladies. Ensuite, ils ont pu obtenir les vraisemblances logarithmiques de chaque entrée avec chaque classe. Dans un deuxième lieu, ils ont construit un vecteur de caractéristiques combinant les vraisemblances logarithmiques avec d'autres caractéristiques statistiques des signaux. Un SVM a ensuite été entraîné à classifier les vecteurs de caractéristiques. Cette méthode leur a permis d'obtenir une exactitude de 74.51%.

Récemment, Zhao et al. [51] ont développé un algorithme d'apprentissage profond pour détecter la maladie de Parkinson. Cette technique leur a permis d'éviter l'extraction manuelle des caractéristiques de la marche. Leur modèle est composé de deux réseaux parallèles. Le premier réseau analyse la distribution spatiale des forces avec un réseau de neurones convolutifs 2D (2D-Convnet). Le second réseau analyse la distribution temporelle à l'aide d'un réseau de neurones récurrents avec des couches de LSTM (*Long short-term memory*). La classification finale est obtenue par la moyenne des deux algorithmes. Ils ont obtenu des exactitudes de l'ordre de 98%.

On remarque donc que les algorithmes ont généralement la même structure présentée au début de la section : construction d'un vecteur de caractéristiques et classification. L'ensemble des travaux présentés est résumé dans le tableau 2.1.

2.4 Travaux antérieurs de classification de signaux biomédicaux

L'analyse de la démarche est un problème qui ressemble beaucoup à l'analyse d'un électrocardiogramme (ECG) ou d'un électroencéphalogramme (EEG). En effet, ces trois analyses ont plusieurs points en commun :

1. Il s'agit d'un signal 1D qui doit être traité et classifié.
2. Il s'agit de signaux physiologiques qui varient périodiquement.
3. Les signaux traités contiennent du bruit qui rend l'analyse plus difficile.
4. Puisqu'il s'agit de données médicales, les bases de données sont relativement petites comparativement à d'autres problèmes d'apprentissage machine.
5. Il y a présence d'une inter-variabilité entre les sujets.

En raison de ces points communs, certaines études dans ces domaines seront examinées.

En ce qui concerne les études sur le EEG, Bhattacharya et al. [4] ont utilisé la transformée en ondelettes empiriques afin de décomposer le signal en différentes fréquences. L'avantage de

Tableau 2.1 Travaux antérieurs de classification de démarche

Algorithme	Entrée	Caractéristiques	Classificateur	Classes
Sarabaz et al. [38] (2012)	Données spatio-temporelles	Aire sous la courbe du spectre fréquentiel	Réseau de neurones	Parkinson et contrôle
Xia et al. [47] (2015)	Données spatio-temporelles	Statistiques	SVM	Huntington, SLA, Parkinson et contrôle
Ertugel et al.[11] (2016)	VGRF	Statistiques de l'histogramme LBP	Apprentissage machine	Parkinson et contrôle
Mannini et al. [27] (2016)	Accélération	Vraisemblance logarithmique de HMM et statistiques	SVM	Huntington, contrôle et AVC
Wu et al.[46] (2017)	Données spatio-temporelles	Entropie	SVM	Parkinson et contrôle
Wang et al. [45] (2017)	Données spatio-temporelles	Vraisemblance logarithmique de k-NN et statistiques	SVM	Parkinson, Huntington et contrôle
Joshi et al.[24] (2017)	Données spatio-temporelles	Ondelette	SVM	Parkinson et contrôle
Zhao et al. [51](2018)	VGRF	-	2D-Convnet, LSTM	Parkinson et contrôle

cet algorithme est qu'il permet de construire une banque de filtres du style ondelettes basée sur les fréquences du signal [15]. Cette approche, combinée avec d'autres outils de traitement de signal, leur a permis de mieux classifier les signaux présentant des anomalies.

Dans le même contexte, Temko et al. [42] ont développé un algorithme de détection des crises d'épilepsie chez les nouveau-nées (figure 2.3). Les données d'entrée de leur algorithme étaient constituées de huit signaux d'EEG. Chacun d'eux a été segmenté en intervalles de temps et analysé séparément. Ensuite, un ensemble de statistiques temporelles ont été extraites de chaque segment pour construire le vecteur de caractéristiques. Ces données ont servi à entraîner un SVM à fournir une probabilité de la présence d'une crise pour chaque intervalle de temps. Cette probabilité a été post-traitée en appliquant un filtre passe-bas par moyenne mobile suivi d'un seuillage pour binariser la sortie (figure 2.4). Les huit sorties ont finalement été combinées par un OU logique. Pour chaque intervalle de temps, s'il y a un signal positif provenant d'une sonde, on signale la présence de crise d'épilepsie.

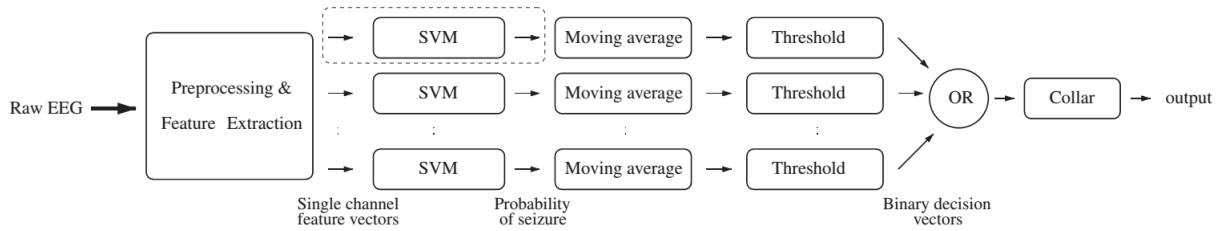


Figure 2.3 Algorithme de détection des crises d'épilepsie : Chaque signal EEG est traité séparément jusqu'à la binarisation de la probabilité de détection. La décision finale est prise par un OU logique sur l'ensemble des signaux EEG ©IEEE [42].

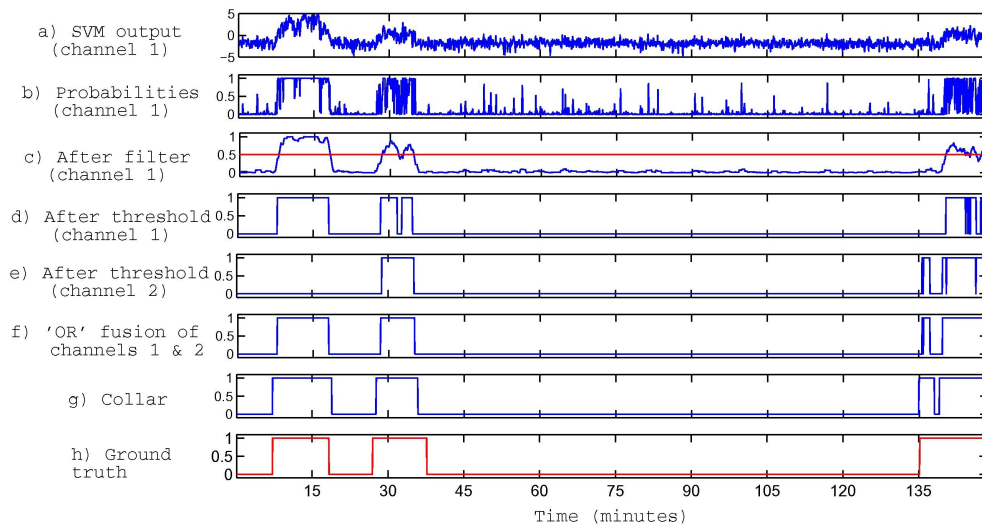


Figure 2.4 Post-traitement pour la détection des crises d'épilepsie ©IEEE [42]

En ce qui concerne le ECG, Acharya et al. [2] ont classifié des signaux ECG de quatre groupes différents (figure 2.5). Ils ont développé un algorithme basé sur l'apprentissage profond. En effet, puisque la recherche d'un vecteur de caractéristiques peut être ardue, l'utilisation d'un réseau de neurones permet un apprentissage automatique de cette étape. Ils ont donc utilisé un réseau de neurones convolutif 1D à plusieurs couches cachées (figure 2.6). Ceci leur a permis d'atteindre des bons résultats de classification, sans utiliser des méthodes d'extraction de caractéristiques.

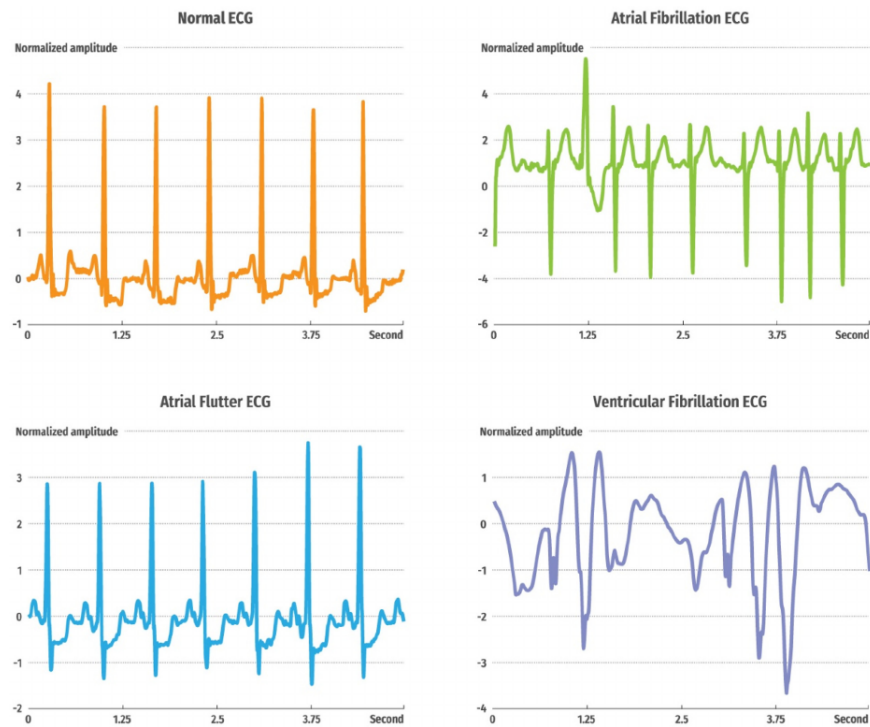


Figure 2.5 Classification d'ECG ©IEEE [2].

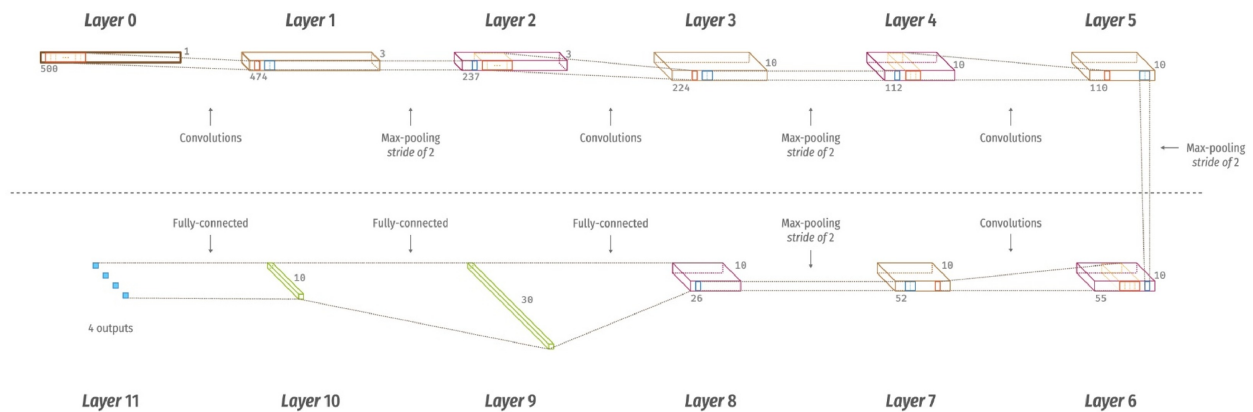


Figure 2.6 Réseau de neurones convolutif de Acharya et al. ©IEEE [2].

Suite à notre présentation des travaux pertinents de classification de la démarche et des signaux biomédicaux, les prochaines sections de ce chapitre porteront un regard plus approfondi sur les algorithmes d'apprentissage supervisé. Dans un premier temps, nous présenterons les algorithmes d'apprentissage machine classiques. Dans un deuxième temps, nous présenterons les algorithmes d'apprentissage profond.

2.5 Méthodes d'apprentissage machine classique

Il existe plusieurs algorithmes d'apprentissage machine classique. Nous allons en aborder trois qui seront utilisés ultérieurement : SVM, arbre de décision, et la forêt d'arbres décisionnels.

2.5.1 Machine à vecteurs de support (SVM)

La classification avec la SVM repose sur la détermination d'un hyperplan qui sépare deux classes. L'hyperplan est défini en fonction de la marge maximale. La marge maximale est la distance entre l'hyperplan et les échantillons les plus proches de chaque classe [43]. Toutefois, afin de ne pas se limiter aux données linéairement séparables, il est possible d'utiliser un noyau afin de séparer les données non linéairement séparables. Cela consiste à transformer la représentation des données dans un espace plus grand où elles seront linéairement séparables. Pour se faire, on utilise le produit scalaire de nos données avec l'espace de plus grandes dimensions. Les noyaux non linéaires permettent de faire la correspondance avec cet espace de plus hautes dimensions. Parmi les types de noyaux communément utilisés, on retrouve la fonction de base radiale (RBF), la fonction polynomiale, et la fonction gaussienne (figure 2.7).

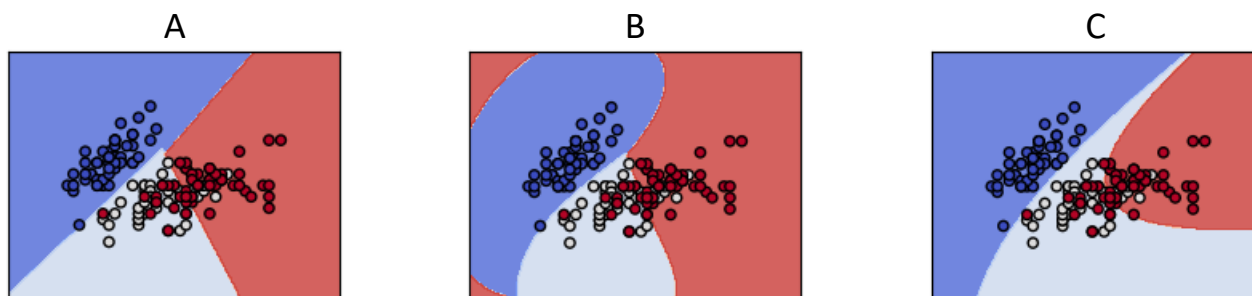


Figure 2.7 Comparaison d'hyperplans avec différents noyaux : A) noyau linéaire, B) noyau RBF C) noyau polynomiale [1].

La régularisation du SVM est une étape importante pour optimiser la classification. Un des facteurs de régularisation les plus importants est le paramètre C . Celui-ci est impliqué dans la fonction de minimisation des poids de l'hyperplan de la manière suivante :

$$\min\left(\frac{1}{2}\|W\|^2 + C\sum_{i=1}^m \zeta\right), \quad (2.1)$$

où W correspond aux poids (coefficients) de l'hyperplan, et ζ est une variable proportionnelle aux erreurs de classification. Ainsi, le paramètre C indique le coût $C\sum_{i=1}^m \zeta$ lorsque l'algorithme fait des erreurs. C'est un indicateur du compromis qu'on est prêt à faire pour maximiser la marge de l'hyperplan vis-à-vis la présence d'erreurs de classification. Par conséquent, des grandes valeurs de C vont faire pencher l'algorithme à choisir des marges petites, si ces marges permettent de bien classer les données d'entraînement. Toutefois, ceci pose un risque de surapprentissage. De l'autre côté, de faibles valeurs de C vont plutôt tendre l'algorithme à maximiser sa marge, même si cette marge diminue la précision de la classification (voir figure 2.8).

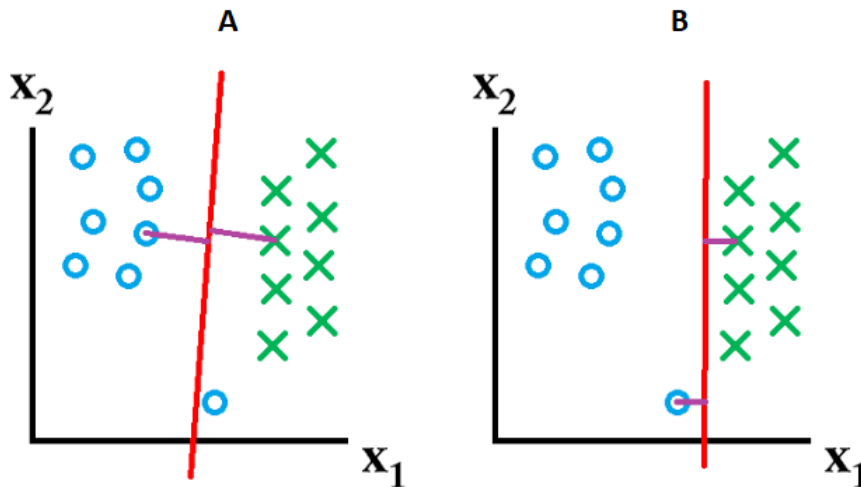


Figure 2.8 Effet du coefficient de régularisation C : A) petites valeurs de C ; B) grandes valeurs de C .

2.5.2 Arbre de décision

L'arbre décisionnel se base sur des règles de décisions dans chacun des nœuds. Ces règles de décisions reposent sur les attributs des caractéristiques. Les feuilles de l'arbre correspondent aux résultats des décisions successives du modèle [36]. La profondeur de l'arbre détermine sa capacité : plus la profondeur est grande, plus la capacité est grande, mais plus le risque de

surapprentissage est grand. Les hyperparamètres, tels que le nombre de données minimum pour diviser un nœud et le nombre minimum de données par feuille, contrôlent la profondeur de l'arbre.

Lors de la construction de l'arbre, la règle de décision de chaque nœud est choisie en fonction de la minimisation de l'entropie ou de l'impureté des branches qui vont en résulter. Ainsi, les caractéristiques qui permettent de mieux distinguer les classes sont choisies dans les premières règles de décision de l'arbre (proche de l'origine).

Il existe plusieurs types d'arbres de décision, les principaux sont ID3, C4.5 et CART.

- ID3 est le premier arbre développé, il a été introduit par Ross Quinlan en 1986 [34]. À chaque nœud, l'algorithme choisit le critère test en évaluant le gain en entropie. Toutefois, cet algorithme n'accepte pas les caractéristiques continues
- L'algorithme C4.5 présente quelques améliorations par rapport à l'algorithme ID3 [35]. Premièrement, l'algorithme accepte les caractéristiques numériques continues. Pour ce faire, l'algorithme procède à une partition dynamique de l'espace continu en intervalles définis. On peut ainsi passer des caractéristiques continues à des caractéristiques discrètes. De plus, cet algorithme contient une étape de régularisation qui permet d'éviter le surapprentissage. Une fois que les règles de décision des nœuds sont définies, l'algorithme compare la performance de l'arbre avec et sans chacun des nœuds. Ainsi, si l'élimination d'un nœud améliore la performance de l'arbre, celui-ci sera supprimé de l'arbre.
- L'algorithme CART est très similaire à C4.5. La principale amélioration ajoutée est qu'il peut être utilisé dans les problèmes de régression [5].

2.5.3 Forêt d'arbres décisionnels

Une forêt d'arbres décisionnels contient plusieurs arbres de décision. La décision finale est prise en considérant la moyenne de tous les arbres. Toutefois, dans l'arbre décisionnel, la règle de décision à chaque nœud est basée sur les meilleures caractéristiques qui donnent le fractionnement le plus homogène dans chacune des divisions. Dans la forêt décisionnelle, la division est choisie sur les meilleures caractéristiques d'un sous-ensemble de caractéristiques choisies aléatoirement. Donc, un arbre individuel dans la forêt d'arbres peut être moins performant que l'arbre décisionnel, mais la moyenne de tous les arbres de la forêt permet de diminuer l'erreur globale et généralement d'avoir des meilleurs résultats [26].

2.6 Méthodes d'apprentissage profond

L'apprentissage profond a connu un essor important pendant la dernière décennie, cela est principalement dû aux bases de données volumineuses et aux outils informatiques de plus en plus performants. En termes d'architecture, l'apprentissage profond repose sur des réseaux de neurones qui contiennent plusieurs couches intermédiaires, communément appelées couches cachées. En partant des données d'entrée vers la sortie, chaque couche cachée extrait l'information de la couche précédente et peut ainsi représenter l'information à un niveau d'abstraction plus élevé, jusqu'à arriver à la sortie. Donc, avec suffisamment de données, le réseau de neurones apprend à extraire les caractéristiques des données d'entrée pendant l'entraînement [17].

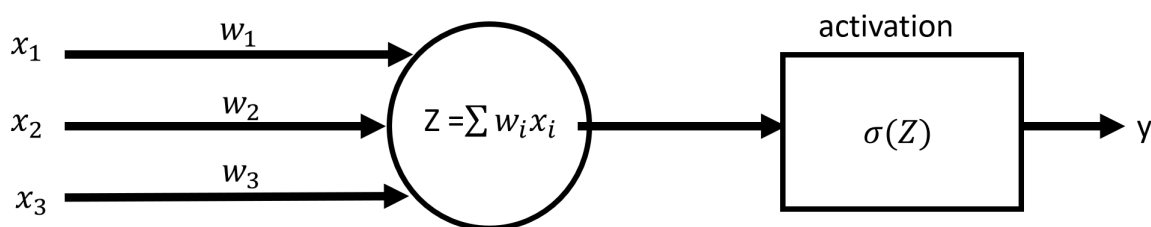


Figure 2.9 Architecture d'un neurone artificiel

Chaque couche est composée d'un ou de plusieurs neurones. Comme illustré à la figure 2.9, le neurone fait une sommation pondérée sur les entrées auxquelles il est connecté, il applique ensuite une fonction d'activation non linéaire sur cette somme pour obtenir sa sortie. Les fonctions non linéaires utilisées sont ReLU, sigmoïde et tanh. La fonction ReLU est la plus utilisée, car elle permet d'entraîner le réseau de neurones plus rapidement [17]. L'apprentissage repose sur l'adaptation des poids W : le réseau de neurones doit apprendre les valeurs W optimales pour obtenir l'estimation la plus proche de la valeur réelle. Initialement, les poids sont assignés aléatoirement.

Pour s'entraîner, le réseau de neurones utilise les données d'entrée (X) et leur sortie (Y). À chaque itération, le réseau de neurones produit une sortie \hat{Y} , cette estimation est comparée avec la valeur de référence d' Y pour calculer l'erreur d'estimation. En fonction de celle-ci, les poids du réseau de neurones sont réajustés par l'algorithme de rétropropagation afin de diminuer l'erreur [17]. Pendant l'entraînement, les minimums locaux sont rares, cependant, l'entraînement peut être difficile avec la présence de points-selles [17]. Il s'agit de points où la performance de l'algorithme ne s'améliore pas pendant plusieurs itérations d'entraînement.

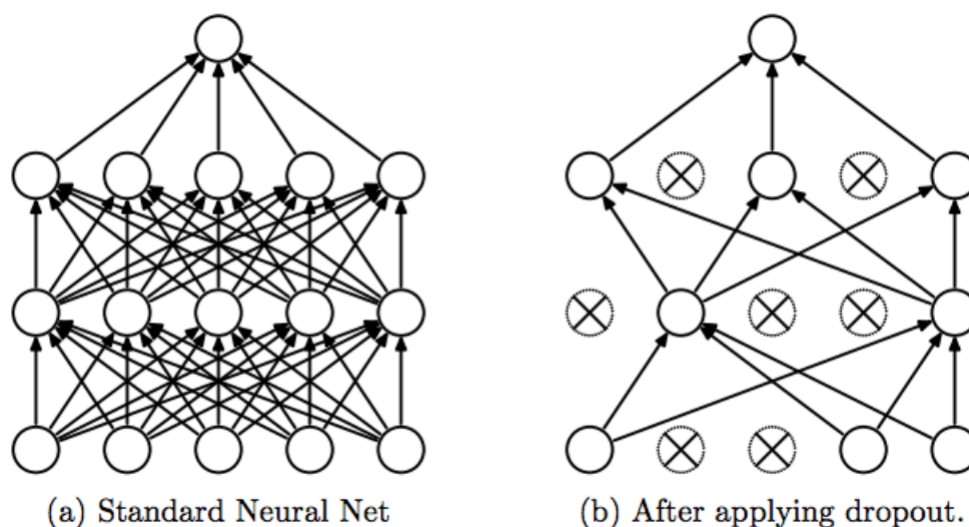


Figure 2.10 *dropout* : a) modèle original b) modèle pendant l'entraînement lors de l'application du *dropout* ©IEEE [40].

Le surentraînement est l'un des défis les plus rencontrés avec les réseaux profonds. Le réseau de neurones a de la difficulté à généraliser aux données qu'il n'a pas vues pendant l'entraînement. Dans ces cas, l'erreur sur l'ensemble d'entraînement est faible, alors que l'erreur sur l'ensemble de test est grande : le réseau apprend à corrélérer au maximum avec l'ensemble d'entraînement et est incapable de généraliser sur d'autres données. Pour éviter cela, on a recours à des techniques de régularisation comme le *dropout* [40]. Cette technique consiste à ignorer un pourcentage de neurones par couches pendant les entraînements, les neurones ignorés seront choisis aléatoirement à chaque itération d'entraînement. Ceci permet l'entraînement d'un plus petit réseau à chaque fois (figure 2.10).

Lorsque la base de données n'est pas très grande, il est possible de faire du transfert d'apprentissage [30]. Cela consiste à utiliser un réseau préentraîné sur une première grande base de données et de l'appliquer dans un autre problème semblable en modifiant uniquement les dernières couches.

Il existe plusieurs types de réseaux de neurones profonds. Dans notre travail, nous utiliserons principalement les réseaux convolutifs neuronaux (ConvNets). Ces derniers apprennent des filtres spatiaux qui pourront extraire l'information pertinente d'un signal. Ainsi, les poids W sont les valeurs des filtres spatiaux que l'algorithme devra apprendre. Il s'agit donc d'une organisation différente des neurones : chaque couche contient un ou plusieurs filtres, dont la dimension et le nombre sont prédéfinis. Les réseaux convolutifs contiennent des couches de convolutions, d'échantillonnage (*max-pooling*) et des couches pleinement connectées :

1. Convolution : Convolution entre le signal d'entrée de la couche et le filtre spatial appris. Le signal résultant passe ensuite par une fonction d'activation (ReLU généralement). Le résultat de cette convolution s'appelle carte de caractéristiques (*feature map*).
2. Échantillonnage : Cette étape permet de diminuer les dimensions des cartes de caractéristiques pour optimiser les calculs et augmenter le niveau d'abstraction. En *max-pooling*, on choisit la valeur maximale dans une fenêtre d'échantillonnage définie. Il y a une perte de précision à cette étape.
3. Couches pleinement connectées : Elles se retrouvent généralement à la fin du réseau. Elles intègrent l'information fournie par les convolutions afin d'arriver à une décision à la dernière couche. Celles-ci sont construites avec des simples neurones comme celui illustré dans la figure 2.9.

Si nous prenons par exemple le réseau de Acharya et al. pour classifier les signaux ECG : le signal traverse en premier des blocs de convolutions et d'échantillonnages jusqu'à la huitième couche. Ensuite, on a des couches pleinement connectées jusqu'à la couche 11. Cette dernière couche donne les probabilités que l'échantillon appartienne à chacune des classes.

2.7 Discussion

La démarche est un mouvement périodique ayant des caractéristiques spatio-temporelles. Les bases de données publiques disponibles portant sur les maladies neurodégénératives ont été enregistrées à l'aide de capteurs de VGRF. En utilisant ces bases de données, plusieurs algorithmes de classification commencent par l'extraction des caractéristiques de démarche. Ces caractéristiques sont par la suite classifiées par des algorithmes d'apprentissage machine. Dans un premier temps, nous allons discuter des différentes caractéristiques de démarche présentées. Ensuite, notre discussion portera sur les différentes méthodes de classification.

2.7.1 Extraction de caractéristiques

Il existe une multitude de caractéristiques de démarche. Certaines études ont utilisé des caractéristiques provenant de comparaisons statistiques et physiologiques entre les groupes de patients, alors que d'autres ont plutôt utilisé des caractéristiques obtenues à l'aide de techniques de traitement de signal. À notre opinion, il est plus rigoureux d'utiliser des caractéristiques provenant d'études statistiques. En effet, cela permet une interprétation des résultats d'un point de vue clinique. De plus, ce choix est d'autant plus important lorsque les bases de données sont petites. En effet, l'algorithme de détection doit généraliser à d'autres

sujets. Le choix de caractéristiques ayant démontré une différence statistiquement significative augmente la probabilité de généralisation. Par contre, une technique de traitement de signal peut offrir de bons résultats avec une base de données spécifique, mais il est possible que la performance diminue pour des bases de données plus grandes.

Par exemple, Wu et al. [46] ont utilisé plusieurs caractéristiques telles que la moyenne et les extremums. Cependant, rien ne prouve que les minimums et maximums sont statistiquement différents dans la démarche parkinsonienne. De plus, il se peut qu'il y ait une redondance entre ces attributs. Pareillement, Ertugrul et al. [11] ont utilisé les motifs LBP-1D. Joshi et al. [38] ont utilisé les caractéristiques provenant d'ondelettes. Mannini et al. [27] et Wang et al. [45] ont utilisé des vraisemblances logarithmiques parmi leurs caractéristiques. Ces algorithmes ont donné des bonnes performances. Cependant, ces caractéristiques n'ont pas préalablement prouvé une différence statistiquement significative entre les groupes étudiés.

D'un autre côté, Sarbaz et al. [38] ont préalablement analysé le spectre fréquentiel des patients. Ceci leur a permis de choisir les caractéristiques fréquentielles qui ont démontré la meilleure distinction entre la démarche parkinsonienne et la démarche normale. Également, Wu et al. [46] ont utilisé les caractéristiques de l'entropie, parce que la démarche parkinsonienne est reconnue par son instabilité. Ils ont démontré dans leur étude que les caractéristiques choisies étaient statistiquement différentes chez les patients atteints de la maladie de Parkinson. Cette démarche est plus rigoureuse d'un point de vue clinique et plus facilement interprétable.

2.7.2 Méthodes de classification

Nous avons présenté les méthodes d'apprentissage machine classiques et les méthodes d'apprentissage profond. Pour les algorithmes d'apprentissage machine classiques, la SVM peut avoir une grande capacité pour classifier les données non-linéairement séparables grâce aux noyaux. Il est par contre plus difficile de trouver les paramètres optimaux pour ce classificateur. C'est d'ailleurs un algorithme qui a souvent été choisi comme classificateur des caractéristiques de démarche [47] [27] [46] [45] [24]. L'arbre décisionnel est assez facile à interpréter, ce qui est très avantageux dans un contexte biomédical. Toutefois, on risque d'avoir un surapprentissage, surtout lorsque la base de données est petite. Finalement, la forêt d'arbres décisionnels permet de minimiser le surapprentissage de l'arbre de décision en utilisant plusieurs arbres. En raison de l'aspect aléatoire dans le choix des tests de décision, les résultats peuvent, cependant, être plus difficilement reproductibles.

Les algorithmes d'apprentissage machine classiques requièrent l'extraction et la sélection des caractéristiques. Il s'agit d'un choix ardu, car certaines caractéristiques peuvent donner des

bonnes performances dans certains cas seulement. En effet, la démarche est un mouvement complexe qui diffère pour chaque individu dépendamment de sa santé, de son âge et d'autres facteurs intrinsèques. Il est donc difficile d'extraire manuellement des caractéristiques robustes aux différences individuelles. L'apprentissage profond permet de faire l'extraction des caractéristiques automatiquement. Les caractéristiques extraites explicitement par les modèles d'apprentissage profond sont toutefois très difficiles à interpréter.

Très peu de travaux ont abordé l'apprentissage profond pour la classification de la démarche. Récemment, les travaux de Zhao et al. [51] ont introduit l'application de cette méthode pour la détection de la maladie de Parkinson. En utilisant l'apprentissage profond, ils ont évité l'extraction manuelle des caractéristiques. Ceci est un grand avantage, car l'algorithme extrait des caractéristiques qui peuvent être très complexes à définir et concevoir manuellement. Avec une telle approche, la base de données doit être suffisamment grande afin d'obtenir un modèle généralisable. Dans leur étude, Zhao et al. [51] ont utilisé la base de données des forces VGRF des patients atteints de la maladie de Parkinson et des sujets sains. Ils disposaient d'environ 305 marches enregistrées avec 12 000 points par signal de VGRF. Ils ont utilisé des 2D-Convnets pour analyser les signaux 1D des VGRF. Comme avantage, cette architecture permet d'analyser l'interaction entre les différents signaux des capteurs de force. Toutefois, l'algorithme apprend les interactions en fonction du positionnement des signaux dans leur matrice d'entrée. Un changement des positions des signaux 1D dans la matrice pourrait ainsi affecter l'apprentissage.

Pour l'analyse des signaux biomédicaux, l'approche de Acharya et al. [2] nous semble intéressante. L'utilisation de plusieurs couches de 1D-Convnet permet l'extraction de motifs pouvant détecter les différentes maladies cardiaques. Les Convnets ont montré une grande performance avec les images. Leur apprentissage est même interprétable qualitativement [50]. Nous pensons que cette approche est prometteuse pour traiter les signaux 1D des VGRF.

En ce basant sur cette observation, nous allons utiliser les techniques d'apprentissage profond dans le but de détecter les sujets atteints de la maladie de Parkinson. La base des données des VGRFs brutes contient assez de données pour permettre l'apprentissage d'un réseau de neurones profond. Dans un deuxième temps, nous allons développer des algorithmes pour détecter les trois maladies neurodégénératives étudiées (Huntington, SLA, Parkinson). Puisque dans ce cas la base de données était relativement petite, il est difficile d'utiliser l'apprentissage profond. Nous allons plutôt explorer les algorithmes d'apprentissage machine classiques en se basant sur les attributs de la démarche de chacune des maladies présentées.

CHAPITRE 3 DÉMARCHE DE L'ENSEMBLE DU TRAVAIL DE RECHERCHE

L'objectif de ce travail est de développer des classificateurs qui permettent de détecter les maladies neurodégénératives à partir de l'analyse de la démarche. Il s'agit d'un problème d'apprentissage supervisé où :

- Entrée X : Données de la démarche
- Sortie Y : Classe du sujet ; contrôle ou atteint de maladies neurodégénératives

Pour accomplir cela, deux bases de données publiques sont à notre disposition :

- La maladie de Parkinson : Cette base de données comprend les données de forces brutes VGRF provenant de capteurs placés sous les pieds des patients atteints de la maladie de Parkinson et des sujets sains.
- Les maladies neurodégénératives : Cette base de données comprend les données spatio-temporelles des patients atteints de la maladie de Huntington, de la maladie de Parkinson, de sclérose latérale amyotrophique et des sujets du groupe contrôle.

Nous avons donc travaillé sur chacune de ces bases de données afin de développer des algorithmes d'apprentissage supervisé capables de détecter les maladies neurodégénératives. Les méthodes développées se divisent ainsi en deux parties, la première portant sur la détection de la maladie de Parkinson. La deuxième méthode développée porte sur la détection de chacune des maladies neurodégénératives.

3.1 Classification Parkinson-Contrôle

Dans cette première partie, nous avons opté pour un algorithme d'apprentissage profond parce que la base de données est suffisamment grande pour entraîner un réseau de neurones profond.

Les données disponibles sont des signaux 1D correspondant à la force exercée sous les pieds pendant la marche. Pour les obtenir, 8 capteurs de force ont été placés sous chaque pied. Au total, 18 séquences temporelles sont disponibles : 16 (8×2) signaux provenant des capteurs de force et deux forces totales pour chaque pied. La figure 3.1 illustre un exemple de force totale sous le pied gauche d'un patient atteint de la maladie de Parkinson, comparativement à un sujet sain.

En moyenne, pour chaque marche, 12 000 points de données sont enregistrés à une fréquence de 100 données par seconde. Sachant qu'un cycle de marche moyen dure 1 seconde, nous

avons échantillonné les signaux en des segments de 100 points de données. Ainsi, chaque échantillon d'entrée contenait 18 séquences temporelles de 100 points chacune.

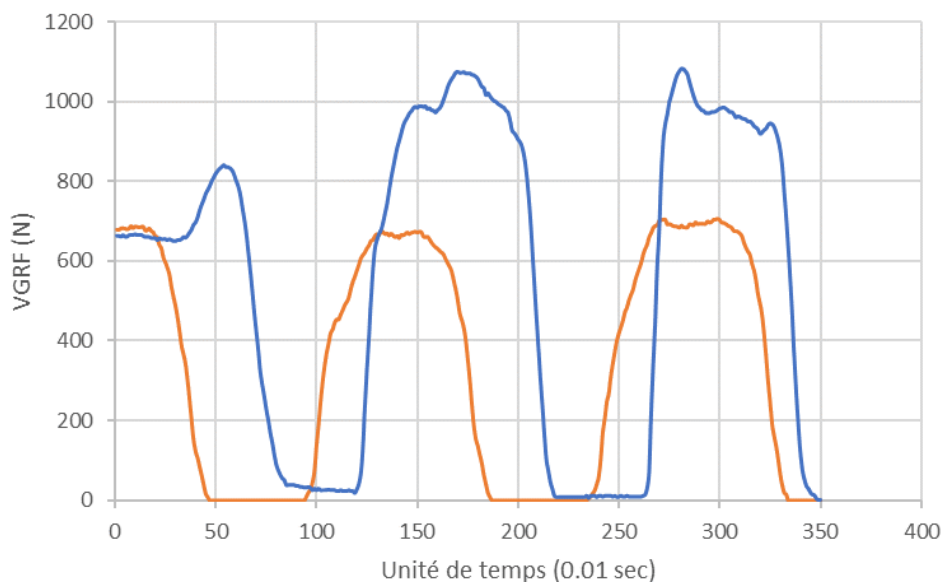


Figure 3.1 VGRF totale du pied gauche d'un patient atteint de la maladie de Parkinson (orange) et d'un sujet sain (bleu).

Afin de classifier ces échantillons, nous avons utilisé un réseau neuronal convolutif. Nous avons conçu l'architecture du réseau selon deux objectifs principaux. Premièrement, nous devons extraire les caractéristiques propres de chaque signal 1D. Deuxièmement, nous devons étudier l'interaction entre les différents signaux. Pour ce faire, nous avons développé un réseau qui contient 18 réseaux convolutifs parallèles, traitant chacun un signal temporel. Chacun de ces réseaux extrait un vecteur caractéristique du signal d'entrée. Ensuite, ces vecteurs sont concaténés et envoyés à un réseau de neurones pleinement connecté, dont la sortie est la classification finale du signal d'entrée.

Cette méthode est expliquée davantage dans le chapitre 4 qui correspond à un article de journal soumis.

3.2 Classification des maladies neurodégénératives

Pour cette deuxième classification, nous avons opté pour des algorithmes d'apprentissage machine classiques, car la base de données disponible est relativement petite. Celle-ci contient

les données spatio-temporelles du cycle de marche des 3 maladies neurodégénératives (Parkinson, Huntington, SLA) et du groupe contrôle. À titre d'exemple, la figure 3.2 présente une comparaison de l'intervalle de pas (msec) pour quatre sujets provenant de chaque groupe étudié.

Nous avons développé quatre classificateurs : Parkinson vs contrôle, SLA vs contrôle, Huntington vs contrôle, et maladies neurodégénératives (SLA, Parkinson, Huntington) vs contrôle. La première partie de notre algorithme est l'extraction d'un vecteur de caractéristiques. Nous avons utilisé les statistiques qui ont démontré une différence significative entre les différents groupes dans la littérature. Ensuite, en utilisant ces attributs, nous avons entraîné différents classificateurs à détecter chacune de ces maladies. Ces travaux seront présentés en détail dans le chapitre 5.

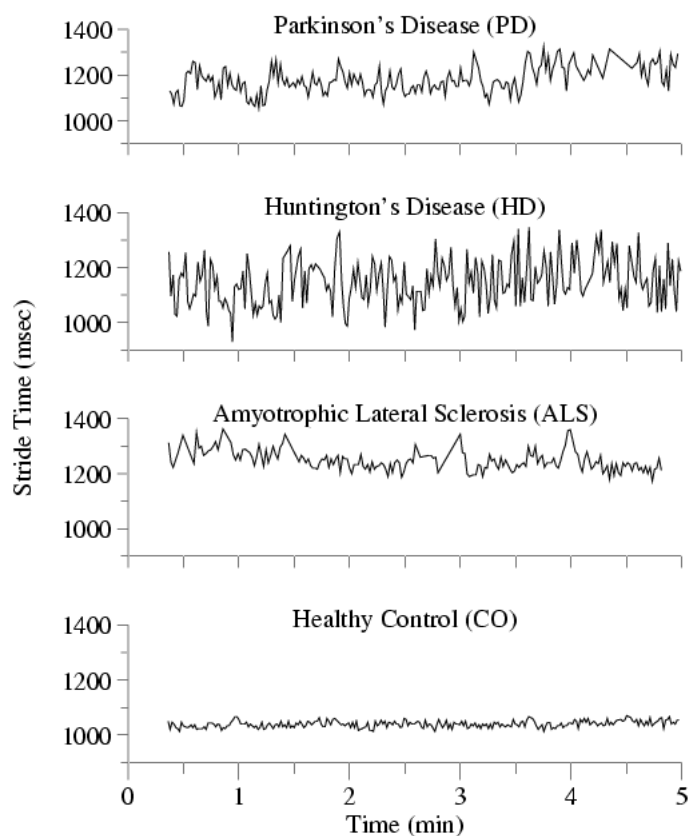


Figure 3.2 Exemple de séquence d'intervalle de pas (en msec) pour (de haut à en bas) des patients atteints de la maladie de Parkinson, Huntington, SLA et des sujets sains. Les graphiques représentent l'intervalle de temps de pas de chaque cycle de marche ©IEEE [32].

CHAPITRE 4 ARTICLE 1 : A DEEP 1D-CONVNET FOR PARKINSON DISEASE DETECTION FROM GAIT

Auteurs : Imanne El Maâchi, Guillaume-Alexandre Bilodeau et Wassim Bouachir

Soumis à : *Pattern Recognition Letters*

4.1 Abstract

Early diagnosis of Parkinson disease is a key step to ensure effective treatment. The analysis of a patient's gait can provide important indicators to allow early diagnosis. However, this physical examination relies on physician expertise and subjectivity. An intelligent algorithm which analyzes a patient's gait and provides a diagnosis may facilitate physician decision making. In this paper, we propose a novel intelligent Parkinson detection system using deep learning techniques. We used 1D convolutional neural network (1D-Convnet) to build a Deep Neural Network (DNN) classifier. The proposed network processes 18 1D-signals coming from foot sensors measuring the vertical ground reaction force (VGRF). The first part of the network consists of 18 parallel 1D-Convnet corresponding to system inputs. The second part is a fully connected network that connects the concatenated outputs of the 1D-Convnets to obtain a final classification. Our experiments demonstrate the high efficiency of the proposed method in the detection of Parkinson disease. In particular, the proposed algorithm achieved an improved accuracy of 90% compared to recent state-of-the-art methods. The results show that the model is able to learn intrinsic characteristics from gait data and to generalize to unseen subjects, which could be helpful in a clinical diagnosis.

4.2 Introduction

Today, over 10 millions people suffer from Parkinson disease. Unfortunately, no cure exists to heal this disorder, that is why early diagnosis is important to improve the patient's treatment. Currently, physicians evaluate some primary symptoms such as shaking, difficulty to initiate movements, slowness, and difficulty to walk. Despite the significant interest in parkinsonian gait analysis, there is no objective tool to assist physicians in their decision making. Since changes in the gait are among the first symptoms of this disease, a powerful gait classifier would be helpful for physicians in the diagnosis process.

Parkinsonian gait is mainly characterized by small steps, a slower gait cycle, a shorter swing phase, a longer stance phase and a flat foot strike instead of toe-to-heel strike [31]. This

abnormal gait is also marked by a reduced stability, which translates to an increased entropy [46].

To detect these characteristics, feature extraction methods have been widely used [11] [6] [38] [47] [27] [45]. Previous studies used temporal [11] [47] [27] [45] or frequential [6] [38] tools to get differentiable patterns between normal and parkinsonian gait. However, gait is a physiological characteristic that differs for each person according to age, health, and other intrinsic factors. Therefore, manual preprocessing and feature extractions will always be limited in their capacity. To avoid hand-crafted signal processing, we propose a novel gait classifier based on deep learning without explicit feature extraction. The proposed Deep Neural Network (DNN) is composed of two parts. The first one consists of 18 parallel 1D-Convnets, each processing a vertical ground reaction force (VGRF) signal coming from a foot sensor. Each 1D-Convnet extracts deep features that are subsequently concatenated together. The second part is a fully connected network that processes the concatenated vector to output the final decision. Thanks to its ability to extract relevant gait features and to generalize to unseen subjects, the proposed model outperforms state-of-the-art methods by achieving an accuracy of 90%.

The main contribution of this work is twofold. Firstly, we developed a 1D-Convnets that extracts relevant deep features for accurate gait classification, which avoids manual feature extraction. Secondly, we show that our proposed method generalizes well to data from unseen subjects, which is extremely valuable for clinical decision support systems.

The rest of the paper is organized as follows. In section 4.3, we present a review of previous works on gait classification, as well as background concepts. Section 4.4 introduces the proposed method. Section 4.5 describes the experimental setup. Section 4.6 presents results and discussion, and we will finally conclude in section 4.7.

4.3 Related work and background

4.3.1 Previous gait classification methods

Various feature extraction and classification approaches have been explored in previous work for gait analysis. Some of them extracted temporal patterns, while others used frequential features. In the temporal domain, Ertuğrul et al. [11] proposed an algorithm based on shifted 1D local binary patterns (1D-LBP) with machine learning classifiers. They used 18 VGRF input signals coming from foot sensors of Parkinson patients and control subjects. For each signal, they applied shifted 1D-LBP to construct 18 histograms of the 1D-LBP patterns from which they extracted statistical features, such as entropy, energy and correlation. Finally,

they concatenated the features from all the 18 histograms and used various supervised classifiers, such as Random Forest and Multi-Layer Perceptron (MLP), to classify the feature vectors. In the frequential domain, Daliri et al. [6] applied a Short-Time Fourier transform (STFT) on each input signal using the same VGRF input data. Then, they extracted the mean frequency and the variance of the frequency to construct a histogram representing the distribution of features extracted from all input signals. This histogram was processed with a feature discriminant ratio (FDR) in order to select the most significant bins. The final bins were then classified by a Support Vector Machine (SVM) with a chi-square kernel.

For other gait disorders, Mannini et al. [27] first trained a Hidden Markov Model (HMM) on acceleration signals recorded with sensors from elderly, post-stroke and Huntington’s disease patients. Then, they constructed a feature vector using the HMM log-likelihood of each patient with each class, combined with various temporal and frequency features. The final vector was sent to an SVM classifier. Wang et al. [45] proposed a similar algorithm, but instead of using an HMM model to extract features, they trained a K-Nearest Neighbor (KNN) classifier and then used the probability given by that classifier for each class as a feature. An SVM was later trained to classify feature vectors containing those probabilities with other statistics. Xia et al. [47] used spatiotemporal features of neurodegenerative disease patients and normal subject patients. For each time series, they computed various statistics, such as the fuzzy entropy, skewness, and kurtosis. They used machine learning classifiers, such as Random Forest, SVM, MLP, and KNN. After an optimization with a features selection algorithm, the best result was obtained with an SVM.

Recently, Zhao et al. [51] developed a deep learning algorithm to detect Parkinson disease. Their model was composed of two parallel networks. The first network analyzed the spatial distribution of forces with a 2D-Convnet. The second network analyzed the temporal distribution with a recurrent neural network (RNN). The final classification was decided by the average of both channels.

4.3.2 Background concepts

The method that we propose is based on deep learning. To be self-explanatory, some basic concepts are briefly explained in this section. The learning in DNN relies on intermediate layers, commonly called hidden layers. The input is processed at each layer in order to reach higher abstraction levels. During the training stage, the DNN increases its accuracy by using stochastic optimizers that decrease the loss at each iteration. Then, with enough data, the DNN learns how to extract features from the input [17]. Each layer is composed of neurons. The neuron computes a weighted summation of the inputs to which it is connected.

A nonlinear activation function is then applied to obtain the neuron output. The learning is based on the adaptation of the weights W : The deep network learns the optimal W values to obtain the closest estimate to the real value [17]. In order to avoid over-fitting, regularization techniques, such as dropout, are commonly used. Dropout reduces the over-fitting by shutting off some neurons chosen randomly at each iteration during the training [40]. In our model, we use convolutional layers, max-pooling layers, and fully-connected layers. Each type of layer is briefly described:

- **Convolutional layers**: they perform a spatial convolution between the inputs and the filters. The filters contain weights which are the learned elements of the layer.
- **Max-pooling layers**: they sample the input to form an output with smaller dimensions by selecting the maximum value element.
- **Fully connected layers (FC)**: they connect the output of the convolution layers to the final output. They allow a convergence of the decision at a higher level.

4.4 Methods

4.4.1 Method overview

The objective of our algorithm is to classify each subject’s walk into one of the categories: Parkinson and control. Each recorded walk was documented with 18 VGRF 1D signals from several foot sensors. Those signals were divided into m segments labeled with the subject category. Those segments are the input samples of our DNN. The DNN is composed of two parts (figure 4.1). The first part consists of 18 parallel 1D-Convnets, while the second part is a fully connected network that operates on the concatenation of the features extracted by the 18 1D-CNN. The final walk classification was decided according to the majority classification of all the subject segments.

4.4.2 Parallel 1D-Convnet branches

The first part of the network is composed of 18 parallel 1D-Convnets (figure 4.2). Each network takes as input a VGRF signal going through 4 convolutional layers, that lead to a fully connected layer. Every two convolutional layers are followed by a max-pooling layer. Each filter in the convolutional layers extracts a specific pattern that allows differentiating between the Parkinson and control group. This 1D-Convnets parallelization allows the treatment of each signal independently. In fact, since each sensor records a specific data from a specific point, each time series has its own deep features and is analyzed separately with the 1D-Convnets parallelization.

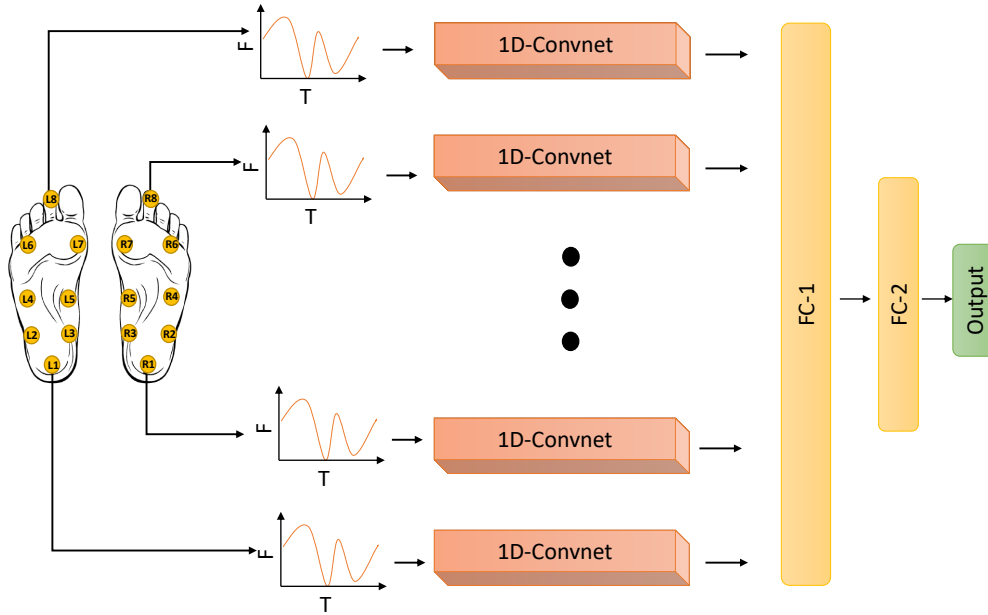


Figure 4.1 The architecture of the DNN: The first part is composed of 18 parallel 1D-Convnets, each processing a VGRF signal (F) coming from foot sensors. The 1D-Convnets are followed by fully connected layers that produce the final classification

4.4.3 Fully connected bloc

The second part of the network is designed to learn the relationship between the spatial feature extracted from the 1D-Convnets and the final output classification. First, the outputs of those 18 parallel 1D-Convnets are all concatenated into one single vector. This deep feature vector pass through two fully connected layers leading to the output layer. The output layer is composed of one neuron that predicts the classification probability. This design has two main advantages:

- Selection: the network selects the most significant time series that contribute the most to the performance of the algorithm.
- Abstraction: time series from different sensors are concatenated only after passing through the 1D-Convnets. This allows getting significant features at the concatenation step, which is more efficient than taking raw data from the input.

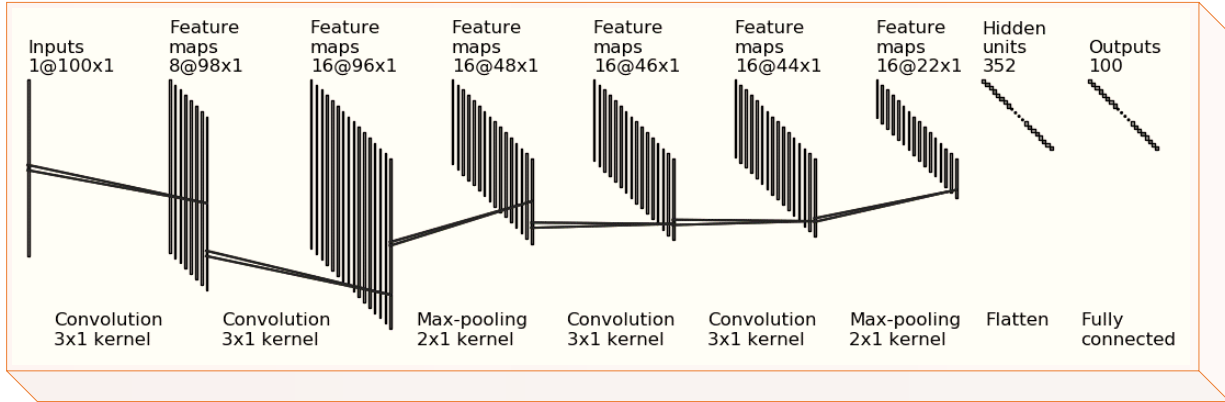


Figure 4.2 1D-Convnet: Each 1D-Convnet block is composed of 4 convolutional layers, 2 max-pooling layers and 1 fully-connected layer.

4.5 Experiments

4.5.1 Database

We used the public database collected by Physionet, which includes data reported by Frenkel et al. [12] [13], Yogev et al. [48] and Hausdorff et al. [19]. In total, 93 parkinsonian patients and 73 control subjects participated in those studies. The dataset also includes double-task time series, where subjects were walking while doing another activity. This results in an unbalanced 305 walks recorded (70% Parkinson - 30% control). For each walk, 18 time series signals are available: 16 (8×2) VGRF recorded from 8 sensors on each foot and 2 total VGRFs under each foot. For all the dataset, the sampling frequency was 100 samples per second.

Since there is no standard evaluation dataset split from previous papers, we did our own division (available online with the source code¹). The dataset was divided into training (80%), validation (10%) and test sets (10%). All folds were unbalanced with 70% Parkinson subjects data and 30% control data subjects. The validation set was used to tune the hyper-parameters, especially the number of layers, number of filters, learning rate, and the optimizer type. The test set was used once the model is finalized.

1. Link will be inserted upon paper acceptance

4.5.2 Evaluation metrics

Each of the validation and test sets is composed of 9 control subjects and 21 Parkinson patients. The control group is identified as the negative (N) group and the Parkinson group is the positive (P) group. We Use the following notations:

- Number of true positive (TP): Parkinson subject correctly classified
- Number of true Negative (TN): control subject correctly classified
- Number of false Positive (FP): control subject misclassified
- Number of false Negative (FN): Parkinson subject misclassified

The specificity (Sp), sensitivity (Se), and accuracy (Acc) are calculated as follows:

$$Se = \frac{TP}{TP + FN} \quad (4.1)$$

$$Sp = \frac{TN}{TN + FP} \quad (4.2)$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.3)$$

4.5.3 Training

Since deep learning algorithms require large dataset, each walk was divided into smaller segments of 100 time steps with 50% overlap. This resulted in 51425 training datas. Each segment was labeled with the patient category for the training. The deep network was trained to classify those segments. During the validation and test phases, the output probability of each segment was binarized (control or Parkinson), then, the preponderant class (>50%) for the given segmented whole walk determined the final classification.

The final hyperparameters of the DNN are presented in table 4.1. A rectified linear unit (ReLU) activation function was used for all the hidden layers and a sigmoid activation was used for the final output. The algorithm was trained using Adam optimizer with a batch size of 800 and an initial learning rate of 0.001. We used early stopping to avoid over-fitting: if the validation accuracy did not improve for 10 epochs, we stopped the training and went back to the weights that gave the best validation accuracy. From there, we decreased the learning rate by a factor of 5 and continued the training. This process was repeated 6 times to complete the training of the model. Also, in order to avoid over-fitting, some dropout was applied to the DNN to regularize the training.

Table 4.1 Layer descriptions: the first part of the DNN is composed of 18 1D-Convnet, and the second part is a fully connected network.

	Layer no	Layer type	Number of units	Kernel size	Dropout
1D-Convnet $\times 18$	1	Convolutional	8	3	0
	2	Convolutional	16	3	0
	3	Max-pooling	-	2	0
	4	Convolutional	16	3	0
	5	Convolutional	16	3	0
	6	Max-pooling	-	2	0
	7	Flatten	-	-	-
	8	FC	100	-	0.5
FC	9	Concatenate	-	-	0.5
	10	FC	100	-	0.5
	11	FC	20	-	0.5
	12	Output	1	-	-

4.5.4 Models used for comparaison

We compared our method with the work of Ertuğrul et al. [11] and Zhao et al. [51]. Ertuğrul et al. work was chosen for comparison because they used a simple manual feature extraction method that achieved accuracy results comparable with other machine learning algorithms. Zhao et al. work was chosen because they also use deep learning algorithms, but with a different network architecture. We reproduced both methods (according to the authors' paper description) in order to compare algorithms on the same training, validation and test sets.

Ertuğrul et al. [11] had employed a feature extraction algorithm summarized as follows:

- For each signal (18 in total):
 - Extract 1D-LBP with 8 neighbors. The output has the same length as the input with values between 2^0 and 2^8 .
 - Construct a histogram of the frequency of the obtained LBP.
 - Extract statistical features from the histogram such as skewness, entropy, energy, a coefficient of variation, kurtosis and correlation.
- Concatenate the features of all the histograms.
- Train a machine learning classifier with the feature vectors created.

For the reproduction of the work of [11], three classifiers were tested: Naive Bayes (NB), MLP and Random Forest. A grid search of hyper-parameters was performed for each classifier in order to find the classifier setup that gives the best accuracy in the validation set.

We have also reproduced the work of Zhao et al. [51]. As described before, they have employed a two-channel network: a 2D-Convnet and an RNN. The 2D-Convnet was built with 2 convolutional layers, 2 max-pooling layers, a FC layer, and an output classification layer. The RNN was built with two Long short-term memory (LSTM) layers, an FC layer, and an output classification layer. The final classification was made by the average of both channels. More implementation details are available in their paper [51]. They used the same 18 VGRF features, to which they added a time vector. Each input sample was an array of $19 \text{ features} \times 100 \text{ time steps}$. Their final output is a classification of the sample input. In order to evaluate their algorithm on a subject base, we took the majority classification of all the segmented walk samples to determine the subject final classification (as we do with our algorithm).

We used Keras with Tensorflow backend and sklearn libraries for models implementation.

4.6 Results & Discussion

4.6.1 Validation and test results

The validation set results are presented in table 4.2. Among the three machine learning classifiers for Ertuğrul et al. model [11], the best results were obtained with random forest classifier with an accuracy of 90%. The DNN proposed in [51] had a specificity of 78%, a sensitivity of 95% and an accuracy of 90%. Our DNN was able to achieve a perfect score by classifying all the validation subjects. A high validation score was expected since we tune the hyperparameters of our DNN according to the validation accuracy. Moreover, since the subject is classified according to the majority (more than 50%) class of individual walk segments, a score of 100% does not mean that all the walk segments were classified correctly but that the majority walk segments of each subject were classified correctly.

The test set was evaluated once all algorithms were tuned. Unexpectedly, the MLP showed better results (accuracy of 80%) on the test set than on the validation set. The Random Forest showed weak results (accuracy of 76%) compared to the validation set. This suggests that it is relatively easy to overfit the hyperparameters of the random forest classifier with the validation set. Zhao et al. [51] algorithm was unable to generalize to the test set. This can be explained by the fact that the DNN learned to memorize the training set and is unable to generalize to the test set. In fact, in their work [51], Zhao et al. reported that each 100 time stamp segment is considered as an independent segment. Therefore, we hypothesize that some walk segments of the same subject were probably divided between the training and test set, which could explain the discrepancy with the accuracy reported in their paper (around

Table 4.2 Validation results: the best results are in bold and the second best results are in italic. (n(P): # of Parkinson subjects, n(N): # of control subjects, TP: # of true positive, FP: # of false positive)

Algorithm	TP (n(P)=21)	FP (n(N)=9)	Sp (%)	Se (%)	Acc (%)
NB [11]	7	<i>1</i>	<i>89</i>	33	50
Random Forest [11]	21	3	67	100	<i>90</i>
MLP [11]	19	7	22	90	70
DNN [51]	<i>20</i>	2	78	<i>95</i>	<i>90</i>
Proposed DNN	21	0	100	100	100

Table 4.3 Test results: the best results are in bold and the second best results are in italic. (n(P): # of Parkinson subjects, n(N): # of control subjects, TP: # of true positive, FP: # of false positive)

Algorithm	TP (n(P)=21)	FP (n(N)=9)	Sp (%)	Se (%)	Acc (%)
NB [11]	4	4	33	19	23
Random Forest [11]	19	5	44	90	76
MLP [11]	<i>17</i>	2	78	<i>81</i>	<i>80</i>
DNN [51]	1	0	100	5	33
Proposed DNN	19	<i>1</i>	<i>90</i>	90	90

98%). Our proposed DNN showed an accuracy of 90%, which means that the algorithm has a good capacity to generalize to unseen patients in the test set, without over-fitting to individual samples.

Compared to other tested models, our algorithm is more adapted to the gait classification problem. First, gait time series are non-linear and noisy signals. The proposed DNN is more appropriate for such signals due to the use of non-linear activation functions. Second, hand-crafted methods are mostly inefficient in extracting discriminative gait features for a specific recognition problem. For example, in the work of [11], LBP patterns were extracted and analyzed starting from VGRF signals. In contrast, the proposed model learns specific spatial kernels based on the data that allows the best distinction between the Parkinson and the control group. It is known that by increasing the model capacity, deep learning algorithms may make the classification vulnerable to overfitting. The algorithm may in this case memorize individual subject gait characteristics instead of memorizing discriminative patterns between the Parkinson and control gaits. Nevertheless, our algorithm showed a high capacity to generalize to the test set, which suggests that our 1D-Convnet has learned discriminative features between the parkinsonian and normal gait.

4.6.2 Ablation study: VGRF signal selection

In order to determine which VGRF time series has the most important impact on the final classification, we deleted each symmetric right and left VGRF branches (R1 & L1 for example) and we trained the model with the remaining 16 1D-Convnets using the same hyperparameters as before.

Table 4.4 Impact of individual time series on the final score. In each row, our DNN was retrained without (w/o) two symmetric inputs Right (R) and Left (L) VGRF signals. Test set results are reported here. Results corresponding to the most significant features are in bold. (n(P): # of Parkinson subjects, n(N): # of control subjects, TP: # of true positive, FP: # of false positive)

VGRF inputs	TP (n(P)=21)	FP (n(N)=9)	Sp (%)	Se (%)	Acc (%)
w/o L1 & R1	18	2	80	85	83.5
w/o L2 & R2	18	1	90	85	86.5
w/o L3 & R3	18	1	90	85	86.5
w/o L4 & R4	18	2	80	85	83.5
w/o L5 & R5	19	1	90	90	90
w/o L6 & R6	17	1	90	80	83
w/o L7 & R7	17	1	90	80	83
w/o L8 & R8	17	1	90	80	83
w/o Total VGRF (R & L)	18	1	90	85	86.5

The results of the VGRF signal selection are presented in table 4.4, The deletion of features L6 & R6, L7 & R7 and L8 & R8 had the most important impact on the performance of the algorithm. Indeed, the overall precision decreased from 90% to 83% following the removal of these features. According to figure 4.1, we observe that the corresponding sensors are mainly placed around the toes. Therefore, we hypothesize that their importance may be due to the flat foot strike of the Parkinson group compared to the toe-to-heel strike of the control group. In fact, the two strikes would have different distributions of VGRF in those points, which makes the corresponding signals quite discriminative for the parkinsonian gait.

4.7 Conclusion

Parkinson diagnosis is still a very challenging problem in medicine because it is theoretically impossible to confirm a Parkinson diagnosis. Doctors can detect this disorder by analyzing patient symptoms through a physical examination. However, this process depends mainly on the physician interpretation. In this paper, we proposed an intelligent Parkinson detector based on gait data. Our algorithm uses deep learning techniques, which avoids the drawbacks

of hand-crafted feature extraction. Since the proposed DNN reached an accuracy of 90% , this system can serve as a practical tool to screen a population in order to detect potential Parkinson patients in a clinical context. We believe that the performance and robustness of the algorithm can still be improved by exploiting larger datasets. For future studies, it would be interesting to get inside the DNN layers and understand what they have learned. Such study would allow a deeper understanding of the parkinsonian gait and its characteristics.

4.8 Acknowledgement

This research is partly funded by Fonds de Recherche du Quebec-Nature et Technologies (FRQ-NT) with grant No. 2016-PR-189250, and an institutional research fund from TELUQ University.

CHAPITRE 5 DÉTECTION DES MALADIES NEURODÉGÉNÉRATIVES

Ce chapitre porte sur les algorithmes développés dans le cadre de la détection des maladies neurodégénératives du Parkinson, Huntington et SLA. L'algorithme que nous avons développé comprend une étape de prétraitement, une étape d'extraction des caractéristiques, et une étape de classification.

5.1 Données et prétraitement

Comme mentionné précédemment, ces groupes se distinguent par les anormalités dans leur démarche. L'objectif de notre étude est de développer un classificateur qui différencie chaque maladie du groupe contrôle. La base de données utilisée a été fournie par Hausdorff et al. [18] et elle est disponible publiquement sur Physionet¹. Elle comprend les données spatio-temporelles recueillies de 64 individus, dont 17 atteints de la maladie de Parkinson, 21 atteints de Huntington, 13 atteints de SLA et 16 individus sains. Chaque sujet a marché pour environ 200 à 300 cycles de marche qui sont documentés de manière séquentielle. Pour chaque cycle, toutes les données spatio-temporelles étaient fournies : les intervalles de temps des pas, les phases d'appui, les phases de balancement et le double support.

Les signaux ont d'abord été prétraités pour enlever les données aberrantes : les données des 20 premières secondes ont été éliminées, car elles étaient biaisées par l'effet du début de la marche. Deuxièmement, un filtre médian a été appliqué aux données qui ont un écart de plus de 3 écarts-types de la moyenne du signal. Ces données correspondaient souvent au moment où les sujets effectuaient des mouvements de rotation [18].

5.2 Extraction des caractéristiques

Ensuite, à partir de ce signal prétraité, nous avons construit un vecteur de caractéristiques propre à chaque patient. Nous avons choisi d'utiliser les caractéristiques étudiées par Hausdorff et al. [18]. En effet, Hausdorff et al. ont publié une étude statistique portant sur la démarche des patients atteints de maladies neurodégénératives. Nous avons décidé d'utiliser leurs résultats et d'aller plus loin en développant des détecteurs des maladies neurodégénératives. Nous avons choisi quatre statistiques qui ont démontré des différences significatives entre les quatre groupes. Ces statistiques décrivent l'amplitude et la dynamique des fluctuations.

1. <http://physionet.fri.uni-lj.si/physiobank/database/gaitnnd/>

1. Amplitude des fluctuations :

- (a) *Écart-type du signal différentiel (STD_DIFF)* : une différence discrète de premier degré a d'abord été appliquée au signal. Ensuite, l'écart-type a été calculé sur cette séquence. L'utilisation du signal différentiel permet de minimiser l'effet des changements locaux.
- (b) *Coefficient de variation (CV)* : (σ : Écart-type, μ : Moyenne)

$$CV = \frac{\sigma}{\mu} \quad (5.1)$$

2. Dynamique des fluctuations :

- (a) *Analyse de fluctuation dérivée (DFA)* : DFA est une méthode d'analyse des marches aléatoires. Elle nous informe sur la variation des propriétés de corrélation à travers différentes fenêtres temporelles. Le calcul se fait de la manière suivante :
- i. Considérant une séquence temporelle $x(t)$ (dans notre cas, c'est la séquence des intervalles de temps du pas) et posant $\langle x \rangle$, la moyenne de $x(t)$, calculer de la somme cumulative :

$$X_t = \sum_{i=1}^t x_i - \langle x \rangle \quad (5.2)$$

- ii. Pour n différentes longueurs de fenêtres ($10 \leq n \leq 20$) :
- A. Avec la méthode des moindres carrés, calculer y_t , une fonction modélisant $X(t)$ dans chaque segment délimité par la fenêtre.
- B. En posant $Y(t)$, la fonction composée des différentes lignes y_t pour modéliser le signal $X(t)$, calculer la fluctuation $F(n)$:

$$F(n) = \sqrt{\frac{1}{N} \sum_{i=1}^t (X_i - Y_t)^2} \quad (5.3)$$

- C. Déterminer le coefficient α tel que $F(n) \propto n^\alpha$. Pour ce faire, construire un graphique *log-log* de $F(n)$ en fonction n , estimer ensuite la pente α de la droite modélisant la fonction $F(n)$.
- D. Utiliser le coefficient α comme caractéristique de la dynamique des fluctuations.
- (b) *Calcul de l'autocorrélation $R(\tau)$* : $R(\tau)$ permet d'estimer la corrélation d'une séquence temporelle avec elle-même avec différents décalages temporels (τ) :

$$R(\tau) = E[X(t) * X(t - \tau)] \quad (5.4)$$

Pour un signal non périodique, comme les données spatio-temporelles de la démarche, plus le décalage de temps devient grand, plus l'autocorrélation diminue. Le τ auquel l'autocorrélation diminue à 0,63 a été choisi comme descripteur d'autocorrélation tel que suggéré par Hausdorff et al. [18]

Ces statistiques ont été calculées uniquement sur l'intervalle de temps du pas gauche. En effet, Hausdorff et al. [18] ont rapporté que les caractéristiques ci-dessus calculées sur les séquences de l'intervalle de temps du pas démontraient des différences significatives entre les quatre groupes. Ainsi, ils devraient être discriminatoires et faciliter la classification. En outre, vu le peu de données disponibles, l'ajout d'autres caractéristiques pourrait causer un surapprentissage.

5.3 Classification

Pour l'étape de classification, nous avons opté pour les méthodes d'apprentissage machine classiques plutôt que d'apprentissage profond vu le peu de données disponibles. Trois classificateurs ont été testés : la SVM, l'arbre décisionnel et la forêt d'arbres décisionnels. La SVM a été choisie en raison de sa grande capacité à classifier des données non linéairement séparables avec l'utilisation des noyaux. L'arbre décisionnel et la forêt d'arbres ont été choisis en raison de leur simplicité et facilité d'interprétation, surtout dans notre cas où nous avons uniquement 4 caractéristiques.

Quatre classifications binaires ont été réalisées : SLA vs contrôle, Huntington vs contrôle, Parkinson vs contrôle et Maladies neurodégénératives (SLA, Parkinson, Huntington) vs contrôle. Pour chaque classification, on a testé les trois classificateurs et une recherche des hyperparamètres a été effectuée pour chaque cas. Par exemple, la SVM optimale pour classifier SLA vs contrôle a des paramètres différents de celui du Parkinson vs contrôle. Les hyperparamètres optimisés sont décrits ci-dessous. Les configurations optimales sont présentées dans le tableau 5.1.

1. SVM :

- (a) Noyau : La fonction de base radiale (RBF) a été utilisée comme noyau pour toutes les SVM. Celle-ci permet d'avoir une plus grande capacité d'adaptation aux données non linéaires. Toutefois, elle peut poser un problème de surapprentissage.
- (b) Paramètre de régularisation C : Les valeurs testées sont entre 0.0001 et 100.

2. Arbre décisionnel : L'arbre utilisé est de type CART.

- (a) Nombre minimal de données par feuille (nFeuille) : [1, 2, 3]

(b) Nombre minimal de données pour diviser un noeud (nNoeuds) : [2, 3, 4, 5]

3. Forêt d'arbres :

(a) Nombre d'arbres (nArbres) : [10, 20, 30, 40, 50]

(b) Nombre minimal de données pour diviser un noeud (nNoeuds) : [2, 3, 4, 5]

Tableau 5.1 Hyperparamètres des classificateurs (Park : Parkinson, Hunt : Huntington, MND : Maladies neurodégénératives (Hunt, Park et SLA), ctrl : contrôle).

	SVM	Arbre décisionnel		Forêt d'arbres	
Classification	C	nFeuilles	nNoeuds	nArbres	nNoeuds
SLA vs. ctrl	0.001	1	3	20	2
Park vs ctrl	0.001	3	3	50	2
Hunt vs ctrl	0.1	3	3	50	2
MND vs ctrl	0.1	1	2	20	4

Les algorithmes ont été testés par une validation croisée de type *leave-one-out cross-validation*. Donc, à chaque fois, l'algorithme était entraîné avec $n - 1$ sujets, et testé sur le sujet restant. Ensuite, les prédictions de tous les patients (lorsque testés) ont été recueillies pour calculer la spécificité, la sensibilité et l'exactitude. Les résultats sont présentés dans les tableaux ci-dessous. Le groupe contrôle est le groupe négatif (N), et les patients atteints de la maladie testée constituent le groupe positif (P).

Pour fin de comparaison, nous rapportons les résultats de Xia et al. [47], car ces derniers avaient obtenu les meilleurs résultats à ce jour. Comme entrée, ils ont calculé la moyenne, l'écart-type, le maximum, le minimum, l'asymétrie et le coefficient d'acuité sur les intervalles de pas, la phase d'appui, et le double support. Les meilleurs résultats ont été obtenus avec une SVM.

5.4 Résultats et discussion

Tableau 5.2 SLA vs contrôle : Les meilleurs résultats (en gras) ont été obtenus avec la SVM (VF/FP : Vrais/Faux positifs, Sp : Spécificité, Se : Sensibilité, Ex : Exactitude, n(P) : # sujets atteints de SLA, n(N) : # sujets contrôle).

Algorithme	VP (n(P)=13)	FP (n(N)=16)	Sp (%)	Se(%)	Ex(%)
SVM	13	0	100	100	100
Arbre décisionnel	10	1	94	91	86.3
Forêt d'arbres	10	1	94	91	86.3
Xia et al. [47]	-	-	94	100	96.5

Tableau 5.3 Parkinson vs contrôle : Les meilleurs résultats sont en gras, nos meilleurs résultats ont été obtenus avec l'arbre décisionnel (VF/FP : Vrais/Faux positifs, Sp : Spécificité, Se : Sensibilité, Ex : Exactitude, n(P) : # sujets atteints de Parkinson, n(N) : # sujets contrôle).

Algorithme	VP (n(P)=15)	FP (n(N)=16)	Sp (%)	Se (%)	Ex (%)
SVM	15	3	81	100	90.3
Arbre décisionnel	14	1	94	100	96.7
Forêt d'arbres	15	1	94	93	93.5
Xia et al. [47]	15	0	100	100	100

Tableau 5.4 Huntington vs Contrôle : Nos meilleurs résultats (en gras) ont été obtenus avec la SVM (VF/FP : Vrais/Faux positifs, Sp : Spécificité, Se : Sensibilité, Ex : Exactitude, n(P) : # sujets atteints de Huntington, n(N) : # sujets contrôle).

Algorithme	VP (n(P)=20)	FP (n(N)=16)	Sp (%)	Se (%)	Ex (%)
SVM	20	0	100	100	100
Arbre décisionnel	19	1	94	94	94.4
Forêt d'arbres	19	1	94	94	94.4
Xia et al. [47]	20	0	100	100	100

Tableau 5.5 Maladies neurodégénératives (Parkinson, Huntington, SLA) vs Contrôle : Les meilleurs résultats (en gras) ont été obtenus avec la SVM (VF/FP : Vrais/Faux positifs, Sp : Spécificité, Se : Sensibilité, Ex : Exactitude, n(P) : # sujets atteints de maladies neurodégénératives, n(N) : # sujets contrôle).

Algorithme	VP (n(P)=48)	FP (n(N)=16)	Sp (%)	Se (%)	Ex (%)
SVM	48	0	100	100	100
Arbre décisionnel	43	3	81	90	87.5
Forêt d'arbres	43	1	94	90	90.6
Xia et al. [47]	-	-	94	98	96.8

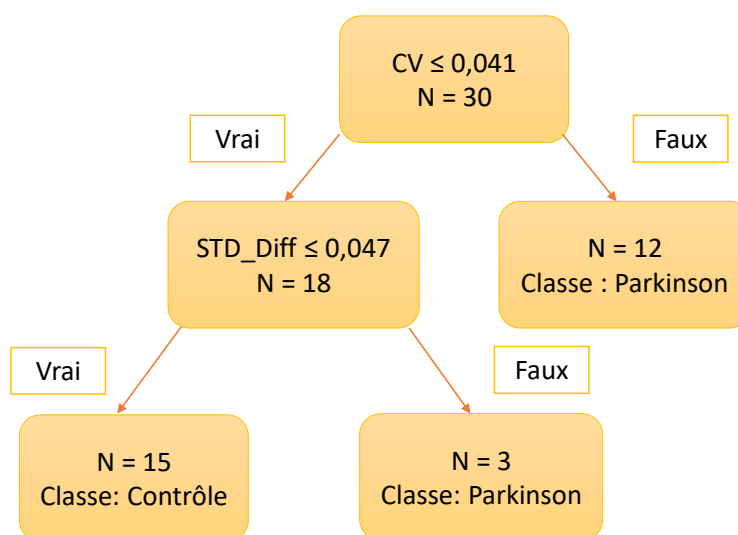


Figure 5.1 Exemple de l'arbre décisionnel après entraînement pour la classification du Parkinson. L'entraînement est fait avec 30 échantillons. L'échantillon restant est utilisé pour le test (sujet contrôle dans ce cas-ci).

À la lumière de ces résultats, nous constatons que la SVM a donné les meilleures performances pour toutes les classifications, à l'exception de la classification Parkinson vs Contrôle, avec des exactitudes de 100%. Ceci a été possible grâce à une optimisation rigoureuse du paramètre C. Les meilleures performances du SVM ont été obtenues avec de faibles valeurs de C (inférieures à 1). Toutefois, lorsque C était très petit (ex : 10^{-20}), l'exactitude diminuait à 75 %. Dans ce cas, ces erreurs étaient possiblement dues à un sous-apprentissage, car l'algorithme acceptait trop d'erreurs pour classifier ses données. D'un autre côté, lorsqu'on augmentait C, on augmente le risque de surentraînement. Par contre, les résultats étaient plus sensibles aux

grandes valeurs de C . Par exemple, dans la classification du Parkinson vs Contrôle, lorsque la valeur du C passait de 0.001 à 0.01, nous avons obtenu une exactitude de 55%. Donc, la régularisation est très importante pour éviter le surapprentissage. Ceci peut être expliqué par le noyau RBF utilisé qui a une grande capacité pour s'adapter aux données non-linéaires.

En ce qui concerne l'arbre de décision, le choix des hyperparamètres dépend du problème de classification. Toutefois, plusieurs combinaisons ont permis d'obtenir les résultats optimaux. L'arbre de décision a donné la meilleure performance (96.5%) pour la classification du Parkinson vs contrôle. Un exemple d'arbre après entraînement est illustré à la figure 5.1. Sa structure est assez simple, car il ne contient que deux niveaux. De plus, ce classificateur utilise uniquement les statistiques de l'amplitude des fluctuations. Nous pouvons ainsi conclure que ce sont ces principales caractéristiques qui permettent de mieux détecter la démarche parkinsonienne. Par conséquent, on peut comprendre que la SVM a eu une plus grande difficulté dans cette classification, car les statistiques de la dynamique des fluctuations (DFA et autocorrélation) ne sont pas assez discriminantes et perturbent la classification. En partant de cette constatation, nous avons donc refait la classification du Parkinson vs Contrôle avec la SVM en utilisant uniquement les statistiques d'amplitude de fluctuation, et nous avons obtenu un résultat parfait (exactitude de 100%). Ce test nous confirme que la démarche parkinsonienne se différencie plutôt par les amplitudes des fluctuations que par leur dynamique. La forêt d'arbres a donné des résultats très semblables à l'arbre décisionnel. Ceci était attendu, car le nombre de caractéristiques disponibles est très restreint (quatre caractéristiques), ce qui limite l'aspect aléatoire de ce modèle.

Les résultats démontrent qu'il est possible de classer les maladies neurodégénératives en utilisant des statistiques simples, mais significatives, pour caractériser les démarches des patients. Notre algorithme atteint des précisions meilleures que celles des travaux précédents [47]. De plus, les caractéristiques que nous utilisons ont préalablement démontré une différence significative pour distinguer la démarche parkinsonienne de la démarche normale. En outre, nos résultats ont le potentiel d'aider à déterminer quelles caractéristiques de démarche permettent de mieux distinguer les maladies neurodégénératives. Toutefois, plus de données seraient nécessaires afin de valider ces résultats.

CHAPITRE 6 AUTRES PISTES EXPLORÉES

Ce chapitre présente les autres méthodes que nous avons essayées pendant le développement de nos algorithmes. Dans un premier temps, nous allons discuter des différentes méthodes qui ont été appliquées à l’algorithme de détection de la maladie de Parkinson, mais qui n’ont pas donné les résultats souhaités. Dans un deuxième temps, nous allons analyser quelques limitations rencontrées pour des algorithmes de classification des maladies neurodégénératives.

6.1 Étude du Parkinson

Le réseau de neurones présenté dans l’article a été obtenu après de multiples essais. Plusieurs configurations n’ont pas réussi à discriminer entre les deux groupes à cause de l’importance du bruit dans les signaux et des différences interindividuelles. Parmi les modèles testés, on retrouve les réseaux de neurones récurrents et le transfert d’apprentissage avec le réseau de convolutions ResNet-50 préentraîné.

6.1.1 Réseaux de neurones récurrents

Les réseaux de neurones récurrents sont très souvent utilisés pour analyser les données séquentielles et temporelles. En effet, une cellule de LSTM est optimisée pour extraire et sélectionner les caractéristiques importantes à long et court terme. Pour cette raison, des couches de LSTM ont été introduites initialement dans le réseau. Celles-ci étaient placées à la fin des 1D-CNNs. Cet emplacement permettait dans un premier temps de filtrer les données avec les couches de convolution, et d’extraire les caractéristiques temporelles dans un deuxième temps. Toutefois, ceci a beaucoup ralenti l’entraînement, sans pour autant améliorer les résultats. Ceci pourrait être expliqué par le manque de structures apparentes dans nos données. En effet, le LSTM a montré des résultats remarquables dans l’interprétation des textes [41], car les patrons sont clairs. Cependant, dans notre cas, l’importance du bruit dans les données et les différences interindividuelles rendent l’apprentissage du LSTM difficile.

6.1.2 ResNet

Nous avons également exploré l’utilisation des réseaux de convolutions 2D préentraînés comme ResNet-50 [22]. En effet, les réseaux préentraînés sur ImageNet ont montré une capacité pour traiter des nouveaux problèmes. Par exemple, en médecine, plusieurs études ont utilisé les réseaux préentraînés pour analyser des images médicales [21], même si ce type d’images ne

fait pas partie de la collection ImageNet [10]. Dans notre cas, les données d'entrée sont les matrices (18 signaux VGRF \times 100 unités de temps). Pour réaliser un transfert d'apprentissage, nous avons modifié les dernières couches pour adapter l'algorithme à notre problème de classification. Nous avons également essayé différentes profondeurs (les n premières couches de convolutions) du réseau, ceci a pour but d'adapter la taille du Resnet-50 à notre problème. En effet, les premières convolutions concernent des traits généraux, tandis que les dernières convolutions concernent des traits spécifiques aux catégories des images d'entraînement (ImageNet). Elles sont donc peu pertinentes pour notre problème de classification.

Toutefois, cette approche n'a pas donné des bons résultats. Cela peut être expliqué par le fait que les motifs présents dans nos données de démarche sont très différents de ceux des images de ImageNet.

6.1.3 Augmentation des données

Nous avons également essayé des méthodes d'augmentation de données pour la classification de la maladie de Parkinson. Dans un premier temps, nous avons essayé d'ajouter du bruit gaussien aux données. Dans un deuxième temps, nous avons essayé de construire des nouveaux signaux en interpolant deux signaux de la même classe. Cependant, ces méthodes n'ont pas amélioré la performance de l'algorithme. En effet, le bruit est une composante importante des données, le fait de rajouter du bruit peut modifier les caractéristiques intrinsèques du signal. De plus, il est fort possible que les données artificiellement ajoutées ne ressemblent aucunement à des patrons possiblement présents dans l'ensemble test. L'augmentation des données s'est avérée ainsi inutile.

6.2 Classification des maladies neurodégénératives

La détection des maladies neurodégénératives a également été obtenue après l'exploration de plusieurs méthodes. Dans un premier temps, nous allons discuter de l'utilisation des méthodes d'apprentissage profond. Ensuite, nous allons aborder la performance de l'algorithme dans un contexte de multiclassification.

6.2.1 Apprentissage profond

Nous avons essayé l'utilisation de l'apprentissage profond pour la classification des maladies neurodégénératives. Nous avons testé des réseaux de neurones de neurones récurrents et des Convnets. Ces algorithmes n'ont pas donné des bonnes performances en raison de la petite taille de la base de données.

6.2.2 Multiclassification

Dans le travail présenté, nous avons travaillé sur le développement de méthodes de classification binaire. En effet, dans un contexte de multiclassification, notre algorithme était incapable de différencier les trois maladies en même temps. Nous pensons que l'ajout d'autres attributs permettra une meilleure différenciation entre les maladies. Nous pensons également que le problème de multiclassification nécessite plus de données.

CHAPITRE 7 DISCUSSION GÉNÉRALE

Ce chapitre présente une discussion générale sur l'ensemble des travaux effectués. D'abord, nous allons évaluer l'atteinte des objectifs. Ensuite, nous allons aborder une réflexion sur les méthodes utilisées. Finalement, nous allons terminer par une discussion sur les méthodes d'évaluation.

7.1 Atteinte des objectifs

L'objectif principal de ce travail est le développement d'algorithmes intelligents pour le diagnostic des maladies neurodégénératives à partir de l'analyse de démarche. Dans un premier temps, nous avons travaillé sur le développement d'un algorithme de détection de la maladie de Parkinson. À la lumière de nos résultats, nous concluons que cet objectif est atteint. L'algorithme présenté dans l'article soumis a démontré une capacité de généraliser à des nouveaux sujets avec une précision de 90%. Il est clair qu'il y a encore place à l'amélioration, mais nous jugeons que c'est un score qui répond entièrement à notre objectif. Dans un deuxième temps, nous visions le développement d'algorithmes de détection des maladies neurodégénératives du Huntington, Parkinson et SLA. Pour chaque maladie, nous avons réussi à atteindre une exactitude de 100%. Nous concluons donc que cet objectif a également été atteint. De plus, nos deux approches ont surpassé les résultats des méthodes de l'état de l'art. Nous avons ainsi atteint notre objectif principal.

7.2 Approches utilisées

Les deux problèmes que nous avons abordés se différencient principalement au niveau de la taille des bases de données. Il s'agit d'un facteur important dans le choix de chaque approche. En effet, lorsqu'une grande base de données est disponible, comme dans la démarche parkinsonienne, celle-ci peut être rentabilisée avec l'utilisation d'un réseau de neurones profond, qui va lui-même extraire les caractéristiques discriminatoires des données brutes. Dans le cas des maladies neurodégénératives, la base de données était plus restreinte. Par conséquent, l'étape d'ingénierie des variables devient importante, car elle permet de faciliter la tâche des classificateurs. L'apprentissage machine classique est également plus adapté pour ce cas, puisque moins de données sont nécessaires pour entraîner les classificateurs. Toutefois, plus la base de données est grande, plus le modèle développé sera robuste et plus facilement généralisable à d'autres sujets. Donc, il est évident que le modèle d'apprentissage profond développé avec

le Parkinson est plus robuste, car il a été entraîné et validé avec une plus grande base de données.

7.3 Méthodes d'évaluation

Le choix de la méthode d'évaluation du classificateur est très important pour la rigueur scientifique de notre travail. Pour nos deux axes de travaux, nous avons choisi des méthodes d'évaluation différentes. Pour la détection de la maladie de Parkinson, nous avons divisé nos données en ensembles d'entraînement, de validation et de test. C'est ce qui est généralement recommandé en apprentissage profond. Ceci permet de s'assurer de la généralisation des hyper-paramètres pour un ensemble de test inconnu par l'algorithme pendant son développement. Toutefois, lorsque les bases de données sont plus petites, cette division est plutôt difficile, car chaque donnée est précieuse pour l'entraînement. En ce qui concerne la classification des maladies neurodégénératives, nous avons opté pour une validation croisée de type *leave-one-out cross-validation*. Dans ce cas-ci, nous n'avons pas procédé à une division de trois ensembles : apprentissage, validation et test, car la base de données était trop petite pour être compartimentée en trois parties. L'utilisation de la validation croisée nous a permis de maximiser les données d'entraînement, tout en étant capables de tester notre algorithme sur des données qui ne sont pas incluses dans l'apprentissage. Toutefois, la généralisation des hyperparamètres n'a pas été testée.

CHAPITRE 8 CONCLUSION

L'objectif de ce projet était de développer des algorithmes de détection des maladies neurodégénératives à partir de l'analyse de la démarche. Ces algorithmes peuvent aider les médecins dans leur examen clinique pour diagnostiquer la présence d'une maladie neurodégénérative. Nous avons fait un survol des algorithmes proposés par les travaux antérieurs. Ces algorithmes se divisent généralement en deux étapes, une première étape d'extraction des caractéristiques et une deuxième étape de classification. Nous avons également présenté les méthodes de classification qui ont inspiré nos travaux : l'apprentissage machine classique et l'apprentissage profond.

Dans un premier temps, nous avons mis au point un algorithme de détection de la maladie de Parkinson basé sur l'apprentissage profond. Les données d'entrée étaient constituées de signaux bruts provenant de capteurs de VGRF placés sous les pieds. Le classificateur utilisé est un réseau de neurones profond composé de 18 1D-Convnets parallèles traitant chacun un signal VGRF. Ces 1D-Convnets sont suivis par un réseau de neurones pleinement connecté qui prend en entrée les caractéristiques extraites par les 1D-Convnets et donne en sortie la classification finale. Cet algorithme a permis de détecter les sujets atteints de la maladie de Parkinson à partir des données de démarche avec une précision de 90%, tout en surpassant des méthodes récentes de la littérature. De plus, le modèle proposé a permis d'éliminer l'étape d'extraction manuelle des caractéristiques. En effet, le réseau de neurones profond effectue cette étape implicitement lors de la phase d'apprentissage.

Dans un deuxième temps, nous avons développé des algorithmes de détection du Parkinson, Huntington et SLA en utilisant les données spatio-temporelles. Puisque la base de données était restreinte, nous avons opté pour les classificateurs plus classiques. La première étape consistait à construire un vecteur de caractéristiques. Nous avons choisi des caractéristiques de l'amplitude et de la dynamique des fluctuations des intervalles de pas. La deuxième étape de l'algorithme consistait à entraîner trois classificateurs : SVM, l'arbre décisionnel et la forêt d'arbres. Le SVM a démontré la meilleure performance en réussissant à classifier tous les sujets (exactitude 100%). Cet algorithme se distingue par sa simplicité et sa capacité d'apprendre à partir d'une base de données relativement petite. En effet, la principale limitation du problème est la taille des bases de données. Une plus grande base de données aurait pu être capitalisée par le développement d'un algorithme d'apprentissage profond. Il s'agit en effet d'un problème récurrent avec les données médicales. Pour aller plus loin, il serait intéressant de développer des algorithmes qui prédisent le stade d'avancement des maladies neurodégénératives chez

les patients à partir de leur démarche.

RÉFÉRENCES

- [1] Support vector machines. <https://scikit-learn.org/stable/modules/svm.html>, 2018.
- [2] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam. Automated detection of arrhythmias using different intervals of tachycardia eeg segments with convolutional neural network. *Information sciences*, 405 :81–90, 2017.
- [3] A. Ascherio, H. Chen, M. G. Weisskopf, E. O’Reilly, M. L. McCullough, E. E. Calle, M. A. Schwarzschild, and M. J. Thun. Pesticide exposure and risk for parkinson’s disease. *Annals of Neurology : Official Journal of the American Neurological Association and the Child Neurology Society*, 60(2) :197–203, 2006.
- [4] A. Bhattacharyya and R. B. Pachori. A multivariate approach for patient-specific eeg seizure detection using empirical wavelet transform. *IEEE Transactions on Biomedical Engineering*, 64(9) :2003–2015, 2017.
- [5] L. Breiman. *Classification and regression trees*. Routledge, 2017.
- [6] M. R. Daliri. Chi-square distance kernel of the gaits for the diagnosis of parkinson’s disease. *Biomedical Signal Processing and Control*, 8(1) :66–70, 2013.
- [7] P. Dayalu and R. L. Albin. Huntington disease : pathogenesis and treatment. *Neurologic clinics*, 33(1) :101–114, 2015.
- [8] M. de Carvalho, R. Dengler, A. Eisen, J. D. England, R. Kaji, J. Kimura, K. Mills, H. Mitsumoto, H. Nodera, J. Shefner, et al. Electrodiagnostic criteria for diagnosis of als. *Clinical neurophysiology*, 119(3) :497–503, 2008.
- [9] L. M. De Lau and M. M. Breteler. Epidemiology of parkinson’s disease. *The Lancet Neurology*, 5(6) :525–535, 2006.
- [10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet : A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.
- [11] Ö. F. Ertuğrul, Y. Kaya, R. Tekin, and M. N. Almalı. Detection of parkinson’s disease by shifted one dimensional local binary patterns from gait. *Expert Systems with Applications*, 56 :156–163, 2016.
- [12] S. Frenkel-Toledo, N. Giladi, C. Peretz, T. Herman, L. Gruendlinger, and J. M. Hausdorff. Effect of gait speed on gait rhythmicity in parkinson’s disease : variability of stride time and swing time respond differently. *Journal of neuroengineering and rehabilitation*, 2(1) :23, 2005.

- [13] S. Frenkel-Toledo, N. Giladi, C. Peretz, T. Herman, L. Gruendlinger, and J. M. Hausdorff. Treadmill walking as an external pacemaker to improve gait rhythm and stability in parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society*, 20(9) :1109–1114, 2005.
- [14] M. Gabel, R. Gilad-Bachrach, E. Renshaw, and A. Schuster. Full body gait analysis with kinect. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 1964–1967. IEEE, 2012.
- [15] J. Gilles et al. Empirical wavelet transform. *IEEE Trans. Signal Processing*, 61(16) :3999–4010, 2013.
- [16] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet : Components of a new research resource for complex physiologic signals. *Circulation*, 101(23) :e215–e220, 2000 (June 13). Circulation Electronic Pages : <http://circ.ahajournals.org/content/101/23/e215.full> PMID :1085218 ; doi : 10.1161/01.CIR.101.23.e215.
- [17] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- [18] J. M. Hausdorff, A. Lertratanakul, M. E. Cudkowicz, A. L. Peterson, D. Kaliton, and A. L. Goldberger. Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. *Journal of applied physiology*, 88(6) :2045–2053, 2000.
- [19] J. M. Hausdorff, J. Lowenthal, T. Herman, L. Gruendlinger, C. Peretz, and N. Giladi. Rhythmic auditory stimulation modulates gait variability in parkinson's disease. *European Journal of Neuroscience*, 26(8) :2369–2375, 2007.
- [20] J. M. Hausdorff, C. Peng, Z. Ladin, J. Y. Wei, and A. L. Goldberger. Is walking a random walk? evidence for long-range correlations in stride interval of human gait. *Journal of Applied Physiology*, 78(1) :349–358, 1995.
- [21] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle. Brain tumor segmentation with deep neural networks. *Medical image analysis*, 35 :18–31, 2017.
- [22] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [23] J. Jankovic. Parkinson's disease : clinical features and diagnosis. *Journal of neurology, neurosurgery & psychiatry*, 79(4) :368–376, 2008.

- [24] D. Joshi, A. Khajuria, and P. Joshi. An automatic non-invasive method for parkinson's disease classification. *Computer methods and programs in biomedicine*, 145 :135–145, 2017.
- [25] M. C. Kiernan, S. Vucic, B. C. Cheah, M. R. Turner, A. Eisen, O. Hardiman, J. R. Burrell, and M. C. Zoing. Amyotrophic lateral sclerosis. *The lancet*, 377(9769) :942–955, 2011.
- [26] A. Liaw, M. Wiener, et al. Classification and regression by randomforest. *R news*, 2(3) :18–22, 2002.
- [27] A. Mannini, D. Trojaniello, A. Cereatti, and A. M. Sabatini. A machine learning framework for gait classification using inertial sensors : Application to elderly, post-stroke and huntington's disease patients. *Sensors*, 16(1) :134, 2016.
- [28] X. Mao, M. T. Ou, S. S. Karuppagounder, T.-I. Kam, X. Yin, Y. Xiong, P. Ge, G. E. Umanah, S. Brahmachari, J.-H. Shin, et al. Pathological α -synuclein transmission initiated by binding lymphocyte-activation gene 3. *Science*, 353(6307) :aah3374, 2016.
- [29] D. A. Neumann. Kinesiology of walking. *Kinesiology of the Musculoskeletal System, St. Louis, Missouri : Mosby*, pages 523–69, 2002.
- [30] S. J. Pan, Q. Yang, et al. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10) :1345–1359, 2010.
- [31] S. V. Perumal and R. Sankar. Gait and tremor assessment for patients with parkinson's disease using wearable sensors. *ICT Express*, 2(4) :168–174, 2016.
- [32] Physionet. Gait dynamics in neuro-degenerative disease data base. <https://www.physionet.org/physiobank/database/gaitnidd/>, 2018.
- [33] S. M. Pincus. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, 88(6) :2297–2301, 1991.
- [34] J. R. Quinlan. Induction of decision trees. *Machine learning*, 1(1) :81–106, 1986.
- [35] J. R. Quinlan. C4. 5 : Programming for machine learning. *Morgan Kaufmann*, 38 :48, 1993.
- [36] S. R. Safavian and D. Landgrebe. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, 21(3) :660–674, 1991.
- [37] A. Samii, J. G. Nutt, and B. R. Ransom. Parkinson's disease. *The Lancet*, 363(9423) :1783 – 1793, 2004.
- [38] Y. Sarbaz, F. Towhidkhah, S. Gharibzadeh, and A. Jafari. Gait spectral analysis : an easy fast quantitative method for diagnosing parkinson's disease. *Journal of Mechanics in Medicine and Biology*, 12(03) :1250041, 2012.

- [39] O. Sofuwa, A. Nieuwboer, K. Desloovere, A.-M. Willems, F. Chavret, and I. Jonkers. Quantitative gait analysis in parkinson's disease : comparison with a healthy control group. *Archives of physical medicine and rehabilitation*, 86(5) :1007–1013, 2005.
- [40] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout : a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1) :1929–1958, 2014.
- [41] M. Sundermeyer, R. Schlüter, and H. Ney. Lstm neural networks for language modeling. In *Thirteenth annual conference of the international speech communication association*, 2012.
- [42] A. Temko, E. Thomas, W. Marnane, G. Lightbody, and G. Boylan. Eeg-based neonatal seizure detection with support vector machines. *Clinical Neurophysiology*, 122(3) :464–473, 2011.
- [43] V. Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013.
- [44] F. O. Walker. Huntington's disease. *The Lancet*, 369(9557) :218–228, 2007.
- [45] X. Wang. *Machine Learning for Gait Classification*. PhD thesis, Universität Bremen, 2017.
- [46] Y. Wu, P. Chen, X. Luo, M. Wu, L. Liao, S. Yang, and R. M. Rangayyan. Measuring signal fluctuations in gait rhythm time series of patients with parkinson's disease using entropy parameters. *Biomedical Signal Processing and Control*, 31 :265–271, 2017.
- [47] Y. Xia, Q. Gao, and Q. Ye. Classification of gait rhythm signals between patients with neuro-degenerative diseases and normal subjects : Experiments with statistical features and different classification models. *Biomedical Signal Processing and Control*, 18 :254–262, 2015.
- [48] G. Yogev, N. Giladi, C. Peretz, S. Springer, E. S. Simon, and J. M. Hausdorff. Dual tasking, gait rhythmicity, and parkinson's disease : which aspects of gait are attention demanding ? *European journal of neuroscience*, 22(5) :1248–1256, 2005.
- [49] S. Zarei, K. Carr, L. Reiley, K. Diaz, O. Guerra, P. F. Altamirano, W. Pagani, D. Lodin, G. Orozco, and A. Chinae. A comprehensive review of amyotrophic lateral sclerosis. *Surgical neurology international*, 6, 2015.
- [50] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.
- [51] A. Zhao, L. Qi, J. Li, J. Dong, and H. Yu. A hybrid spatio-temporal model for detection and severity rating of parkinson's disease from gait data. *Neurocomputing*, 2018.