Automatic detection of daily living activities in people with Parkinson's disease using kinematic-driven data

Luís Filipe Simões de Abreu

October 2022

Luís Filipe Simões de Abreu

Automatic detection of daily living activities in people with Parkinson's disease using kinematic-driven data

October 2022

Universidade do Minho
Escola de Engenharia

Luís Filipe Simões de Abreu

Automatic detection of daily living activities in people with Parkinson's disease using kinematic-driven data

Individual Project
Integrated Master's in Informatics Engineering

Dissertation supervised by
Professora Doutora Cristina Manuela Peixoto Santos

October 2022

DECLARATION

Name: Luís Filipe Simões de Abreu

Dissertation Title: Automatic detection of daily living activities in people with Parkinson's disease using kinematic-driven data

Mentors: Cristina Manuela Peixoto Santos

Conclusion Year: 2022

Master Designation: Mestrado Integrado em Engenharia Informática

Master Branch: Machine Learning e Ciência de dados

I declare that I grant to the University of Minho and its agents a non-exclusive license to file and make available through its repository, in the conditions indicated below, my dissertation, as a whole or partially, in digital support.

I declare that I authorize the University of Minho to file more than one copy of the dissertation and, without altering its contents, to convert the dissertation to any format or support, for the purpose of preservation and access.

Furthermore, I retain all copyrights related to the dissertation and the right to use it in future works.

I authorize the partial reproduction of this dissertation for the purpose of investigation by means of a written declaration of the interested person or entity.

This is an academic work that can be used by third parties if internationally accepted rules and good practice with regard to copyright and related rights are respected.

Thus, the present work can be used under the terms of the license indicated below.

In case the user needs permission to be able to make use of the work in conditions not foreseen in the indicated licensing, he should contact the author through the RepositóriUM of the University of Minho.

Universidade do Minho, _____/_____/_____

Signature: _____

Acknowledgements

I want to start by thanking my parents and sister for all support they have given me all my life. They are my role models, and I hope one day to be as great as they are.

I want to thank Henrique Faria and João Marques, for being top notch “artists” and for becoming my friends for a lifetime.

To Cecília Soares and Alfranio Correia Júnior, I learned to program with you in the multiple evenings before deadline, thank you for your incredible friendship, company, sympathy, optimism, and encouragement. I can’t forget your four-legged son, Dennis, his energy, friendship, and comradeship is awesome, and he was always there throughout our journey. I really love throwing him the Duck and petting him.

I also want to highlight the incredible importance that Rafaela Castro has had in this journey of mine, her patience, tenderness, and fierce spirit inspired me and led me to overcome all the challenges. With you by my side everything is possible.

I would also like to thank Professor Cristina Santos and Helena Gonçalves for guiding me through this challenge that is the master's thesis, your indications and suggestions were crucial and always right. To the Birdies of BiRDLab, your constant good mood and hardworking spirit is contagious and unforgettable, thank you for always answering all my questions and doubts.

Many more people remain to be mentioned, but all of them were, are and will always be important in my journey through life. I always have you in my heart, thank you for everything.

Finally, I want to thank to Services and Advanced Research Computing with HTC/HPC clusters (SeARCH) for enabling computational power for all the deep learning models building stages.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

Universidade do Minho, _____/_____/_____

Signature: _____

ABSTRACT

Parkinson's Disease (PD) is a neurodegenerative disorder of the central nervous system. Resting tremor, akinesia, and bradykinesia (slow movements), rigidity, shuffling walking, and postural instability are some of the symptoms that not only negatively impacts patients' life, but also the life of people around them.

Current approaches for monitoring patients' motor autonomy are limited to the observer and self-reported methods. The observer-based examinations, patients perform a set of standard PD examinations. The self-reported method relies on patients' daily activities diaries. These approaches are commonly used, but are limited to a few sessions per year, they do not address common motor daily tasks, and their results are object of subjective interpretation by the clinical expert.

By combining kinematic-driven data from wearable sensor with AI, the main goal of this dissertation is to develop an automatic software for recognition of human activities (e.g., walking, standing, turning, sitting, and lying) in PD to assist the clinical experts with objective and concrete data. A data collection protocol was developed and captured, resulting in a database comprised of data collected from eighteen PD patients who performed three trials of six different daily activities: *walk*; *180º turning*; *sit on chair*; *get up from chair*; *lay on bed* and *get up from bed*.

A Deep Learning (DL) framework based on Convolutional Neural Network capable of recognizing daily activities was developed and attained a performance of F1 Score equal to 0.90892.

As a complementary goal an automatic software for human walk initial contact (IC) and final contact (FC) recognition using kinematic data was also developed. IC and FC are tremendously important to provide patient on-demand motor assistance and estimation of walking-associated metrics.

A Deep Learning framework based on Bidirectional Long Short-Term Memory Neural Network capable of walking IC/FC events detection was developed and attained a performance of MCC Score equal to 0.538386.

Promising results were attained for both DL frameworks, however, this dissertation suggests that there is still room for further improvements. Enriching the dataset with more data from different patient, data balancing and feature extraction techniques, experimenting new models' architectures should be considered in future works.

Keywords: Parkinson's Disease, Deep Learning, Human Activity Recognition

RESUMO

A Doença de Parkinson (DP) é uma doença neurodegenerativa do sistema nervoso central. O tremor em repouso, acinesia e bradicinesia (movimentos lentos), rigidez, marcha e postura instável são alguns dos sintomas que afetam negativamente a vida dos pacientes e também as pessoas à sua volta.

As monitorizações da autonomia motora dos pacientes estão limitadas aos métodos presenciais e auto-relatados. Em exames presenciais, os pacientes realizam um conjunto de exames padrão de DP. O método auto-relatado baseia-se nas agendas de atividades diárias dos pacientes. Estas abordagens são comuns, mas são limitadas a algumas sessões por ano, não abordam tarefas motoras diárias comuns, e os seus resultados dependem da interpretação subjetiva do perito clínico.

O principal objetivo desta dissertação é desenvolver um software automático para e reconhecimento de atividades humanas (por exemplo, andar, estar em pé, virar, sentar e deitar) na DP, que combine dados cinemáticos de sensores vestíveis com inteligência artificial para ajudar os especialistas clínicos a obterem dados objetivos e concretos. Foi desenvolvido um protocolo de recolha de dados, resultando numa base de dados constituída por dados recolhidos de dezoito pacientes de DP que realizaram três ensaios de seis atividades diárias diferentes: *caminhar*; *virar 180ª*; *sentar-se na cadeira*; *levantar-se da cadeira*; *deitar-se na cama* e *levantar-se da cama*.

Foi desenvolvida uma estrutura de Deep Learning (DL) baseada em Convolutional Neural Network capaz de reconhecer as atividades diárias e atingir um desempenho de F1 score igual a 0,9089.

Como objetivo complementar, foi também desenvolvido um software automático para o reconhecimento do contacto inicial (CI) e final (FC) do andar humano, utilizando dados cinemáticos. O CI e a CF são tremendamente importantes para fornecer assistência motora em tempo real e estimativa da métrica associada à marcha do paciente.

Foi desenvolvida uma estrutura de DL baseada em Bidirectional Long Short-Term Memory Neural network capaz de detetar eventos de IC/FC durante o andar e atingir um desempenho MCC score igual a 0,5384.

Foram alcançados resultados promissores para ambas as estruturas DL, contudo, esta dissertação sugere que ainda há espaço para mais melhorias. Enriquecer o conjunto de dados com mais dados de diferentes pacientes, técnicas de balanceamento de dados e extração de características, experimentar modelos com diferentes arquiteturas deve ser considerado em trabalhos futuros.

Palavras-Chave: Doença de Parkinson's, Deep Learning, Reconhecimento de Atividade Humana

Table of Contents

1 Introduction 1

1.1 Motivation and problem statement 2

1.2 Goals and research questions 4

1.3 Contribution to knowledge 6

1.4 Dissertation structure 6

2 Literature Review 8

2.1 Introduction 9

2.2 Methods 10

2.2.1 Data sources, search strategy and studies selection 10

2.3 Results 11

2.3.1 General results 11

2.3.2 Daily motor activities 12

2.3.3 Data input 13

2.3.4 AI-based models 18

2.4 Discussion 26

2.4.1 (I) Which daily living activities are recognized and for what purpose? 26

2.4.2 (II) Which type of data was used in AI models input for human activities recognition systems? 26

2.4.3 (III) Which AI models and how they were implemented in human activities recognition systems? 28

2.5 Conclusion and future directions 29

3 Solution Overview 31

3.1 Introduction 32

3.2 Proposed solution framed into +Sense 34

3.2.1 +sMotion: Combination of Kinematic-drive Data from Wearable Sensors with AI for motion assessment in PD 35

3.3 Conclusions 36

4 Deep learning Framework for Human Activities Recognition 38

4.1 Introduction 39

4.2 Data Input 39

4.2.1 Data acquisition 39
4.2.2 Dataset description 42
4.2.3 Data input preparation 45
4.3 Model Pipeline 49
4.3.1 Convolutional Neural Network 50
4.3.2 Proposed Convolutional Neural Network Model 51
4.3.3 Input and Output data format 52
4.3.4 Model Evaluation Metric 53
4.3.5 Fine-tunning Stage 55
4.3.6 Ablation Model Index 56
4.4 Results 56
4.4.1 Fine-tuning Stage Performance 56
4.4.2 Model Training and Validation Performance 57
4.4.3 Test Performance 58
4.4.4 Ablation Index Performance 59
4.5 Discussion 60
4.5.1 Proposed DL HAR Model Performance 60
4.5.2 Ablation Index Performance 62
4.6 Conclusion 62
5 Deep-Learning Framework for Human Walk IC/FC Detection 64
5.1 Introduction 65
5.2 Related Works: brief State of the Art 65
5.2.1 Conclusion 68
5.3 Data Input 69
5.3.1 Data acquisition 69
5.3.2 Dataset description 69
5.3.3 Data Input preparation 71
5.4 Model pipeline 74
5.4.1 Bidirectional Long Short-Term Memory 74
5.4.2 Proposed Bidirectional Long Short-Term Memory Model 76
5.4.3 Input and Output Data Format 77
5.4.4 Model Evaluation Metric 78
5.4.5 Fine-tunning Stage 79
5.5 Results 79

5.5.1 Fine-tuning Stage Performance 79
5.5.2 Model Training and Validation Performance 80
5.5.3 Test Performance 81
5.6 Discussion 82
5.7 Conclusion 84
6 Conclusion 85
7 References 89
Appendix A 95
Appendix B 102

LIST OF FIGURES

Figure 2-1 Flowchart for the search strategy based on Preferred Reporting Items for Systematic Reviews and Meta-Analyses(PRISMA). 11
Figure 3-1 +Sense project modular composition. 35
Figure 3-2 +Sense hardware and software setup. 35
Figure 3-3 Scheme about how this dissertation was framed in +Sense. Data collection using wearable motion LAB and AI analysis provided with this dissertation. 36
Figure 4-1 HAR schema layout solution. 39
Figure 4-2 Waistband containing the MPU-6050 Inertial Measurement Unit sensor (WBD)..... 41
Figure 4-3 Trials performed by the volunteers. A – Sit and get up from chair; B – Lay and get up from bed; C – Walking; D -Walking and perform 180º turn. 42
Figure 4-4 Comparison between Healthy person *walking* and PD patient *walking* activities. Only Acceleration values from X axis are displayed for better understanding. 43
Figure 4-5 Comparison between Healthy person *180º turning* and PD patient *180º turning* activities. Only Acceleration values from X axis are displayed for better understanding. 43
Figure 4-6 Comparison between Healthy person *sitting on chair* and PD patient *sitting on chair* activities. Only Acceleration values from X axis are displayed for better understanding..... 44
Figure 4-7 Comparison between Healthy person *get up from chair* and PD patient *get up from chair* activities. Only Acceleration values from X axis are displayed for better understanding..... 44
Figure 4-8 Comparison between Healthy person *lay on bed* and PD patient *lay on bed* activities. Only Acceleration values from X axis are displayed for better understanding. 44
Figure 4-9 Comparison between Healthy person *get up bed* and PD patient *get up bed* activities. Only Acceleration values from X axis are displayed for better understanding. 44
Figure 4-10 Human activities recognition data input handling pipeline. 45
Figure 4-11 Patient trial with corrupted captured kinematic data. 45
Figure 4-12 With a light blue color the PD patient *walk* kinematic data labelling and with a dark blue PD patient Standing still labelling kinematic data. Only Acceleration values from X axis are displayed for better understanding. 46
Figure 4-13 With red color the PD patient *180º turning* activities kinematic data labelling. With light blue color are labelled as *walking* activities and with dark blue color PD patient *standing still* kinematic data labelling. Only Acceleration values from X axis are displayed for better understanding. 46

Figure 4-14 With red color the PD patient *sitting on chair* activity kinematic data labelled. With dark blue color PD patient *standing still* activity kinematic data labelling. Only Acceleration values from X axis are displayed for better understanding. 47
Figure 4-15 With green color the *get up from chair* activity labelled. With dark blue color PD patient *standing still* activity labelling kinematic data. Only Acceleration values from X axis are displayed for better understanding. 47
Figure 4-16 With red color the *sitting on chair* activity labelled. Pink color represents the Lay on bed activity and dark blue color *standing still* activity labelled kinematic data. Only Acceleration values from X axis are displayed for better understanding. 47
Figure 4-17 With yellow color the labelled *get up from bed* activity. The green color represents the *get up from chair* activity and dark blue color *standing still* activity labelled kinematic data. Only Acceleration values from X axis are displayed for better understanding. 48
Figure 4-18 Schematic representation of the dataset splitting for the different HAR DL model development stages. 49
Figure 4-19 Model Pipeline Schema representing the development of the final HAR DL model for PD patients' solution. 50
Figure 4-20 Example of a basic CNN architecture. 50
Figure 4-21 CNN example of functioning through convolutions, down sampling with pooling and finally the classification of input data. 51
Figure 4-22 Proposed HAR DL model architecture. 52
Figure 4-23 F1-score value evolution for train and validation dataset. 57
Figure 4-24 Loss value evolution for train and validation dataset. 57
Figure 4-25 Confusion matrix results of the DL model trained for 100 epochs. 58
Figure 4-26 Confusion matrix test results, using Test dataset. 58
Figure 4-27 Confusion matrix result for architecture Ablation 1, Ablation 2 in upper row. Confusion matrix results for architectures Ablation 3 and 4 in lower row, respectively. 60
Figure 5-1 Human walk IC/FC detection framework schematic view. 65
Figure 5-2 Walking cycle and walking events. 66
Figure 5-3 Patients started the trial on (1), walked 10m and stopped at (2). 69
Figure 5-4 Patients started the trial on (1), walked 30m, turned right at (2), walk forward 2m (3), turned around (4), walking forward 2m (5), runed left at (6), made a 32m walk (7), tuner at left (8), walking 2m (9) until returning to the trial initial position. 69

Figure 5-5 Walking events occurrence during the trial plotted over the x-axis values of the acceleration based on [28]. 70
Figure 5-6 Human Walking IC/FC detection data input handling pipeline. 71
Figure 5-7 Correction to walking events timestamps. Green color circle representing a FC event timestamp was placed at green color dot timestamp. Orange color cross representing a IC event timestamp was placed at red color cross timestamp. 71
Figure 5-8 PD patient Trial labelled example. With red color timesteps labelled as IC. With green color timesteps labelled as FC. 73
Figure 5-9 Schematic representation of the dataset splitting for the different Human walk IC/FC detection model development stages. 73
Figure 5-10 Model Pipeline Schema representing the development of the final Human walking IC/FC DL model for PD patients' solution. 74
Figure 5-11 Architecture of a LSTM cell. 75
Figure 5-12 LSTM and BiLSTM architectures workflow. 75
Figure 5-13 Proposed Human Walk IC/FC DL model architecture. 76
Figure 5-14 Counting the Initial foot Contact sequence size. 77
Figure 5-15 Counting the Final foot Contact sequence size. 77
Figure 5-16 Matthews Correlation Coefficient formula. TN: True Negative; TP: True Positive; FN: False Negative; FP: False Positive. 79
Figure 5-17 MCC Score value evolution for train and validation dataset. 80
Figure 5-18 Loss value evolution for train and validation dataset. 80
Figure 5-19 Confusion matrix results of the DL model trained for 10 epochs. 81
Figure 5-20 Confusion matrix test results, using Test dataset. 81
Figure A-1 Study chronology (W: week). 96
Figure A-2 - Experimental procedure tasks for human motor-related activities: (A) Sit in a chair, (B) Get up from a char, (C) Lie on a bed, (D) Get up from a bed, (E) Walk/Stand, (F) Right/Left 180° turning. . 101

List of Tables

Table 2-1 Overview of the most represented studied activities. 12
Table 2-2 Sensors, data input and protocols observed in literature. 15
Table 2-3 Summary of machine learning models proposed by several authors, and their data processing and algorithms performances. 19
Table 2-4 - Limitations and guidelines to overcome them. 29
Table 3-1 Human activities to be recognized by our proposed DL classification framework. 32
Table 4-1 Demographic and clinic characteristics of study participants. 40
Table 4-2 Daily activities label identification. 48
Table 4-3 Mean, minimum, maximum length of each daily activity. 53
Table 4-4 Different size windows input experimented on. 53
Table 4-5 Confusion Matrix. 53
Table 4-6 Fine-tunning phase parameters explored. 55
Table 4-7 Ablation study for this dissertation proposed architecture. 56
Table 4-8 Fine-Tuning stage optimal parameters found, and the results obtained with them. 56
Table 4-9 Metrics evaluation for 100 epochs trained model. 58
Table 4-10 DL model final evaluation using the Test dataset. 59
Table 4-11 Accuracy values from HAR model final test, by each recognized activity. 59
Table 4-12 Ablation study results using test data for this dissertation proposed DL model architecture. 59
Table 5-1 Types of walking segmentation. 66
Table 5-2 Sensors and features presented in SoA research. 67
Table 5-3 AI algorithms presented in SoA research. 67
Table 5-4 Limitations and guidelines to overcome them. 68
Table 5-5 Walking events to be recognized label ids' 72
Table 5-6 Initial Contact and Final Contact sequence size. 78
Table 5-7 Optimization hyper-parameters. 79
Table 5-8 Fine-Tuning stage optimal parameters found, and the results obtained with them. 79
Table 5-9 Metrics evaluation for 10 epochs trained model. 81
Table 5-10 DL model final evaluation using the Test dataset. 82
Table 5-11 Accuracy values for Human Walk IC/FC Detection model final test, for each walking event recognized. 82

List of Abbreviations and acronyms

A

AE	Auto Encoder
AI	Artificial Intelligence
ANN	Artificial Neural Network
APP	Software application
Acc	Accuracy
Acce	Acceleration
ARF	Activity Recognition Framework

C

CNN	Convolution Neural Networks
CoM	Center of Mass
CC	Combinational Classifier
Cov	Covariance

D

deg	degree
DL	Deep Learning
DNN	Deep Neural Network
DRNN	deep recurrent neural network
DRNNs	deep recurrent neural networks
DT	Decision Tree

E

EAE	Ensemble of Auto-Encoders

F

FC	Final foot Contact
FoG	Freezing of Walking

G

GMM	Gaussian Mixture Modelling
GRU	Gate Recurrent Unit
Gyr	Gyroscope

H

HA	Human Activity
HAR	Human Activity Recognition
HW-IC/FC	Human Walk Initial foot Contact and Final foot Contact
HIER	Hierarchical agglomerative clustering
HP's	Health professionals

I

IC	Initial foot Contact
IMU	Inertial Measurement Unit
INN	Inception neural network

K

k	Number of activities
KNN	k-nearest neighbours

L

LDA	Linear Discriminant Analysis
LR	Linear Regression
LSTM	Long Short-Term Memory

M

Max	Maximum
MAS	Mean Absolute Error
ML	Machine Learning
MLP	Multilayer Perceptron
Min	Minimum
MinMax	Minimum-Maximum data normalization
MI	Minimum Intensity

P

PD	Parkinson's Disease
PSO	Particle Swarm Optimization
PwP	Patient with Parkinson's
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses

R

RF	Random Forest
RF	Radiofrequency
RNN	Recurrent Neural Network
RMS	Root Mean Square

S

SoA	State-of-the-Art
SD	Standard Deviation
SVM	Support Vector Machine

U

UPDRS-III	Unified Parkinson's Disease Rating Scale Part III

W

WBD	Instrumental Waistband
WS	Wearable Sensors

1 Introduction

1.1 MOTIVATION AND PROBLEM STATEMENT

Parkinson's Disease (PD) is a neurodegenerative disorder of the central nervous system [1], which is characterized by the loss of non-motor and motor functions. The non-motor symptoms include cognitive changes, behavioural/neuropsychiatric changes, autonomic nervous system failure, and sensory and sleep disturbances, while the motor symptoms are associated with resting tremor (initially unilateral), akinesia and bradykinesia (slow movements), rigidity, shuffling walking, and postural instability [2].

According to Sociedade Portguesa de Doentes de Parkinson (APDP) akinesia is the central symptom of Parkinson's Disease, meaning difficulty in initiating a movement, with a progressive reduction in speed, amplitude, and difficulty in executing a motor plan. Bradykinesia is the slowing down and poverty of movement. Akinesia and bradykinesia are translated into difficulty and slowness in opening and closing the hands, in holding a glass, to get up from a chair and to start walking, difficulty in writing which tends to become progressively smaller [3].

Rest tremor is a rhythmic movement of the extremities and is present in the hands, face, lip and chin, mainly in the resting position. It tends to appear asymmetrically, in the left or right hand, and may affect, with the progression and worsening of the disease, both sides and be present in the action of movements (e.g. holding a fork to the mouth) [3]. Rigidity consists of resistance to passive mobilization of a human body segment (neck, trunk, and limbs), remaining constant throughout the possible flexion or extension of the limb, leading to bradykinesia. This stiffness is accentually felt while combing the hair, brushing the teeth or when a freezing of walking happens [3]. Patients tend to adopt a curved posture, which, if not corrected, at long term causes changes in the curvature of the spine. Postural instability is a consequence of disturbed postural reflexes, due to a slower response to changes in the positioning of the center of body mass. Some patients have walking with small steps with decreased arm balance, poor foot elevation and sometimes an involuntary increase in the speed of increasingly shorter steps. Other patients have blockages during walking, at the beginning or during walking, with their feet sticking to the ground and difficulty in restarting the walking. In everyday life the Parkinson's patient is subject to several postural challenges of balance such as uneven flooring, changes in direction, sudden stops, unevenness, several tasks at the same time (e.g., talking, carrying a bag), inattention due to auditory (e.g. someone calling our name), visual and motor planning stimuli (e.g. moving around a chair; passing through a door between two rooms; imminence of an approaching obstacle on the floor). In these situations the parkinsonian patient has a higher risk of falling [3].

According to World Health Organization (WHO), disability and death due to PD are increasing faster than for any other neurological disorder, globally [4]. WHO estimated in 2019 that over 8.5 million individuals suffered from PD disability, which has doubled in the past 25 years.

In Portugal, a study published by Ferreira et al.[5] in 2017, estimated a disease prevalence of 180 cases by 100 000 habitants, which might indicate that over 185 000 persons is suffering from this disability.

PD not only impacts the life of the patients, but also the life of people around them. Often, family members and friends provide informal care to the patient due to financial incapacity. Spending many hours daily providing care for the patient living with PD, besides incurring financial losses, physical and emotional stress can be unsurmountable to these informal caregivers. As such, providing more autonomy for PD patients can be life changer.

Current approaches for monitoring patients' motor autonomy are limited to the observer and self-reported methods during the routine consultations. In the observer-based examinations, patients are required to perform a set of scored traditional PD examinations, such as Unified Parkinson's Disease Rating Scale Part III (UPDRS-III). The self-reported method relies in requiring patients to answer a list of questions about their daily activities, counting with the individuals' self-memory.

Despite these approaches being commonly used, they are limited to a few sessions per year, they do not address common motor daily tasks, and their results are object of subjective interpretation by the clinical expert. Indeed, walking, sitting on a chair or lying on a bed are examples of daily basic activities that are affected by PD, representing a spectrum of relevant information that if recognized and analysed could characterize patients' motor function. Thus, there is a need to develop systems able to provide quantitative measures of ambulatory performance. Indeed, being motor symptoms the main trademark of PD, physicians will benefit from clinical support tech-tools capable of supervise and gather data about patients' motor performance while performing daily physical activities. With this information physicians will be able to adopt a closed follow-up of the true motor stage of their patients, enhancing the power of prescribed treatments.

Due to constantly decreasing production cost, availability, portability, miniaturized size and lower power consumption of wearable sensors (WS), scientific breakthroughs in fields such as machine learning (ML) and internet of things, these sensors emerged as a powerful tool to enable the human activity recognition (HAR) [6]. WS, like accelerometer, gyroscope, pedometers, can be defined as devices which can be worn by humans to track human health and can provide continuous monitoring of health parameters without visiting hospitals frequently [7].

By combining the kinematic-driven outcomes from WS with advanced methods of ML, an activity recognition system could be implemented to provide physicians a quantitative profile of motor function behaviour in natural settings and over prolonged periods. Also, real-time implementations could be integrated on assistance/rehabilitation devices to deliver sensory information about their posture or risk of falling during motor activities aiming to help patients to mitigate motor conditions.

This emphasizes the significance of developing task recognition frameworks capable of provide long-term and objective data to physicians, but also be able to be integrated on wearable assistance devices. For example, inside the HAR field of study, walking analysis is probably the most studied and has the broadest scope of applications including medicine, rehabilitation or sports [8]. In walking analysis, the capability of detecting initial and final foot contact during walking cycle can endow the frameworks of on-demand motor assistance and estimation of walking-associated metrics. This showcases the HAR importance.

Although HAR systems based on combination of WS and ML have received significant attention on the last years, few studies have been dedicated to PD. It is required further studies about which motor tasks could better describe patients' motor function, as also more clinical evidence is needed. Advanced artificial intelligence (AI) models, such as ML, need more analysis regarding data input processing, models' features, and benchmarking analysis on models' performance. Also, there is a need to consider a HAR framework able to identify human activities from a minimum number of WS, aiming to be a usable wearable technology that is comfortable and easily accepted by patients.

Therefore, future challenges on HAR systems based on AI combined with WS data include: **(i)** study HAR in PD; **(ii)** create an open-source multimodal dataset of physical activities in PD based on 3D motion data and kinematic-driven walking parameters acquisitions through wearable miniaturized inertial sensors; **(iii)** study the hypothesis to standardize protocols to capture the meaningful motor tasks able to describe patients' motor function, increasing at the same time the clinical evidence; **(iv)** explore advanced AI models; **(v)** include raw data input signals from a single wearable sensor. From this dissertation it is expected to answer these **key-constraints**.

1.2 GOALS AND RESEARCH QUESTIONS

By combining wearable sensor data with AI, the **main goal** of this dissertation is to **develop an automatic software for recognition of human activities** (e.g., walking, standing, turning, sitting, and lying) in PD. Wearable sensory acquisition systems will rely on the **shortest number of inertial**

sensors and in the lesser transformed data as possible to provide information about users' motion without compromising patients' movement. With the **application of algorithms of AI over these kinematic data**, it is expected to identify different physical activities that provide complete information about users' motion. It is expected to contribute with an AI-based model capable of accurately identify different motor activities performed by a PD patient in his/her day-to-day life activities.

There are various objectives that will allow to pursue the **ultimate goals** of this master's dissertation, which are outlined below:

Goal 1 - Critical analysis on the state-of-the-art (SoA) HAR in PD patients, to answer the following questions: **(i)** "Which are the most relevant daily activities performed by PD patients?" **(ii)** "Which sensors are used to generate relevant data?" **(iii)** "What are the best algorithms to use in HAR?". In Chapter 0 we address this issues in depth.

Goal 2 - Determine which data is relevant to be used in the recognition of patient daily activities. This goal is addressed by Chapter 0, where in combination with the literature we will decide the best way to handle these data. **A new data collection protocol is to be implemented**, the resultant dataset should have data from twenty PD patients, each performing three trials of sit and get up from chair, get up and lay down from bed, walking, 180° right and left turning activities.

Goal 3 - Define and **implement a framework capable to recognize HAR in PD patients**. By combining the information from the read literature, Chapter 0, an attempt to create a framework for HAR have been made. This framework must have a similar or better performance than the 0,946 F1 score found in the literature and be the most efficient has possible. This implementation is described in Chapter 4.

Goal 4 - Define and **implement a framework for PD walking segmentation**, considering that the detection of **initial (IC) and final (FC) foot contact** of a walking cycle is tremendously important to provide on-demand motor assistance and estimation of walking-associated metrics. In depth description and implementation are discussed in Chapter 5.

Considering the final goal of this dissertation and the step-goals mentioned before, several research questions were identified to be investigated and answered:

RQ1: What is the importance of daily activities recognition for motor assessment in PD? How have human activities been detected? And of IC/FCs?

RQ2: Which DL model produces best result in recognizing daily human activities? And in IC/FC walking events recognition?

RQ3: How effective and robust are the proposed DL solutions toward new patients?

The first question relates to the main Goal of this dissertation, on why and how this dissertation can contribute in a meaningful way to society. Throughout the chapters, **Chapter 4** and **Chapter 5, answers to RQ2** were discussed in depth. This RQs' are related to Goal 2 and Goal 3, and asks about different DL models architectures, training and validating performances of the proposed DL models. **RQ3**, also related to Goal 3, is **answered trough Chapter 4 and Chapter 5** by testing the proposed model with data from unknown PD patients.

1.3 CONTRIBUTION TO KNOWLEDGE

The main contributions of this dissertation to knowledge are:

1. **Review on SoA literature about HAR on PD**, with the purpose to provide a closed, continuous, and objective clinical support tool using WS as input of AI-based models
2. A **new database** with data from daily life physical activities performed by **PD subjects**.
3. A framework based on **DL algorithms capable of accurately recognize HAR** performed by PD patients.
4. A framework based on **DL algorithms capable of accurately detect and recognize human walk IC/FC events** performed by PD patients in real time.
5. By **changing the constitution of the final DL models**, for both the HAR model and the human walk IC/FC detection model, an **attempt to explain** how their various constituent layers affect their building was made.

1.4 DISSERTATION STRUCTURE

This manuscript is divided in 6 chapters, including this as a first chapter, Chapter 1, which provides the motivation and context of the problem in study, what is the main goal of this dissertation and the main questions that emerged while achieving it, and finally how this dissertation is structured and what is the new knowledge this dissertation brought to the scientific community.

In Chapter 0, is presented the theoretical knowledge of concepts and technologies addressed in this work. In this chapter a SoA review is presented, giving a background about the usage of WS to collect data and the AI models capable of recognizing human activities. A discussion about the answers to the

research questions made in Chapter 0, what was concluded after better understanding of the SoA and possible future direction of HAR discussion is also debated in this chapter.

In the Chapter 3, an overview of the solution presented by this dissertation is exposed.

In chapters 4 and 5 a description on how this dissertation main and complementary goals were achieved is elaborated. These chapters include the different stages needed to achieve the final AI classification model. Data collection and exploration, model training, validation and test performance evaluation are here presented and discussed.

On Chapter 6 a conclusion about the important role of the AI in the close and continuous monitoring over PD patients is discussed while providing a brief analysis of the results obtained by this dissertation proposed solution.

2 Literature Review

2.1 INTRODUCTION

Difficulty in initiating movements, slowing down and poverty of movement, resting tremors in body extremities, rigidity in movements, postural instability and freeze of walking (FoG) episodes are the main motors disabilities that PD patients suffer in day-to-day life. Thus, trivial motor activities are affected by these motor disabilities, and consequently patients' quality of life.

Nowadays only two options are available to clinical personnel for monitoring and evaluating the evolution of the disease in patients. By recurring to patients self-reported diaries, which are often incomplete, or by routine consultations, that sometimes can be sparse in time, where the clinical staff assesses in a subjective way the disease degree by applying already pre-existing and impersonal scales like UPDRS-III. As such, it is emergent and crucial an alternative technological solution capable of a constant monitoring, addressing long-term acquisitions on domiciliary contexts to provide a patient customized clinical support decision tool.

Wearable sensing technology brings an emerging healthcare paradigm that assimilates the benefits of smart sensing devices, communication technologies and clinical diagnosis. This technology provides reliable and robust health monitoring services and reduces human hand in a cost effective manner [9]. In fact, in PD domain, WS can monitor patients' motor function and help physicians implement an adequate treatment to improve motor symptoms. An activity recognition system could provide clinicians with a quantitative profile of patients' motor behavior in home scenarios and over prolonged periods of time. Also, real-time monitoring of patients' locomotion could be used to apply timely interventions and prevent associated walking injuries leading to patients' motor autonomy.

By combining WS data with advanced AI algorithms, research and development efforts have accomplished new HAR frameworks capable of a continuous patient monitoring. There are different body worn sensors and even more AI algorithms capable of successfully performing HAR. Inertial sensors or reflective markers are a good example of how wearable sensorial technology in conjugation with AI algorithms can classify steady state locomotion's, such as sitting, standing up/down or laying on bed, or non-steady state locomotion's, like ascending/descending stairs and walking or to predict and classify PD symptoms [10][11][12][13].

Considering the need to better understand the SoA, it was accomplished a comprehensive review on the scientific contributions about wearable devices combined with AI for HAR. From this critical review, the following questions were investigated and answered: **(i)** *Which locomotion modes are recognized and for what purpose?* **(ii)** *Which type of input data was used in AI models for locomotion modes recognition*

systems? and **(iii)** *Which AI models and how were they implemented in locomotion modes recognition systems*?

The first question indicates which are the most relevant motor activities. This knowledge enabled to build a framework focused on what is important to measure on PD patients. The second and the third research questions offer a more technological point of view over the SoA of HAR, allowing to better manipulate and understand how the most recent WS and promising AI algorithms have been used for HAR.

2.2 METHODS

2.2.1 DATA SOURCES, SEARCH STRATEGY AND STUDIES SELECTION

An electronic systematic search was carried out on databases as Google Scholar, Research Gate, Science Direct, Scopus and Web of Science. The survey was conducted according to the guidelines of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), as depicted on Figure 2-1.

For that purpose, key-words matching headings were used: [“locomotion modes recognition”]; [“locomotion modes in parkinson patients”]; [“parkinson's wearable”]; [“inertial AND sensors AND parkinson AND machine AND learning”]; [“parkinson machine learning”]; [“detecting parkinsons movements”]; [“inertial AND Parkinson”]; [“activity recognition”]; [“gaist recognition”]; [“locomotion recognition”]; [“parkinson's disease review”]; [“activities daily living”]; [“sensor motion analysis”]; [“sensor motion analysis deep learning”]; [“parkinson patient activity recognition deep learning”] ;[“wearable sensors deep learning activity”]; [“locomotion wearable Parkinson”]; [“activity recognition Parkinson”] ;[“deep learning walking events”]; [“deep learning walking segmentation”]; [“wearable sensors”] and [“wearable sensors in healthcare”].

Studies were included if they fulfilled the following **inclusion criteria**: **(i)** studies of idiopathic PD; **(ii)** preferable use of WS to classify daily activities **(iii)** use of DL models to recognize usual HA, **(iv)** data used in the paper was collected from healthy patients or PD patients or both, **(v)** results were published in the English language and within the past 11 years.

The **exclusion criteria** were: **(i)** not using inertial sensors data to perform HAR, **(ii)** not recognizing steady and non-steady activities, **(iii)** not using models of deep learn in the classification of activities. Some articles’ reference lists were searched for additional support.

2.3 RESULTS

2.3.1 GENERAL RESULTS

A total of 196 articles were identified: Google Scholar (n= 165), Research Gate (n= 5), Science Direct (n= 3), SCOPUS (n= 14) and Web Of Science (n= 9) databases. Duplicates were removed (n= 87). Articles were excluded if they have the following keywords on titles (n= 18) and abstract: FoG, objective assessment for PD, tremor activity and limb amputation. In case the abstract of an article did not provide enough information to determine its eligibility, the full article was reviewed. Next, the full-text papers were reviewed to meet the inclusion criteria and 13 articles met the eligibility criteria and were included in this review.

These reports were discussed according to their different aspects in performing HAR, i.e., What (1) locomotion modes and motor activities should be recognized, (2) AI model input to use and (3) AI models implemented for in HAR. As a result, 83 article were eligible to be part of this discussion although in order to cover all the paradigms of HAR, 13 proposed AI-models' articles were identified to serve as base of discussion about HAR. Firstly, we discriminated the human activities to be recognized, secondly the sensor and types of data they produce, acquisition protocols and datasets used, lastly process of features selection/extraction and implementation AI-models and their performance.

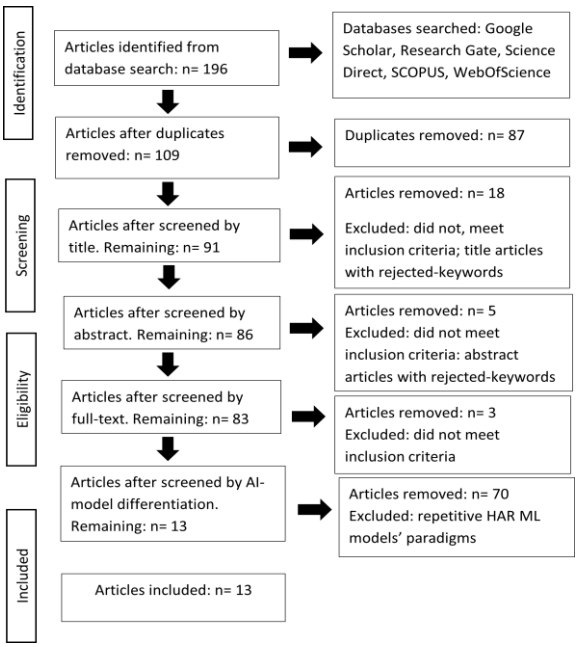


Figure 2-1 Flowchart for the search strategy based on Preferred Reporting Items for Systematic Reviews and Meta-Analyses(PRISMA).

2.3.2 DAILY MOTOR ACTIVITIES

Table 2-1 provides an overview of the human activities recognized by the selected studies during the SoA survey, as well as a synthesis of the studies objectives.

Table 2-1 Overview of the most represented studied activities.

Paper	Target	Motor activities
[10]	PD	Ramp ascent/descent, upstairs, downstairs, and level-ground walking
[14]	Healthy	Walking, jogging, upstairs, downstairs, sitting, standing and others
[15]	Healthy	Walking, jogging, upstairs, downstairs, sitting, standing and others
[16]	Healthy	Walking, jogging, upstairs, downstairs, sitting, standing and others
[17]	Healthy	Walk, jogging, skip, upstairs, downstairs, standing, sitting, and lying
[18]	Healthy	Walking, upstairs, downstairs, sitting, lying, running, jumping, sleeping
[19]	Healthy	Sitting, standing, walking, lying, sitting, walking, upstairs, downstairs and others
[20]	Healthy	Sitting, standing, walking, lying, sitting, standing, walking, upstairs, downstairs and others
[21]	Healthy	Downstairs, upstairs, sitting, standing, walking, jogging
[22]	Healthy	Downstairs, upstairs, sitting, standing, walking, jogging and others
[23]	Healthy	Walking, running, sitting, standing, and lying
[24]	Healthy and PD	Tandem walk, sitting, standing, standing, backwards walking, walk
[25]	Healthy	Walking, jogging, upstairs, downstairs, sitting, standing and others

Eleven of the thirteen selected articles included healthy people as their target audience. Only [10] and [24]focused on the recognition of activities performed by people with PD. These studies produced DL models capable of identifying human activities from the simplest (walking, running, sitting, standing, lying down, climbing, or descending stairs) to the most complex (Nordic walking, car driving, playing soccer or rope jumping).

The studies [10],[17],[18],[21],[23],[24] focused their models on the most recurrent activities to be performed in daily life (walking, running, sitting, standing, lying down, climbing or descending stairs) that may provide the more meaningful information from a clinical point of view. The focus on these

common activities, despite being more accessible to identify and already providing important information about patients' motor status, it may also limit the physicians from the possibility of a more profound follow-up of their patients. Indeed, [14],[15],[16],[19],[20],[22],[25] complemented this simplicity of activities with the addition of data related to more complex human activities, such as folding laundry, prepare a sandwich or house cleaning. This more complex activities data might be of assistance to clinical experts to assess the impact of the disease in their patient's daily life, well as the disease progression. However, the addition of these more complex activities, makes the classification models more complex and thus negatively affecting their classification ability. It is a necessity to find a trade-off between the daily activities and the model's architecture utilized to recognize these activities.

2.3.3 DATA INPUT

Table 2-2 provides an overview of the data used to train the models for HAR in the selected studies. Also identifies which devices were used to collect the data, how they were implemented, their availability and what transformations they underwent to be used to train the AI models.

In study [10], the researchers created a new dataset with data collected from five healthy subjects (4 males, 1 females, age 25.2±2.5 years, height 1.75±0.11 m, mass 66.8±12.2 kg) and five individuals with early stage PD (2 males, 3 females, age 62.8±3.9 years, height 1.72±0.03 m, mass 77.5±17.88 kg, Hoehn and Yahr stage 1 or 2) participated in the study. Sixty-six **reflective markers were attached to anatomical body** locations to track 12 body segments of the arms, legs, and torso. The experimental setup consisted of a **"terrain park" circuit** including an over-ground walkway, a four-step staircase with step height of 0.15 m and depth of 0.30 m, a 2.5 m ramp inclined at 10°, and elevated platforms to connect the stairs and ramp. The platform contained a single step of height 0.15 m. Individuals with **PD** were asked to walk at their comfortable speed and perform **five trials of the circuit** for both left leading and right leading legs in the following orders: stair ascent/ramp descent and ramp ascent/stair descent and, they were instructed to use handrails when desired. **Healthy** subjects performed **five trails** of the circuit while using the handrails, and five sets without using the handrails.

Another research where the researchers created their own dataset was [23], where no information about the number of participants from which the data was collected is provided but information about the sensor used was disclosed, in this case a **single smartphone** was kept in a pants pocket **to collect the sensor data** during **five common physical activities, each of which was performed for ten minutes.** To eliminate noise from the sensor data a simple low-pass filter was applied. Also, [24] created their own dataset, which consisted of records from **10 healthy elderly**

volunteers (mean age 75 years) and **two Patients with Parkinson's** (PwP) (mean age 76 years). They performed six various mobility tests in the laboratory including Tandem walk, stand to sit, sit to stand, standing, backwards walking test and 3m walk. Each activity was performed three times. They collected participants movement by a video camera and a wearable sensor (includes an accelerometer and a gyroscope, set to capture three axes at 50 Hz) attached to participant's lumbar spine.

Table 2-2 Sensors, data input and protocols observed in literature.

Paper	Sensors			Data acquisition		Dataset		
	Which one?	**How many?**	**Where?**	**Participants**	**Protocol**	**Setting**	**Open source?**	**Pre-processing**
[10]	reflective markers	66	arms, legs and torso	5 H 5 PD	ramp ascent/descent, upstairs, downstairs, and level-ground walking	L		
[14]	smartphone IMU	1		36 H	walking, jogging, upstairs, downstairs, sitting, standing and others	L	✓	
[15]	smartphone IMU	1		36 H	walking, jogging, upstairs, downstairs, sitting, standing and others	L	✓	
[16]	smartphone IMU	1		36 H	walking, jogging, upstairs, downstairs, sitting, standing and others	L	✓	
[17]	smartphone IMU	1		570 H	walk, jogging, skip, upstairs, downstairs, standing, sitting, and lying	H	✓	
[18]	smartphone IMU	2	right hip	44 H	walking, upstairs, downstairs, sitting, lying, running, jumping, sleeping	H	✓	
[19]	IMU (Accelerometer, Gyroscope)	26	wrist, arm, chest, dominant side's ankle, upper body, shoes, legs, hip, left/right front/back shoulder	13 H	sitting, standing, walking, lying, sitting, walking, upstairs, downstairs and others	H	✓	

IMU: inertial measurement unit; PD – patients with PD; H – healthy subjects; L – laboratory; H -home; Acce: accelerometer; Gyro: Gyroscope.

Paper	Sensors			Data acquisition		Dataset		
	Which one?	**How many?**	**Where?**	**Participants**	**Protocol**	**Setting**	**Open source?**	**Pre-processing**
[20]	IMU	26	wrist, arm, chest, dominant side's ankle, upper body, shoes, legs, hip, left/right front/back shoulder	13 H	sitting, standing, walking, lying, sitting, standing, walking, upstairs, downstairs and others	H	✓	null handling, normalization [0,1]
[21]	IMU	1	smartphone	24	downstairs, upstairs, sitting, standing, walking, jogging	H	✓	2nd notch filter
[22]	IMU	26	wrist, arm, chest, dominant side's ankle, upper body, shoes, legs, hip, left/right front/back shoulder	13 h	downstairs, upstairs, sitting, standing, walking, jogging and others	L	✓	Normalization [0,1]
[23]	smartphone IMU	1	hip	-	walking, running, sitting, standing, and lying	H		low-pass filter, Normalization [0,1]
[24]	IMU	1	body centre of mass	10 H 2 PD	tandem walk, sitting, standing, standing, backwards walking, walk	L		
[25]	smartphone IMU and IMUs	4	wrist, arm, chest, ankle	46 H	walking, jogging, upstairs, downstairs, sitting, standing and others	H	✓	

IMU: inertial measurement unit; PD – patients with PD; H – healthy subjects; L – laboratory; H -home; Acce: accelerometer; Gyro: Gyroscope.

[14],[15],[16] used the **WISDM public dataset**. This dataset is often used in papers where it is intended to make an evaluation of the quality of the proposed model in comparison with other existing ones. WISDM dataset was built by using the raw accelerometer and gyroscope sensor **data collected from the smartphone and smartwatch** at a rate of 20Hz. It was collected from 51 test subjects as they perform 18 activities for 3 minutes each. In [15] only the data from six activities (**walking, jogging, upstairs, downstairs, sitting and standing**) from 36 different subjects was used and in [16] the same activities of [15] were selected but only the accelerometer data from 29 different subjects was used.

Inoue et al [17] and Murad et al [18] both used the public **dataset HAR Using Smartphones**. This **dataset is composed of experiments performed by 30 healthy volunteers** within an age bracket of 19-48 years. Each person performed six activities (**walking, walking upstairs, walking downstairs, sitting, standing, laying**) wearing a **smartphone on the waist**. Using its embedded accelerometer and gyroscope, they captured **3-axial linear acceleration and 3-axial angular velocity** at a constant rate of 50Hz. To label the data of the experiments, manual work supported by video-recordings was done. Additionally, in [18] the **public USC-SIPI Human Activity Dataset** was used. This dataset was collected from **14 subjects** (7 male, 7 female) with a age ranged from 21 – 49 years, height ranged between 160 – 185 cm and weight between 43 – 80 kg, by using a high performance inertial measurement unit (**IMU**) (3D accelerometer and gyroscope) sensor **positioned on volunteers' front right hips**. The dataset contains 12 basic human activities: **walking forward, walking left, walking right, walking upstairs, walking downstairs, running forward, jumping up, sitting, standing, sleeping, in elevator up, and in elevator down**. Here the researchers considered 11 classes by combining the last two activities into a single "in elevator" activity.

[19],[20],[22] validated their proposed models in two public HAR benchmark datasets, **PAMAP2** and **Opportunity datasets**. PAMAP2 dataset contains data of 18 different physical activities (such as walking, cycling, playing soccer, etc.), performed by 9 subjects wearing 3 IMU placed over the wrist on the dominant arm, on the chest, on the dominant side's ankle and with a sampling frequency: ~9Hz. Opportunity dataset comprises records from 4 users wearing 7 IMU, 12 3D acceleration sensors, 4 3D localization information performing 18 different activities. In [20] data pre-processing was needed in order to fill the missing values of the sensor, to do that a linear fitting was executed, and the data for each sensor channel was normalized to the [0,1]. In [19] and [22] there are no records of data pre-processing being applied.

In order to prove the proposed algorithm, in [21], the **Motion Sense** benchmark public dataset was used. This dataset was built with data obtained through the sensor of a **smartphone**, which was not fixed on the user and data was produced in a controlled laboratory environment. For the dataset, data from 6 distinct activities (**walk downstairs, walk upstairs, sit, stand and jogging**) were collected from 24 healthy users, 14 men and 10 women with ages between 18 and 46 years old, heights between 1.64 and 1.90 m and body weight between 48 and 102 kg. Before starting the model training, a second order notch (band-stop) filter was applied to the sensors to reduce the noise in the signal.

2.3.4 AI-BASED MODELS

Table 2-3 shows which AI models have been proposed for HAR, the attributes used in the training of the respective models and the results obtained through them.

Garcia et al [25] used for benchmark tests of the proposed algorithm public datasets already previously mentioned, WISDM and PAMAP2. For both datasets exposed to the proposed algorithm, only the acceleration data were used in the model training process.

Kazemimoghadam et al [10] implemented and compared two different algorithms, LDA and LSTM, using the F1 score metric, to evaluate the quality of the developed models. Each of the models were exposed to 3 different training processes, one of these training processes consisted in using data from healthy users for training and using the data from users with PD as validation, the second consisted in using the data from users with PD for training and using the leave-one-subject-out technique with data from healthy users and users with PD, lastly the training and testing used the data from users with PD and applying the cross-validation technique. To classify the tasks using LDA, six time-domain features from accelerometer and gyroscope including minimum, maximum, mean, standard deviation, first and last sample of each window were extracted. To classify the tasks using LSTM raw data without employing feature extraction was used, and the number of neurons in the input layer was adjusted according to the number of input signals and also as parameters used for optimizing the LSTM are batch size, number of epochs, and number of hidden units. The article presents a table of the several results with the several training paradigms defined by the authors, and in general the hypothesis that an LSTM is a better classifier than an LDA model is confirmed, having LSTM obtained an average F1 score of 0.89 and LDA an average F1 score in the recognition of activities of 0.69.

Table 2-3 Summary of machine learning models proposed by several authors, and their data processing and algorithms performances.

Paper	Model	Features	feature selection /extraction	Language/SW	Performance	Real-time?
[10]	LDA e LSTM	Acce, Gyr	Min, max, mean, SD	-	LDA: F1 score=~0.64 LSTM: F1 score = ~0,88	
[14]	CNN with shallow features	Acce, Gyr	Interquartile range, amplitude, kurtosis, RMS, variance, mean, SD, skewness, min, mean-cross, median, max, and zero-cross		Acc = 98,6%	
[15]	J48, MLP and LR	Acce, Gyr	Min, max, average, SD, zero crossing and correlation between axis	WEKA machine learning toolkit	J48: Acc = 96,7% LR: Acc = 84,7%, MLP: Acc = 94,7% CC: Acc = 97,2%	
[16]	CNN	Acce, Gyr	-	FFTW3 library and the Torch	Acc =98,2	✓
[17]	DT, SVM, RF, RNN	Acce, Gyr	Mean, variance, MAS of the, first and second eigenvalue of the Cov matrix, sum of the vertical component ratios for the intensity, Cov ratio, variance ratio, mean FFT-domain energy, mean FFT-domain energy of the intensity, FFT-domain entropy, FFT-domain entropy of the intensity, number of mean crosses of the MI, number of crosses of the zone of the MI ± 0.1 G, and number of samples outside the zone of the MI	-	DT: Acc = ~59%, SVM: Acc = ~65%, RF: Acc = ~70%, RNN: Acc = ~98%	
[18]	RNN	Acce, Gyr	-	-	UCI-HAD Proposed RNN: Acc= ~96,7, USC-HAD proposed RNN: Acc= ~97.8	No

LSTM: Long Short-Term Memory; LDA: Linear Discriminant Analysis; CNN: Convolutional Neural network; RNN: Recurrent Neural Network; MLP: Multilayer Perceptron; Min: minimum; Max: maximum; RMS: Root Mean Square; MAS: mean absolute sum SD: standard deviation; Acce: Accelerometer; Acc: accuracy; CC: Combinational Classifier; LR: Linear Regression; DT: Decision Tree; SVM: Support Vector Machine; RF: Random Forest; GMM: Gaussian Mixture Modelling; HIER: average-linkage hierarchical agglomerative clustering; EAE: Ensemble Auto Encoder; MI: mean intensity; Cov: covariance

Paper	Model	Features	feature selection /extraction	Language/SW	Performance	Real-time?
[19]	Ensemble LSTM	All features presented on datasets	-	-	Opportunity: F1 score and SD: 0.726+-0,008, PAMAP2: F1 score and SD: 0,854+-0,026	No
[20]	combination of inception neural network and recurrent neural network	Opportunity: Acce Gyr; PAMAP2: All features	-	Keras 2- Tensorflow framework, Python	PAMAP2: F1 score= 0,935; Opportunity: F1 score= 0,946	Yes
[21]	SVM with PSO	Acce, Gyr	mean, median, harmonic mean, sin, cosine, position vector and MFCC (Mel frequency cepstral coefficients)	-	Accuracy: 87,5	no
[22]	CNN	All features	-	-	Opportunity: Acc= 89,66 F1 score=89,57; PAMAP2: Acc=92,55 , F1 score=92,60	No
[23]	k-means clustering, GMM, and HIER	Acce, Gyr	mean, SD	WEKA machine learning toolkit	k-means: Acc=0,7198; GMM: Acc=1; HIER: Acc=0,7998	No
[24]	Naive Bayes, LogitBoost, Random Forest, SVM	Acce, Gyr	correlation based feature selection, forward and backwards features selection methods and a wrapper feature selection method based on random forest algorithm	WEKA machine learning toolkit	naive bayes: Acc=81,24; LogitBoost: Acc=89,69; Random Forest: Acc= 92,29; SVM: Acc=84,17	no
[25]	EAE, EkVN model	Acce and Gyr;	mean, SD and Pearson correlation	-	WISDM: EkVN Acc=0,73, EAE Accy=0,82; PAMAP: EkVN Acc=0,71 , EAE Acc= 0,63	No

LSTM: Long Short-Term Memory; LDA: Linear Discriminant Analysis; CNN: Convolutional Neural network; RNN: Recurrent Neural Network; MLP: Multilayer Perceptron; Min: minimum; Max: maximum; RMS: Root Mean Square; MAS: mean absolute sum SD: standard deviation; Acce: Accelerometer; Acc: accuracy; CC: Combinational Classifier; LR: Linear Regression; DT: Decision Tree; SVM: Support Vector Machine; RF: Ranfom Forest; GMM: Gaussian Mixture Modelling; HIER: average-linkage hierarchical agglomerative clustering; EAE: Ensemble Auto Encoder; MI: mean intensity; Cov: covariance

A proposal to combine a set of shallow features with those obtained from DL was made on [14]. The authors propose a pipeline that combines both shallow and deep learnt features, the raw datasets measured by the inertial sensors (accelerometer and gyroscope) are collected and divided into segments. The automatically learnt features and the shallow features are extracted in processes A and B, respectively. In the last block, the features are combined and classified using a fully connected layer and a soft-max layer of the deep learning model. In process A, a set of deep features is automatically extracted using the proposed deep learning model to convert the input data into a spectrogram representation that will feed a 1-D convolutional kernels with the same principle that CNN. In process B predefined shallow features are considered. The features are Interquartile Range, Amplitude, Kurtosis, Root Mean Square, Variance, Mean, Standard Deviation, Skewness, Min, Mean-cross, Median, Max, Zero-cross and are extracted separately from each segment of each axis, creating a vector representation for the considered segment. Average accuracy was used as comparing metric, and the proposed model achieved 98,6% of accuracy.

In [15], the authors implemented three experiments using WEKA software, one using J48 algorithm, the second using MLP algorithm, LR algorithm and finally an ensemble of the three previous algorithms. In the combinational classifier the voting algorithm based on their average probabilities was used to classify the activity. Before implementing the models, data extraction from the sensors data was needed, extracting data like minimum value, maximum value, average, standard deviation, zero crossing rate for every axis and correlation between axes. By applying the previous models to testing set the algorithm J48 obtained 96.7% accuracy, LR model achieved 74.7% accuracy, MLP classifier reached 94.7% accuracy and the combinational model obtained a mean accuracy of 97.2% in the classification of the multiple activities.

In [16] a comparison between their proposed model against SoA models was also made. The proposed DL architecture applied a filter to the pre-arranged spectrograms of the input, and the weighted sums of the convolved signal at each time are computed in the temporal convolution layer and finally, the fully connected layer and soft-max layer are used for classification. To extract the spectrogram from the data FFTW3 library was used and the Torch framework was needed to implement the DL model. The proposed model achieved an average accuracy of 98.2%

Inoue et al [17] proposed a method of HAR from raw accelerometer data applying a recurrent neural network (RNN). The results obtained were then compare with more classic ML model such as Decision Tree (DT), Random Forest (RF) and Support Vector Machine (SVM). The quality measure

instrument was the accuracy. To train the more classic model were extracted from data 27 new features values, but to find the most relevant features a stepwise-feature selection using logistic regression was applied resulting on the 13 following features: mean value of x, y axis, variance of z axis, mean sum of the absolute values of each axis, second eigenvalue of the covariance matrix between the axes, covariance ratio in the x- and y-direction for the z-component variance of each axis, (14–16) variance ratio of the back and forth difference in the x- and y-direction for the variance of the back and forth difference in the z-direction of y axis, mean FFT-domain energy of the intensity, FFT-domain entropy of x axis, FFT-domain entropy of the intensity, number of crosses of the zone of the mean intensity. The authors constructed a deep recurrent neural network (DRNN) such that the three-axis acceleration data of each time corresponded to the three-dimensional input layer, and six activity classes to the six-dimensional output layer. Each unit of each internal layer was an LSTM unit. The activation function of the output layer and the error function were defined by a softmax function and a cross entropy function, respectively. The truncated BPTT under the mini-batch stochastic gradient descent method was used to update the weights at the time of training. Chainer framework, which uses Python language was used to build the RNN model. Parameters such as the numbers of layers and units, truncated time, and dropout rate, have been optimized to obtain the best result possible and so the proposed model obtained an accuracy of 95.42% which is higher 35.18%, 27.76% and 22.35% than those obtained from the comparative models of DT, SVM, and RF, respectively.

A proposal of the use of deep recurrent neural networks (DRNNs) for building recognition models that are capable of capturing long-range dependencies in variable-length input sequences is presented in [18]. This manuscript presents us a unidirectional, bidirectional, and cascaded architectures based on long short-term memory (LSTM) DRNNs and evaluate their effectiveness on miscellaneous benchmark datasets. In all this models the inputs are raw signals obtained from multimodal-sensors, segmented into windows of length T and fed into LSTM-based DRNN model. The models outputs class prediction scores for each timestep, which are then merged via late-fusion and fed into the softmax layer to determine class membership probability. The Unidirectional LSTM-Based DRNNs model, consists of an input layer, several hidden layers, and an output layer and the number of hidden layers is a hyperparameter that is tuned during training. The bidirectional LSTM-based DRNN, includes two parallel LSTM tracks: forward and backward loops for exploiting context from the past and future of a specific time step to predict its label, and the number of hidden layers is also tuned during training. Cascaded unidirectional and bidirectional LSTM-based DRNN models are a cascaded structure in which the first layer is a bidirectional RNN, and the upper layers are unidirectional, with the number of layers being tuned also during training. To verify

the performance of the proposed models, the authors employed four widely used evaluation metrics for multi-class classification, precision, recall, accuracy, and F1-score. The model who attained the best performance was Unidirectional DRNN with an average accuracy=96.7%, precision= 96.8%, average recall= 96.7%, F1 score= 0.96 in UCI dataset and in USC-HAD dataset an accuracy = 97.8%, precision = 97.4.0%, average recall = 97.4% and F1 score = 0.97.

In [19] an attempt to demonstrate that Ensembles of deep LSTM learners outperform individual LSTM networks on human activity recognition using wearable sensors is made. Each LSTM model used present in the Ensemble was constituted by two-layer LSTM networks, with each layer containing 256 LSTM units. Dropout was performed at the first and second hidden layers both with a probability of 0.5. We used the ADAM updating function, with a learning rate of 0.001. The metrics used to evaluate the performance of the Ensemble was F1 score and crosse entropy. In the process of training the authors made different combinations of Ensembles by changing the number of ensembles (1,10,20) and changing the evaluation score (F1 score, Cross entropy). The overall best model configuration, is an Ensemble with 20 LSTM models mixing cross entropy and F1 score + F1 as loss functions and with 10 base learners each, reaching the best F1 score 0.726, 0.854 on opportunity dataset and PAMAP respectively.

A proposition of a multi-level neural network structure model based on the combination of Inception Neural Network (INN) and Gate Recurrent Units (GRU) is made in [20]. The model proposed is a pipeline of 4 consecutive INN, followed by one max-pooling layer to help the network better eliminate misjudgment caused by noise disturbance, and finally by to 2 GRU layers, so that the model can better extract the sequential temporal dependencies, INN is a new and innovative network structure proposed by Google in the second half of 2014, composed by a 1×1 convolution kernel to directly activate the combination of multi-channel information and pass it to the next layer. Two convolution kernels of 1 × 3 and 1 × 5 are cascaded respectively by a 1 × 1 convolution kernel. The network structure was built in Python with the assistance of Keras 2 - Tensorflow framework. The authors considered that classification accuracy was not an appropriate index for performance evaluation and so they decided to use F1 score as evaluation metric. The proposed solution obtained a F1 score of 0.946, 0.935 on Opportunity and PAMAP2 datasets respectively. In addition, the authors performed a real-time system evaluation of the model, reaching a predicted speed 65.09 sliding windows/s.

In Batool et al [21], an SVM model based on Particle Swarm Optimization (PSO) is proposed. In the proposed model, six different feature like mean signal, median signal, harmonic mean, sine and cosine signal, position vector signal, frequency cepstral coefficients methods were extracted from the data

and are applied on the already pre-processed data (acceleration, gyroscope) for the analysis of human activity in frequency domain. The model proposed by the authors consists in applying the PSO algorithm that works as a pre-classifier, and the resulting particles are then used to feed the SVM model. A comparison of SoA methods with proposed method was made, by using accuracy metric to compare. The SoA methods were k-nearest neighbors (KNN) (accuracy= 74.0%), artificial neural network (ANN) (accuracy=75.3%) and a standard SVM (accuracy= 84.4%) classifier, and the proposed model, SVM with PSO, obtained 87.5% of accuracy.

An architecture that processes time-series of multiple IMUs separately was proposed in [22]. However, this architecture contains multiple parallel processing branches, following the idea of wider rather than deeper networks. This architecture is referred as CNN-IMU network, and it contains parallel branches, each with temporal-convolutions, subsequent pooling operations, and an additional fully connected layer. Each branch contains N blocks, each of them having two stacked 5×1 temporal convolution followed by 2×1 max-pooling operations, and, eventually, concatenated by a fully connected layer. As referred Instead of scaling the network deeper, these layers are processed in parallel for each IMU, increasing the network's descriptiveness. The network combines these intermediate representations into a global one by means of a subsequent fully connected layer. In [10], CNN-IMU is adapted to not realize max-pooling operations while CNN-IMU-2 does, and this is the biggest difference between the 2 models proposed by the author. The results reported from author was by training the models while changing the learning rate, accuracy and F1 score was used as metric to evaluate the performance of the models, and no data features extraction or pre-processing was needed to do. With CNN-IMU the average accuracy obtained was 92.37% and the average F1 score was 0.92 while for model CNN-IMU-2 the average accuracy reached 92.77% and the average F1 score was .93.

In [23] features were extracted from the sensors data, by applying fast Fourier transform the authors extracted 12 features from time domain (six from acceleration and six from angular velocities) and 12 features from frequency domain (six from acceleration and six from angular velocities). To evaluate the performance of unsupervised learning methods for activity recognition, two kinds of experiments were carried out based on the knowledge of the number of activities (k). First, verification whether unsupervised learning is a reasonable way for activity recognition when k is known, and second, showed whether unsupervised learning can still be useful when k is unknown, indicating its potential to distinguish an arbitrary k. To evaluate the quality of the clustering algorithms accuracy and normalized mutual information was used. WEKA software was used to run the experiments. When k is known the author chose three clustering algorithms, k-means clustering (acc= 0.7198, NMI= 0.8670), mixture of Gaussian

(GMM) (acc= 1.0000, NMI= 1.0000), and average-linkage hierarchical agglomerative clustering (HIER) (acc= 0.7998, NMI= 0.9092). When k is unknown four clustering algorithms: k-means clustering, mixture of Gaussian (GMM), average-linkage hierarchical agglomerative clustering (HIER), and DBSCAN were chosen. Since the results for when k is known are presented in graphics it's hard to put them into numbers in this small review of the article.

In [24] the authors aimed to study was to examine whether machine learning algorithms trained on data of tri-axial accelerometer alone or with gyroscope could identify human physical activity, to do this the authors applied several algorithms such as Naive Bayes, LogitBoost, RF and SVM. To feed and train this model besides accelerometer and gyroscope data, data extraction was needed to be done, extracting features as mean, autocorrelation, power spectral density, spectral power, Entropy, Sum Power Det Coeff, Inter Quartile range, Spectral variance, Main frequency, Intensity, Zero-crossing rate, Skewness and Correlation coefficient. A comparison by accessing the performance of the different models through precision, recall and average accuracy was made, where was confirmed that by using accelerometer data with gyroscope better results are achieved. Overall, the RF algorithm obtained the better results with precision=0.99, recall=0.99 and an average accuracy of 92.29%, while LogitBoost obtained precision=0.99, recall=0.99 and an average accuracy of 89.69%, followed by SVM with precision=0.95, recall=1 and an average accuracy of 84.17% and finally Naive Bayes with precision=1, recall=0.97 and an average accuracy of 81.24%. These algorithms were implemented through WEKA software.

In article [25] the authors proposed a multi-class algorithm which consists of an ensemble of auto-encoders (EAE) where each auto-encoder is associated with a unique class. Here the authors conducted several experiments to compare the predictive performance of the EAE, with others CNN SoA approaches and with a previous developed model by the authors called EkVN model which consists of an ensemble of models (KNN, Very Fast Decision Tree (VFDT) and Naive Bayes), in the context of human activity recognition. The author in the comparison used the accuracy and time (train/test) of the models as decisive metrics. For the CNN results the author used the values presented on their respective investigation, but for the EkVN model the author had to implement it from scratch, thus extracted features such has standard deviation and Pearson Correlation were added to data. The author on EAE trained it and tested it in an online and in an offline stage only by using accelerometer data as input. EAE online was trained/tested during 209.9/36s and obtained 0.82 of average accuracy, the EAE offline had a duration on train/test about 209.9/31s and an accuracy of 0.75 and finally the EkVN model achieved an accuracy of 0.97 with a time of 79.3/1.1 in train/test.

2.4 DISCUSSION

2.4.1 (I) WHICH DAILY LIVING ACTIVITIES ARE RECOGNIZED AND FOR WHAT PURPOSE?

Every day, human beings carry out a huge number of different physical activities. These activities range from the simplest, such as walking, sitting, and lying down, to the most complex such as brushing our teeth, jumping rope or vacuuming the room. In the health sector, being able to recognize human activity automatically through computers has special importance because it allows to know the habits and health conditions of a patient, thus improving the ability of diagnosis and treatment by health professionals. In the literature reviewed, it was found that **most authors tried to develop their research on HAR for healthy people**. Only [10] and [24] tried to develop HAR models for people with PD. This fact corroborates the statement made by Kazemimoghadam et al [10] that, although the area of HAR received enormous attention from scientists, there is still a long way to apply these methods to people with mobility disorders.

It was observed that the most human activities commonly classified are **walking**, **climbing,** and **descending stairs** and **sitting**. **Standing** is also an activity slightly less represented [14],[15],[16],[17],[19],[20],[21],[22],[23],[24],[25] as also observed for activity of **lying** which was only studied in [17],[18],[19],[20] and [23].

Most of the selected articles refer to HAR systems in healthy people, being a need to extrapolate these findings to other field on medicine. Thus, **it is required to develop and customize HAR systems to recognize basic physical activities of people with PD**, as walking, standing, turning, upstairs/downstairs, sitting or lying, **but also more complex activities** like running, jumping, or carrying out weights.

2.4.2 (II) WHICH TYPE OF DATA WAS USED IN AI MODELS INPUT FOR HUMAN ACTIVITIES RECOGNITION SYSTEMS?

The crescent evolution of WS technologies has given rise to WS in the field of human health monitoring. This literature review verified what data these sensors produce, in which part of the body they should be placed, how many sensors need to be used and what is the experimental protocol for collecting meaningful data for HAR.

There were two major types of systems to perform HAR, one system used WS and the other used external devices such as cameras [10].

It was observed that **IMU**s were frequently applied to measure users' motion during motor tasks. IMUs are an ideal option considering the **low-cost, low power-consumption, portable, size, weight and easily accessibility and usability, not being technology invasive to users**. The other options such as reflective markers would entail greater monetary investment and would limit patient data collection to indoor environments due to the static nature of the system resulting from the need to use cameras for data collection. Regarding to the use of smartphones as a source of data, it has against it the fact that the data obtained is affected by its orientation and placement on the body and this hinders the processing of these same data into meaningful and interpretable outputs [26].

IMUs can integrate accelerometers, gyroscopes and magnetometer and produce kinematic-driven data from the segment which are coupled, such as 3D acceleration, angular velocity and orientation [27]. These sensors were mostly placed on the **legs**, **ankles** and **trunk varying in number between 3 and 7 IMU's used**. As read in [14], [15], [16], [17], [21], [23], [24] by using a smartphone containing **one IMU sensor it is possible to obtain satisfactory results** regarding daily activity recognition and according to [28] by **using one IMU on the lower back, it is also possible to obtain a complete analysis about the walking cycle**

In the selected articles it is found that mostly datasets used are public and are composed of data collected on healthy people, and mostly the data comes from smartphones. In the articles where the data came from people with PD [10],[24], few participants were included for data collection, 5 and 2 respectively. **Using data from a reduced number of subjects may raise problems in the reliability of the AI-based models' performance**, due to the possibility of overfitting not only in the training process but also in the validation process. Therefore, **there is a need to consider using data from a higher number of patients with PD** [29].

It was verified that on the datasets used, whether publicly available or built by the authors, there is **vague documentation about the protocols defined for the collection of information**. Also, it was observed that all datasets included data collection methods from different protocols. Thus, there is a need to standardize the protocols to capture the most significant activities able to provide information about patients' motion function.

2.4.3 (III) WHICH AI MODELS AND HOW THEY WERE IMPLEMENTED IN HUMAN ACTIVITIES RECOGNITION SYSTEMS?

HAR generally exploits time series data from inertial sensors to identify the actions being performed [14]. In recent years, due to their popularity, DL technologies, for instance CNN and RNN, have been successfully introduced in HAR applications [10]. What distinguishes these from the classical statistical ML methods is that **they can extract relevant features from raw data although this is achieved by adding more layers and nodes of classification which increase computational complexity** [16]**.** The classical statistical ML, such SVM or DT, are trained to identify different activities using handcrafted features which means that these approaches rely on human domain knowledge which might affect the models' performance [25].

On literature, it was observed that **many different DL architectures** have been proposed, such as CNN, RNN, AE, and all have in common the fact that they obtain better results in HAR than the classic statistical ML models. To evaluate the quality of the DL models, the authors chose to use as metrics the accuracy [14],[15],[16],[17],[18],[21],[22], [23],[24],[25] and the metric F1 scores [30],[19],[20],[22] . Particularly, **F1 scores were used on the studies which included unbalanced datasets** in the number of records of each classified activity. Thus, it is required to use protocols which provide data repeatability to overcome this data collection asymmetry. Besides F1 scores, accuracy, recall and MCC can be also considered to evaluate the performance of the proposed model.

One of the applications of a **HAR system for PwP** is its integration in motor assistance devices, which can, for example, unlock FoG. Therefore, it is necessary that HAR performs in real time, but in the literature only Chen Xu et al [20] successfully performed tests with online implementations. Thus, after the models are trained and validated, it is recognized the need to use HAR systems in **real time** to be integrated in **motor assistance devices**.

It was noticed that all the **models proposed for HAR limit themselves to recognize simples' activities such as walk, standing or sitting and are not able to be flexible enough to go beyond and reach further into more distinguishable continuous monitoring of human behavior**, as segment between side-by-side activities or activities transitions, such as right/left turning or standing/walking or standing/sitting/getting up from a chair. To solve this problem the application of a **tree diagram**, suggested by Nguyen et al [31], enables the classification between dynamic/static sub-activities without requiring high power computational consumes and allows more scalability to more precise activities such as raise left arm or raise one finger. Although this tree diagram might be a solution,

some concerns arise when happens to fail its first step of classification because it means that the last classification step will always be a wrong classification.

2.5 CONCLUSION AND FUTURE DIRECTIONS

From SoA review, it was concluded that **HAR focused on PD was not very studied**, proof of that is a total absence of public HA dataset composed by data collected from PD patients. Current approaches for evaluating the motor functions on individuals with PD are limited to observer-based and self-report methods [10]. **IMUs were the sensors most chose** by authors to provide data input signals on their HAR models. There is no standard method for data collection and for sensors placements. Except for the cases where smartphones were used, a great number of sensors are used for data collection which is impractical for day-to-day usage by the patients.

In respect of DL models, **CNN and LSTM networks were the most popular models implemented**, as also the usage of accuracy as metric to evaluate the performance of the recognition models. At the training phase of this models the majority of the research papers [16], [18], [19], [20], [22] opted for **not using feature extraction in the learning process**.

It was noted an **absence of a standard framework to perform HAR**, being the proposed models limited to identify basic human activities. Also, the proposed frameworks were not capable of stepping further into autonomous human monitorization, not being able to, for example, recognize if a person is walking slow or faster or even if is performing a second activity while is laying. Achieve this degree of HAR will be beneficial for the most varied scientific areas.

Table 2-4 summarizes the identified limitations regarding technological, adopted strategies, and validation methodology issues and it is also provided guidelines for their mitigation. Therefore, a systematic approach will be followed to identify the requirements of the system, from the point of view of the user and technologies, considering the limitations identified in the literature review, allowing to move on to the next dissertation tasks.

Table 2-4 - Limitations and guidelines to overcome them.

Limitation	End user requirements	Guidelines
Datasets composed only by data from healthy subjects	Software capable of monitoring PD patients	Develop a dataset with PD patients' data
Non clear body configuration of WS	Portability, comfort, easy set-up	Find a trade-off between the number/location of sensors without losing significant data

Poor data collection on home-based conditions	Personalized treatments	Perform experimental tests on home-based scenarios
No assessment of HAR usability	Acceptability of the device	Include the users' opinion in the development of the proposed solution and assess its acceptability and usability
Flexibility and scalability of DL models	Improve and deepen the HAR	Implement a hierarchical framework capable with minor changes perform in depth continuous monitorization

3 Solution Overview

3.1 INTRODUCTION

Aiming to provide a closed, continuous, and objective clinical support tool using WS data as input for AI-based models, to reach the main goal of this dissertation, two main objectives were outlined:

1. Build and develop a database with kinematic data from daily life physical activities performed by PD subjects.
2. Develop AI framework based on deep learning (DL) algorithms capable of accurately recognize HAR performed by PD patients

To this end, from a SoA review, it was defined the achievements and limitations of current related works, being able to clarify the steps to follow in this dissertation. A critical literature review was accomplished enabling to highlight the technological and end-users' requirements to consider on the design of the proposed solution. Thus, the conceptual overview of the proposed solution in this dissertation includes three main aspects:

1. ***Human activities to recognize in PD domain***. The activities defined to collect kinematic data from are referred in Table 3-1: walking, stand, stand to sit, sit to stand, stand to lying, lying to stand and 180° turning. Although all activities add important information to monitor patients motor function, walking and turning recognition are key activities in a motor assistance device, since scientific community observed these situations as main triggers to walking-associated disabilities.

Table 3-1 Human activities to be recognized by our proposed DL classification framework.

Human activities	Monitoring	Assistance
Standing still	✓	
Siting on chair	✓	
Get up from chair	✓	
Lying on bed	✓	
Get up from bed	✓	
Walking	✓	✓
180° turning	✓	✓

2. ***Data input.*** To be more comfortable and less invasive to patients, a small number of body sensors is used to collect data. **One inertial measurement unit** (IMU) that integrates an accelerometer and a gyroscope, capable of capturing lower trunk acceleration and angular velocity data, will be used to collect data from the patient. The **kinematic raw data captured by the IMU sensor is to be used directly by DL model proposed by this dissertation**, **reducing the extra computational power used in the features extraction implemented by others AI models**. Advantageously, **this IMU sensor location can capture data from a complete walking cycle with a single sensor**, while sensors configurations on lower limbs require more sensors to capture the same data. Indeed, inertial data have been extremely used on the literature, since it includes kinematic data about users' motion and use low power-computation, low-cost and portable devices. **Since there are no public datasets available with kinematic data from IMU pertaining to daily activities performed by PD patients, this dissertation built one**. To do so, a collection protocol was defined and implemented as described on Appendix A.

3. ***HAR AI-based model.*** The proposed framework consists of one AI DL model. This Convolutional Neural Network (CNN) model based is responsible for classifying human activities such as: *walking*, *standing still*, *sit on chair*, *get up from chair*, *lay on bed*, *get up from bed* and *180° turning*.

Besides the two main objectives previously mentioned, a **secondary objective** was delineated during the dissertation implementation process, since it was observed the importance of detecting walking**, Initial foot Contact (IC)** and **Final foot Contact (FC)**, to provide on-demand motor assistance and estimate walking-associated metrics. This **complementary objective** focused on the **recognition of walking events** using AI

Human Walk IC/FC Detection (HW-IC/FC) AI-based model. This AI DL model is based on Bidirectional Long-Short Term Memory (BiLSTM). BiLSTM are upgraded version of the regular Long-Short Term Memory (LSTM) and are efficient in data sequence processing. This dissertation solution can recognize IC and FC events from the PD patients walk,

This dissertation frames in project research titled by +SENSE: Sensory biofeedback devices for patients with PD, which aims towards high-tech solutions to mitigate motor symptoms in PD. Therefore, with this dissertation it is expected to contribute with a HAR framework capable of recognizing daily

activities such as walking, sit and get up from chair, lay and get up from bed and to recognize when a patient performs a 180° turn. The complementary HW-IC/FC framework is capable of recognizing walking events, namely IC and FC. Both frameworks will be used to monitor PD patients' motor function, and to be integrated on +SENSE assistance devices. Further description about +Sense project functioning, how this dissertation fits in +Sense project and how the HAR and HW-IC/FC frameworks were developed in this work will be presented in the following chapters. To process and transform the data, Python language with its multiple libraries dedicated to data handling, exploration and visualization will be used. Framework TensorFlow with its libraries will also be utilized to implement the HAR and HW-IC/FC DL models.

3.2 PROPOSED SOLUTION FRAMED INTO +SENSE

This dissertation is integrated into the **+Sense project**. +Sense presents front-end high-tech solutions based on wearable biofeedback devices which rely on acquisition, interpretation, and feedback of PD patients' sensorimotor information. By promoting patients motor autonomy, +Sense projects research envisions to improve patients' life quality. There are four +sense modules, as shown in Figure 3-1: (1) +sBiofeedback; (2) +sMotion; (3) +sC-support and (4) +sImmersive. This dissertation **contributed** to the +sMotion module by **building a framework capable of transforming the received data and using it to classify the daily activity (HAR) and walking IC/FC. HAR recognition enable to automatically detect which activity patients are performing and assess their performance during motor trivial quotidian tasks.** Thus, a closed follow-up and assessment of patients can be achieved. Also, this framework by having the capability of recognizing walking IC/FC, also has the potential to assist PD patients in their locomotion's. The application of this dissertation proposed framework indirectly contributes to +sBiofeedback module which provides on-demand motor assistance biofeedback tools. Further, the estimation of walking IC/FC complements +sMotion for automatically estimate walking-associated metrics, such as velocity, step length/time, cadence, walking variability or walking asymmetry.

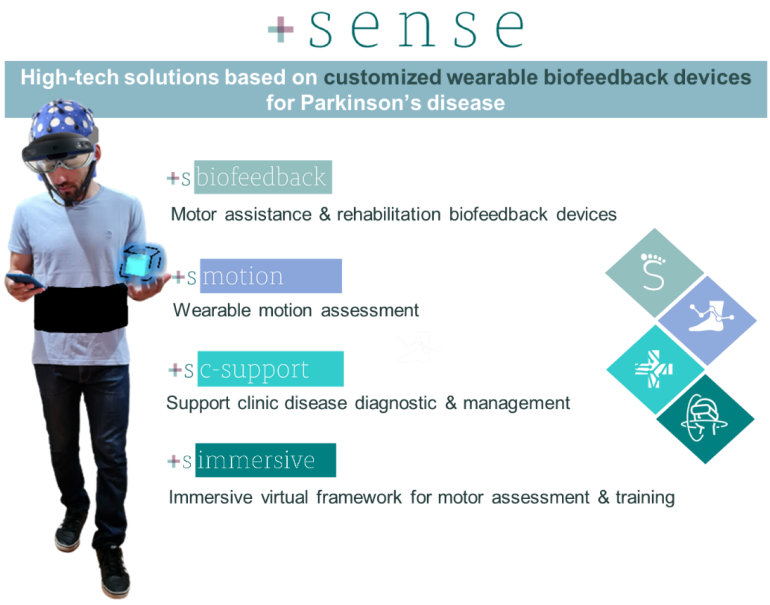


Figure 3-1 +Sense project modular composition.

3.2.1 +sMotion: Combination of Kinematic-drive Data from Wearable Sensors with AI for motion assessment in PD

+sMotion module includes a wearable walking analysis LAB which is a part of an instrumented waistband (WBD), comprising the 1) Sensory Acquisition Unit; 2) Processing Unit; 3) Data Storage Unit; 4) Mobile APP, as depicted in Figure 3-2. The wearable walking analysis LAB is **placed close to the center of mass area**, being capable of capturing an entire gait cycle, this positioning confers stability to the device and adds more comfort to the user bye being discrete and light. Is responsible for a non-invasive kinematic data acquisition and monitoring lower trunk inertial signal, post-processing walking analysis and walking-associated metrics estimations of PD patients.

Figure 3-2 +Sense hardware and software setup.

The **+sMotion module sensory acquisition relies** on the use of only **one** MPU-6050 **Inertial Measurement Unit** to acquire **acceleration and angular velocity data**. The computing unit comprises a STM32F4-Discovery that processes the acquired kinematic data in real-time to segment the walking cycle. The acquired inertial data and identified events are saved in the Data Storage Unit, a OTG USB driver. The Mobile APP is an Android APP that wirelessly communicates with the processing unit, via Bluetooth, enabling to start/stop data acquisition, control operability settings and plotting the acquired data.

To further extend the +sMotion module utilities, this dissertation contributed by taking advantage of the already existing data sensory acquisition hardware and Mobile APP software, with two AI-based DL models, one for human activity recognition (HAR) and another to human walk IC/FC detection classification. +Sense device collects kinematic data that will be processed by DL HAR and HW-IC/FC frameworks, to support the clinical expert with relevant metrics regarding the PD patient, exemplified by Figure 3-3.

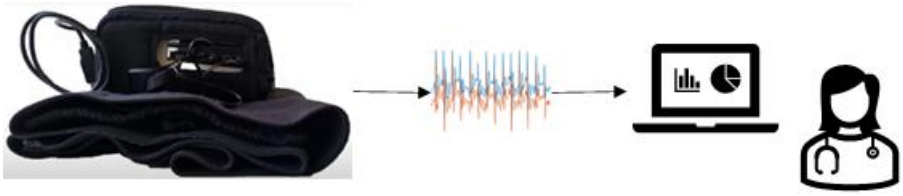


Figure 3-3 Scheme about how this dissertation was framed in +Sense. Data collection using wearable motion LAB and AI analysis provided with this dissertation.

3.3 Conclusions

This dissertation was developed with the objective to automatically detect daily activities performed by people with PD using kinematic data through the implementation of an AI DL model framework. A complementary AI DL model was also built to detect IC/FC during PD patients walk using kinematic data. It is expected that these different frameworks to be a complement to +Sense project.

PD patients' assessments depend exclusively on presential consults or self-made report by patients. Usually, medical consults are sparce in time and the disease keep on evolving, often the drugs prescribed to lesser the symptoms of the disease becomes less effective. Auto-reports made by the patients of their day-to-day life are usually incomplete due to complacence or incapability of the patient, becoming in an unreliable assessment tool.

This dissertation also defends that the existence of a real-time walking device is important, since walking impairments are one of the majors' symptoms of PD. By complementing the main objective of this work with AI model capable of human walk IC/FC events detection, it is expected to contribute with an alternative to the already existing solution in +sBiofeedback module.

Having an automatic All-in-One device capable of monitoring and support the PD patients in its daily life is crucial and game changer.

Although in SoA research papers mentioned the usage of smartphones to collect data from the subject, in daily life that isn't an option because it demands to always use it near the body of the subject. This handicap would affect one of the purposes of +Sense project which is the empowering of the autonomy of PD patients, especially the patients who like to wear their smartphones on their purses.

As an alternative to smartphones in data collection, the SoA research proposed the usage of IMU sensors. The quantity of the sensors to be used ranged between 3 to 26. This high number of sensors might be impractical and uncomfortable to be used in the different everyday activities all day.

Another critical issue mentioned in SoA is the body part positioning of the sensors, either smartphone or IMU device. Because different sensor body positioning generates distinct data for the same activity it will affect the performance of the device in its capability of recognizing human activity.

+Sense is the answer to all handicaps mentioned above. With an elastic waistband adjustable to any person waist, it can be easily concealed beneath clothes, delivering total freedom of movements and full autonomy in daily activities. Since it imbeds one IMU sensor, that is always placed at the lower trunk when the WBD is placed on, the problems caused by placing the sensor in different/wrong places and the sensor being uncomfortable are also solved.

As we will see later in this dissertation, **good results** using AI DL model for the **recognition of PD patients' daily activity and walking IC/FC were obtained**, being the **kinematic data collected through the +Sense WBD that is equipped with one IMU located at the lower trunk body position**.

4 Deep learning Framework for Human Activities Recognition

4.1 Introduction

As found in SoA research, generally, DL models can achieve better results in recognizing human activities through kinematic data. As such to build a DL framework capable of recognizing daily activities performed by PD patients it is this dissertation main objective. To do so, a series of steps needed to be taken:

- know what activities were relevant to be monitored;
- what type of data should be collected;
- how the data should be processed;
- how the pipeline for the DL model training and evaluation would be implemented.

Always with the support of the literature, these steps will be presented and discussed throughout this chapter.

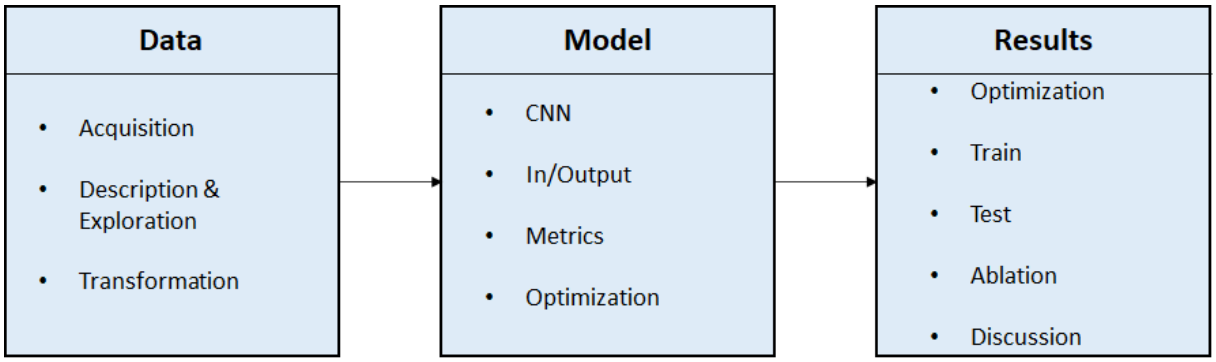


Figure 4-1 HAR schema layout solution.

4.2 Data Input

4.2.1 Data acquisition

Several public datasets for activity recognition were found publicly available, like PAMAP2 [32], Opportunity [33] and UCI HAR dataset [34]. Nevertheless, these never presented all the required configurations needed to accomplish the aim of this dissertation, since most of them used smartphones to collect data, and all of them collected data from healthy people. As we will see later on this dissertation, the data signal collected from a PD patient is noticeably different from a healthy person. Therefore, a custom dataset was created. Data acquisition was conducted under the ethical procedures of the Ethics

Committee in Life and Health Sciences (CEICVS 147/2021), following the Helsinki Declaration and the Oviedo Convention. All participants gave their informed consent to be part of the study. Data was collected at the Hospital of Braga – 2CA Braga Academic Clinical Center. A complete protocol is available at Appendix A.

4.2.1.1 PARTICIPANTS

Eighteen participants with PD (ten males and eight females) were recruited and accepted to participate in this data collection. A list of inclusion criteria was outlined to conduct the experimental data collection.

Table 4-1 Demographic and clinic characteristics of study participants.

Characteristics		**Participants (N=18)**
Age [years] (mean ± SD)		67.89 ±8.72
Height [cm] (mean ± SD)		166.21166.21 ±7.47
Weight [Kg] (mean ± SD)		70.6870.68 ±11.63
Gender	Female	N=8
	Male	N=10
H&Y	(mean ± SD)	2.322.32 ±1.13
	I	N= 6
	II	N= 4
	III	N= 4
	IV	N= 4
UPDRS-III	(mean ± SD)	36.63±24.40
	Low (1-12)	N=8
	Mild (13-22)	N=6
	High (≥23)	N=4
SD: standard deviation.		

4.2.1.2 ACQUISITION INSTRUMENTS

The volunteers performed the daily physical activities while using the WBD that contains one MPU-6050 Inertial Measurement Unit to acquire acceleration and angular velocity data. This equipment was used in all trial performed by the PD patients, Figure 4-2 presents the used WBD.

Figure 4-2 Waistband containing the MPU-6050 Inertial Measurement Unit sensor (WBD).

4.2.1.3 ACQUISITION PROTOCOL

Considering the literature found in SoA and hearing medical staff opinion, eight different physical daily activities were recognized by de DL model. Those activities are:

- Walk
- Turn 180° degrees while walking
- Sit on chair
- Get up from chair
- Lie on bed
- Get up from bed
- Walk upstairs
- Walk downstairs

Due to the lack of physical conditions at 2CA Braga Academic Clinical Center department, the activities of climbing stairs and walk downstair were excluded from the trials, but to the activities of turning while walking were differentiated to distinguish between turn right or left while walking. As such, a final protocol was made to collect data of the following physical activities:

- Walk
- Turn 180° degrees right while walking
- Turn 180° degrees left while walking
- Sit on chair

- Get up from chair
- Lie on bed
- Get up from bed

Each activity trial was performed three times by every study participant, and the execution times of the activity were manually recorded for later assistance in labelling the data, Figure 4-3. A comprehensive scientific protocol is available at Appendix A.

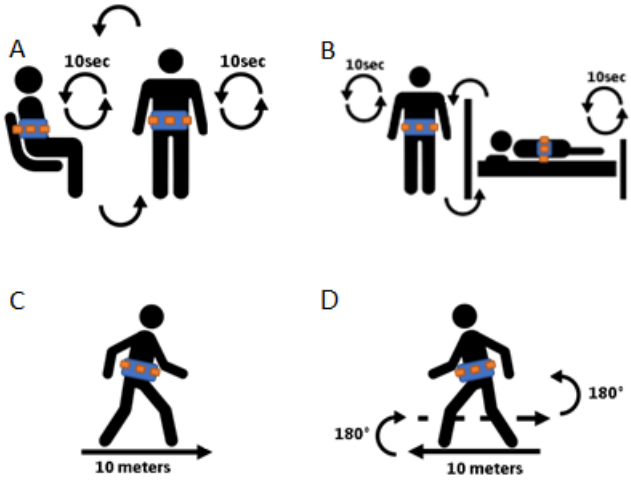


Figure 4-3 Trials performed by the volunteers. A – Sit and get up from chair; B – Lay and get up from bed; C – Walking; D - Walking and perform 180° turn.

4.2.2 DATASET DESCRIPTION

To build a DL model, possessing a good knowledge of the kinematic data that we are working on is very important, and so an initial data exploration was done.

The Parkinson's Disease Human Activity Recognition Database is a collection of eighteen PD patients' data. Each patient performed three trials of eight different activities:

- Walk;
- Turn 180° degrees right while walking;
- Turn 180° degrees left while walking;
- Sit on chair;
- Get up from chair;
- Lie on bed;

- Get up from bed.

The data was collected using the WBD containing one MPU-6050 IMU, which collected tri-axial data for acceleration and angular velocity at 100 Hz frequency, meaning 100 observations of the activity were registered per second. Each trial datafile contains eight independent variables:

- time_sampling: internal time of the processing unit responsible for acquiring the IMU data
- Sample: id number of the sample
- acc_x: acceleration value in x axis
- acc_y: acceleration value in y axis
- acc_z: acceleration value in z axis
- gyr_x: angular velocity value in x axis
- gyr_y: angular velocity value in y axis
- gyr_z: angular velocity value in z axis

Figure 4-4, Figure 4-5, Figure 4-6, Figure 4-7, Figure 4-8, Figure 4-9 express how different PD patients activities kinematic collected data are different from Healthy persons performing the same activities.

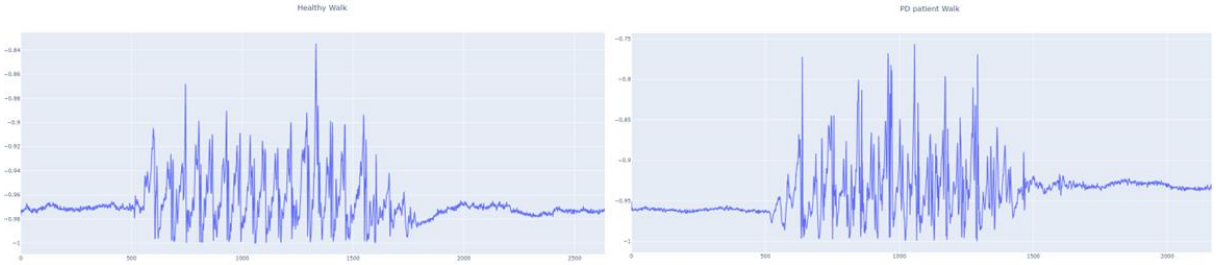


Figure 4-4 Comparison between Healthy person *walking* and PD patient *walking* activities. Only Acceleration values from X axis are displayed for better understanding.

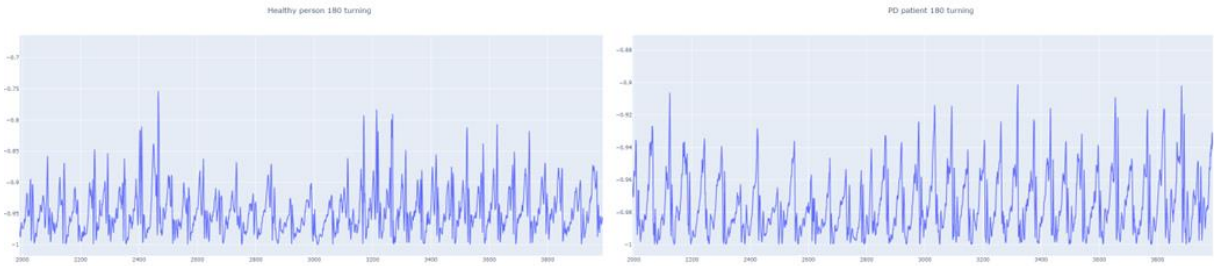


Figure 4-5 Comparison between Healthy person *180° turning* and PD patient *180° turning* activities. Only Acceleration values from X axis are displayed for better understanding.

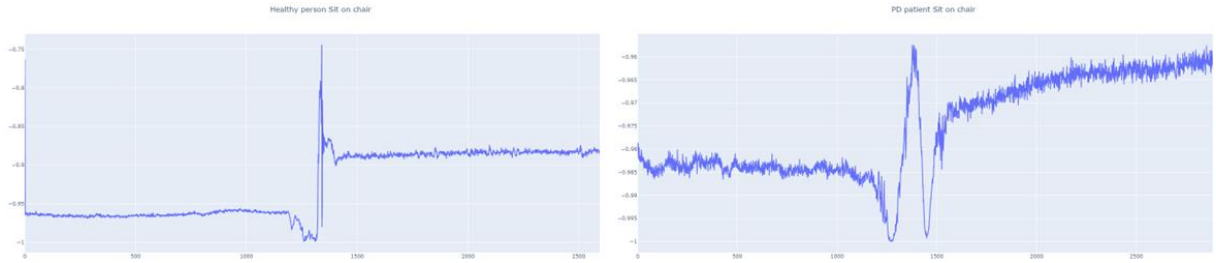


Figure 4-6 Comparison between Healthy person *sitting on chair* and PD patient *sitting on chair* activities. Only Acceleration values from X axis are displayed for better understanding.

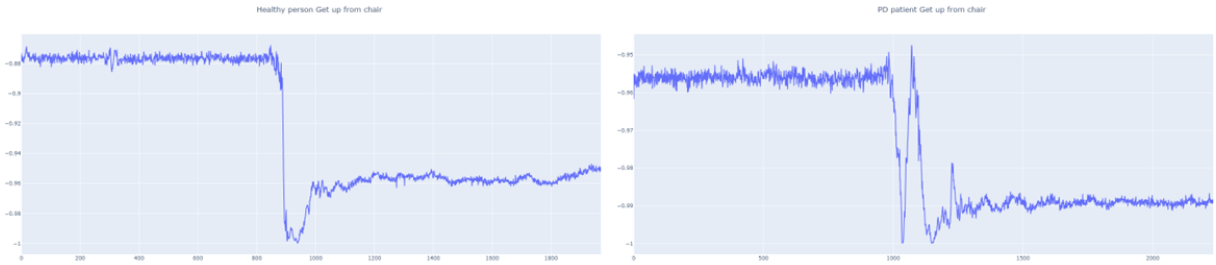


Figure 4-7 Comparison between Healthy person *get up from chair* and PD patient *get up from chair* activities. Only Acceleration values from X axis are displayed for better understanding.

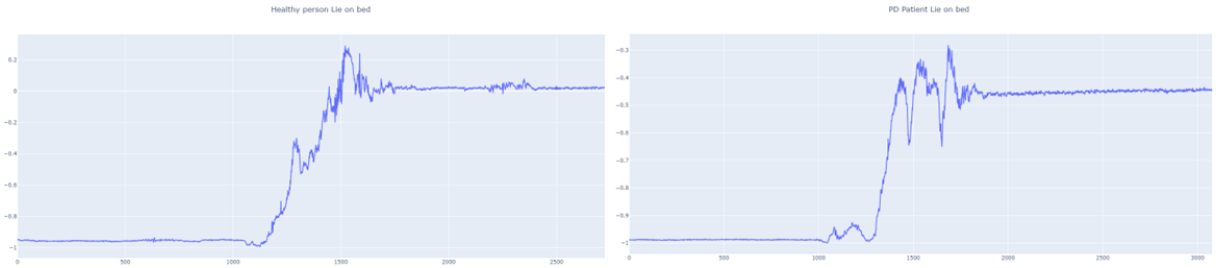


Figure 4-8 Comparison between Healthy person *lay on bed* and PD patient *lay on bed* activities. Only Acceleration values from X axis are displayed for better understanding.

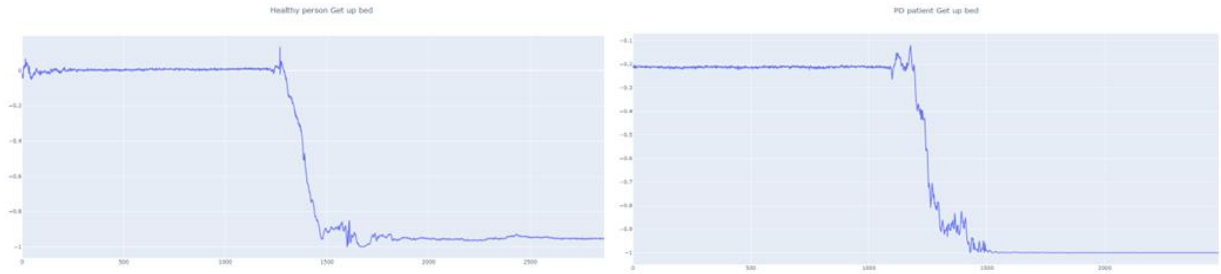


Figure 4-9 Comparison between Healthy person *get up bed* and PD patient *get up bed* activities. Only Acceleration values from X axis are displayed for better understanding.

4.2.3 DATA INPUT PREPARATION

Before begin training, validating, and testing the proposed DL model of this dissertation, several steps were needed to take based on literature, such as: data verification, data labelling and data splitting, were needed to take.

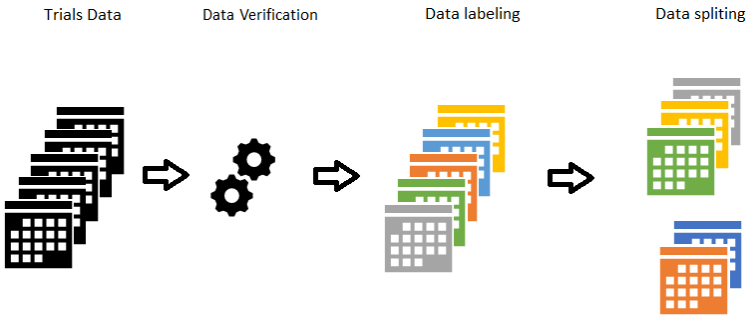


Figure 4-10 Human activities recognition data input handling pipeline.

4.2.3.1 DATA VERIFICATION

To assess if the data files collected contained corrupted data, a manual search in all dataset files were accomplished, as depicted in Figure 4-11. It is notable that the corrupted data was always found at the end of data file, being easily removed for further AI process. An example of corrupted kinematic data collected by the IMU sensor is present in Figure 4-11, where the signal selected by the black frame corresponds to a movement outside of the trial purpose.

Figure 4-11 Patient trial with corrupted captured kinematic data.

4.2.3.2 Features Extraction

Traditional machine learning approaches use handmade features to train their models [10], [15], [21], [22], [24] recent advancements in neural networks allow for automatic feature extraction. DL supervised training methods [10], [17], [18], [19], [20], [22], [25] were proved to be able to obtain better performances than the traditional machine learning models without the need for additional features beyond the raw basic data. As such **no handcrafted features were added to the training dataset**.

4.2.3.3 Data Labelling

The data labeling process was done manually based on synchronized notes of activities start/stop taken during data acquisition. During data acquisitions, a timer ran at the same time the WBD was activated to collect patients' kinematic data and a researcher registered the time in which each participant started and stopped the execution of an activity. Data labeling used these notations and to avoid bias, during data acquisitions, the same researcher accomplished these annotations and data labelling.

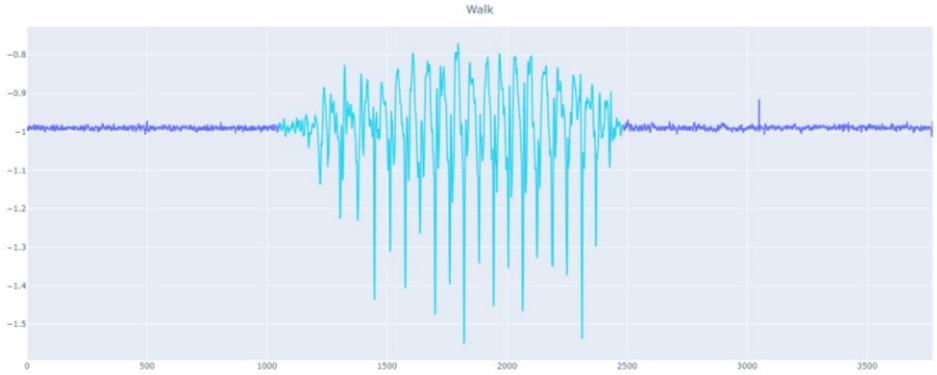


Figure 4-12 With a light blue color the PD patient *walk* kinematic data labelling and with a dark blue PD patient Standing still labelling kinematic data. Only Acceleration values from X axis are displayed for better understanding.

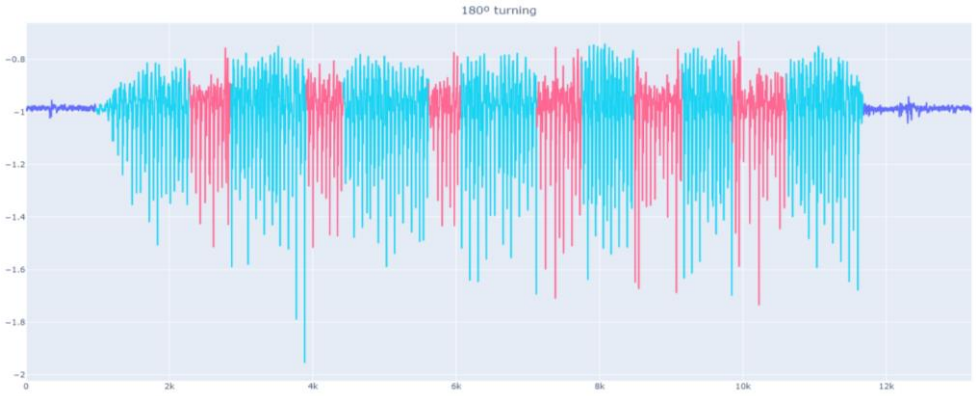


Figure 4-13 With red color the PD patient *180º turning* activities kinematic data labelling. With light blue color are labelled as *walking* activities and with dark blue color PD patient *standing still* kinematic data labelling. Only Acceleration values from X axis are displayed for better understanding.

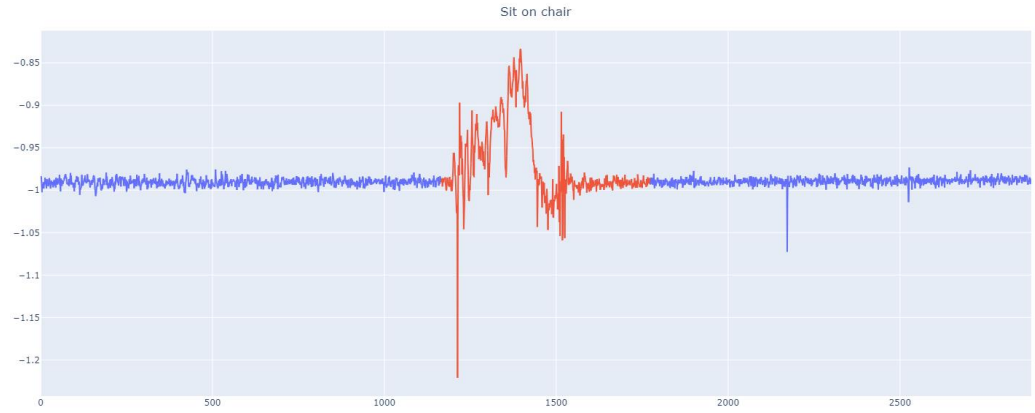


Figure 4-14 With red color the PD patient *sitting on chair* activity kinematic data labelled. With dark blue color PD patient *standing still* activity kinematic data labelling. Only Acceleration values from X axis are displayed for better understanding.

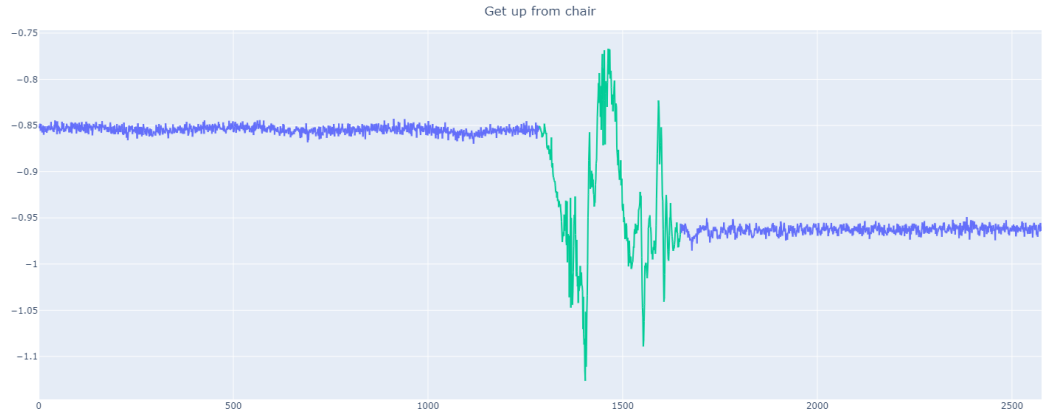


Figure 4-15 With green color the *get up from chair* activity labelled. With dark blue color PD patient *standing still* activity labelling kinematic data. Only Acceleration values from X axis are displayed for better understanding.

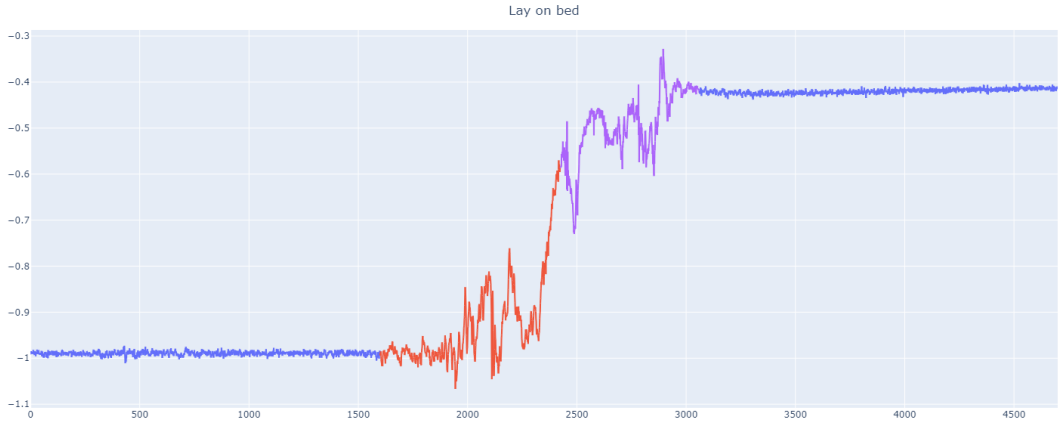


Figure 4-16 With red color the *sitting on chair* activity labelled. Pink color represents the Lay on bed activity and dark blue color *standing still* activity labelled kinematic data. Only Acceleration values from X axis are displayed for better understanding.

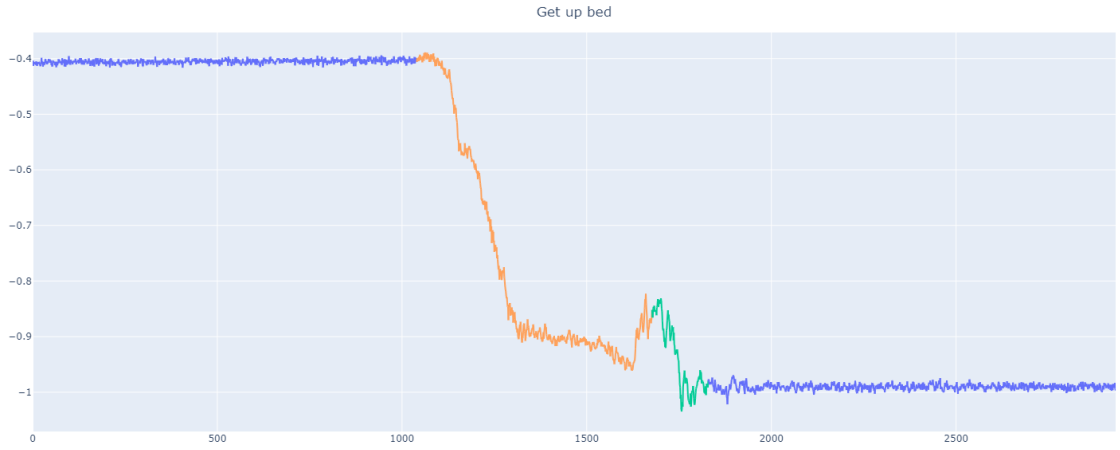


Figure 4-17 With yellow color the labelled *get up from bed* activity. The green color represents the *get up from chair* activity and dark blue color *standing still* activity labelled kinematic data. Only Acceleration values from X axis are displayed for better understanding.

With this choice of labelling, the activities to be recognized by the DL model proposed by this dissertation are enumerated at Table 4-2.

Table 4-2 Daily activities label identification.

Label identification	Activity
0	Standing
1	Sit on chair
2	Stand from chair
3	Lay on bed
4	Get up from bed
5	Walk
6	180° turning

4.2.3.4 Data Splitting

The whole dataset was divided into four different individual datasets. More specifically, 20% was used as test dataset (3 patients) for the proposed model final evaluation, 80% (15 patients) of the data was used in the fine-tuning stage of the proposed model. This 80% of data is further divided into 80% (12 patients) training and 20% (3 patients) validation datasets with the purpose of training and validating the final model using the best hyper-parameters found in fine-tunning stage phase. The testing dataset is only used once as unseen data for a final assessment of the model performance.

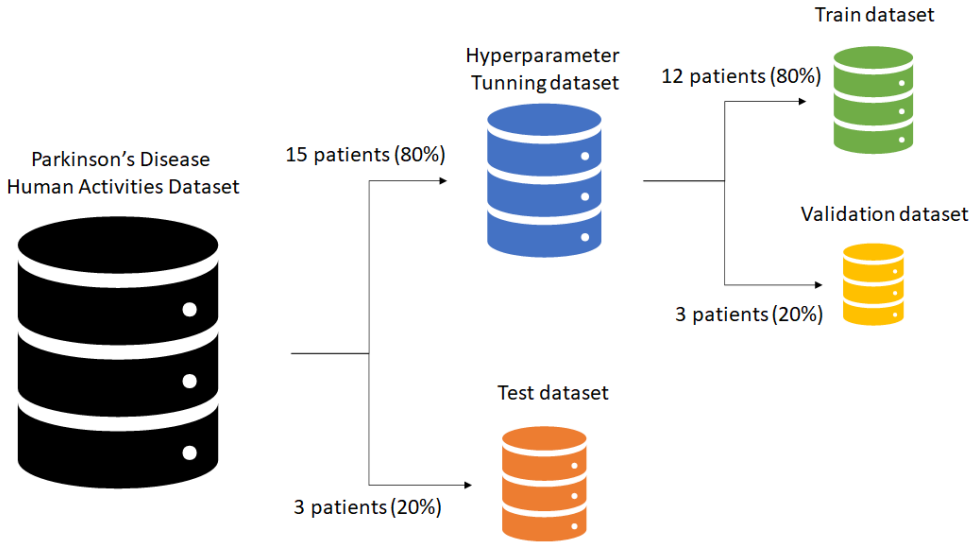


Figure 4-18 Schematic representation of the dataset splitting for the different HAR DL model development stages.

After Data Splitting a **Minimum-Maximum (MinMax) data normalization was applied**. **Variables measured** at **different scales do not contribute equally to the model fitting** and model learned function and might end up **creating a bias**, thus normalization is a needed procedure.

Although many methods exist for scaling data, on the chosen SoA literature only in [20], [22], [23] it was mentioned that the data was scaled to fit [0,1] range, and such this dissertation adopted this concept. **MinMax data normalization was done feature-wise in an independent way.**

4.3 Model Pipeline

This dissertation proposed DL algorithm for the recognition of daily activities performed by PD patients is a Convolution Neural Network (CNN) DL model based on Jason Brownlee proposed architecture for HAR [35]. Model pipeline included model parameters optimization through grid search process, model validation using the best parameters found in grid search process and a final performance evaluation of the model using unknown patient kinematic data. The following sub-chapters will present and describe the used model, its hyperparameters, its input/output data formats and the metric used to judge the model performance.

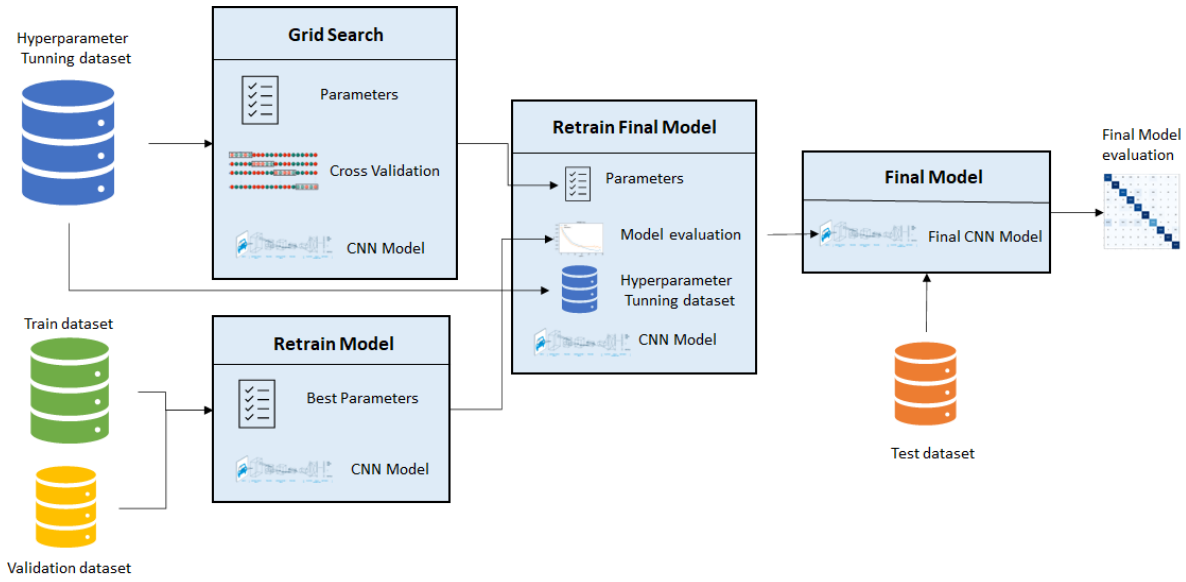


Figure 4-19 Model Pipeline Schema representing the development of the final HAR DL model for PD patients' solution.

4.3.1 Convolutional Neural Network

Convolutional Neural Network (CNN) is a class of neural Network that can process data in a grid-like way. Usually is very associated with image processing but it can also process IMU sensor kinematic data. According to [36], CNNs are comprised of three types of layers. These are convolutional layers, pooling layers and fully connected layers. When these layers are stacked, a CNN architecture has been formed.

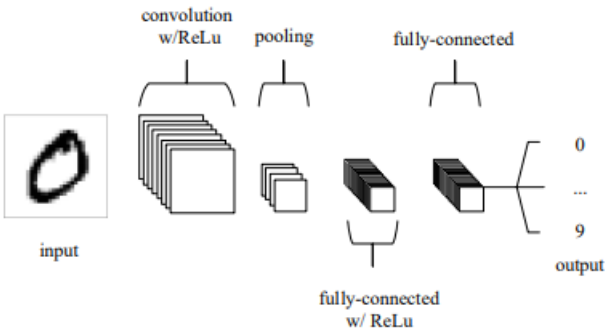


Figure 4-20 Example of a basic CNN architecture.

The convolutional layer input, through the calculation of the scalar product between the neuron weights and the region connected to the input data, is responsible of determining the output to be utilized by the following layer.

Is common this convolutional layer be followed by a pooling layer. This pooling layer will perform a down sampling along the spatial dimensionality of the given input, further reducing the number of parameters within that activation.

The fully connected layers will then perform the same duties found in standard neural networks and attempt to produce class scores from the activations, to be used for classification.

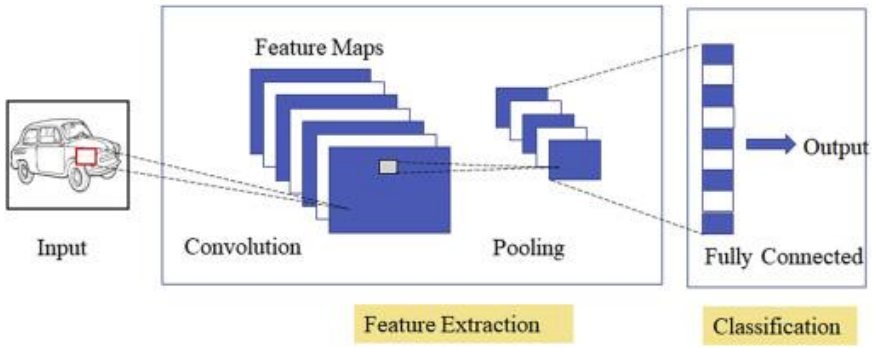


Figure 4-21 CNN example of functioning through convolutions, down sampling with pooling and finally the classification of input data.

CNN architecture has multiple parameters, but three parameters can be considered key to its performance: number of filters, kernel size and pooling size. CNN use several different filters to convolve over a given data spatially to detect features. These filters learn to capture spatial features from the data based on the learned weights through backpropagation. Stacked layers of filters can than turn a given data into a highly abstracted representation of the original data, easier to predict the outcome. The kernel size dictates the size of the filter, often kernel and filter are mentioned as being the same thing. A filter is a matrix of weights with which we convolve the input, measuring how close a patch of input resembles a feature. Smaller filters collect as much local information as possible, bigger filters represent more global, high-level, and representative information. The pooling size refers to how much the data should be down sampled, reducing its dimensionality by keeping the sub-regions activated features binned.

4.3.2 Proposed Convolutional Neural Network Model

Many CNN models were experimented, but with the collected data the best results were obtained by Jason Brownlee CNN architecture for HAR [35].

The model is composed of two 1D CNN layers, followed by a dropout layer, rate 0.1, for regularization, then a pooling layer. According [35] is common to define CNN layers in groups of two, because gives the model a better chance of learning features from the input data. The dropout layer is intended to help slow down the learning process of the CNN layers. The pooling layer reduces the learned features to the most essential elements. After the CNN and pooling, the learned features are flattened, by a Flatten layer, to one long vector and pass through a fully connected layer before the output layer

used to make a prediction. The fully connected layer works as a buffer between the learned features and the output with the intent of interpreting the learned features before making a prediction. For this model, we will use a kernel size of two, **to optimize the network Adam version of stochastic gradient descent algorithm** is used, and the **categorical cross entropy loss function** is applied since we are learning a multi-class classification problem.

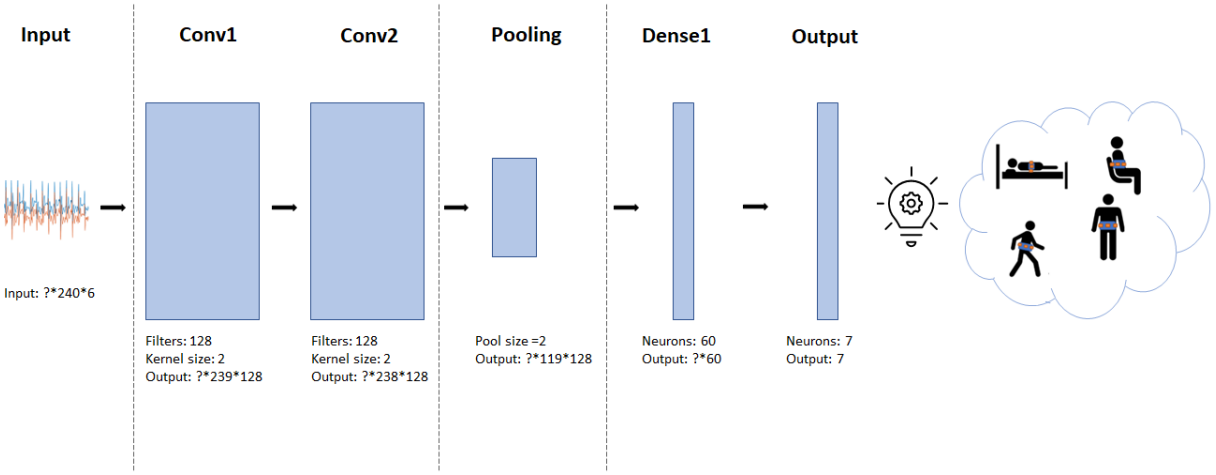


Figure 4-22 Proposed HAR DL model architecture.

4.3.3 Input and Output Data Format

The model requires a three-dimensional input with [n° of samples, timesteps, n° of features]. Meaning, each input sample is one window of time that will have a pre-determined number of timestep, and a timesteps has a fixed number of variable or features.

The output for the model will be a six-element vector containing the probability of a given window belonging to each of the seven activity types.

These input and output dimensions are required when fitting the model, and we can extract them from the provided training dataset.

Different sizes of input windows were experimented on, Table 4-4, but all had **50% overlapping data**. The overlap size was determined by the baseline model authors [35]. Overlapping time windows are intended to handle transitions more accurately. The different sizes of input windows tested were influenced by the time duration of the physical activities. In this dissertation a DL model based on Convolutional Neural Network algorithm was made, as the algorithm works, in a roughly way, by performing convolutions over sliding windows of data. Knowing an approximated optimal size of those windows should help on finding the best parameters of the DL model. Minimum, maximum, and mean size of the sequence for each activity present on the dataset were calculated, resulting on Table 4-3.

Table 4-3 Mean, minimum, maximum length of each daily activity.

Activity	Min	Mean	Max
Standing still	459	1770.872	4184
Sitting on chair	18	388.2857	1442
Get up from chair	116	303.6	1010
Lay on bed	118	512.3208	1470
Get up from bed	188	481.4259	1477
Walk	470	1316.937	5487
180º turn	231	512.2857	3507

Based on the mean, minimum and maximum length calculated for each daily activity, four different size window input were experimented on, as indicated in Table 4-4.

Table 4-4 Different size windows input experimented on.

Size window Input			
5	56	112	240

When fitting the model, it is also required a two-dimensional input [nº of samples, label]. Meaning, each sample have the corresponding label. The **label is in one-hot-encoding** format and chosen by selecting the most frequent label present in each sample window.

4.3.4 Model Evaluation Metric

In general, the prevalent assessing metric found in SoA research was the Accuracy, but SoA research papers where Neural Networks models were implemented and evaluated, F1 score metric was on an even foot against Accuracy metric. For a better result visualization commonly a Confusion Matrix is used, and from this instrument it is easier to calculate some of the most frequent evaluation metrics, such as F1 Score, Precision, Recall and Accuracy.

Table 4-5 Confusion Matrix.

		Predicted	
Actual		Positive Class	Negative Class
	Positive Class	True Positive (TP)	False negative (FN)
	Negative Class	False Positive (FP)	True negative (TN)

Precision: It is implied as the measure of the correctly identified positive cases from all the predicted positive cases. Thus, it is useful when the costs of False Positives is high.

$$Precision = \frac{TP}{(TP + FP)}$$

Recall: It is the measure of the correctly identified positive cases from all the actual positive cases. It is important when the cost of False Negatives is high.

$$Recall = \frac{TP}{(TP + FN)}$$

Accuracy: One of the easiest metrics to calculate, it is the measure of all the correctly identified cases. It is most used when all the classes are equally important.

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

F1-score: This is the harmonic mean of Precision and Recall and gives a better measure of the incorrectly classified cases than the Accuracy Metric.

$$F1\ Score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)}$$

According to Hossim et al [37] Accuracy metric is the measure of all the correctly identified cases, while F1 score metric is the harmonic mean of Precision, measure for correctly identified positive cases from all the predicted positive cases, and Recall, measure of the correctly identified positive cases from all the actual positive cases [38], giving, in the perspective of this work, a better measurement of the incorrectly classified cases than the Accuracy metric. Since this work is using an unbalanced dataset, to train the suggested model, **F1 score is used as the main measure to analyze the performance quality**.

4.3.5 FINE-TUNNING STAGE

To obtain better results from the initial chosen model, a common process in deep learning called Grid Search was implemented. With this algorithm every different combination of the chosen hyper-parameters is tested, while calculating the performance metrics using cross-validation, to find a possible better model. GridSearchCV method from Scikit-learn library, was implemented.

The implementation of this method required the compiled selected CNN model. Since this method returns a ranked report of the best results obtained by the different hyper-parameters combinations a scoring function was needed. A parameter grid in a Python dictionary format containing all the hyper-parameters to be tested was added. In GridSearchCV() method the parameters of the model used to apply these methods are optimized by cross-validated grid-search over the parameter grid.

In the SoA literature there is no definite answer on which hyper-parameters must be tested to achieve a good performance. This dissertation has chosen the **number of timesteps per sample, number of filters, number of neurons, pool size, data batch size, dropout rate, and learning rate** as the different hyper-parameters to be tested. The values experimented on the different parameters are listed on Table 4-6.

Table 4-6 Fine-tunning phase parameters explored.

Parameters	Values
Window sample timestep size	5, 56, 112, 240
Number of filters	32, 64, 128
Number of neurons	60, 100
Pooling size	2
Data batch size	256, 512
Dropout rate	0.3, 0.5
Model learning rate	0.001, 0.0001

There are different techniques that may be used to **cross-validate** a model. During literature research hold-out cross-validation [17][19][20][22] and k-Fold cross-validation [14][15][16][24] were the most common cross-validation used. According Payam et al [39] Hold-out validation avoids the overlap between training data and test data, yielding a more accurate estimate for the generalization performance of the algorithm. The downside is that this procedure does not use all the available data and the results are highly dependent on the choice for the training/test split. In k-fold cross-validation the data is first partitioned into k equally (or nearly equally) sized segments or folds. Subsequently k iterations of training and validation are performed such that within each iteration a different fold of the data is held-out for

validation while the remaining k - 1 folds are used for learning [39], thus using all data available. In this model tunning phase, was decided that a **tenfold cross-validation would be implemented**.

4.3.6 Ablation Model Index

To know the contribution that each layer to the general performance of the model, an ablation study was made. This study consists of systematically removing layers from the general model and checking its performance at each removal. The different ablations performed to the compete architecture can be seen in Table 4-7.

At each removal, the model was trained with all training dataset available, and with the same parameters utilized to build the general model. The results were obtained, by testing this new model versions against the test dataset, see Table 4-12.

Table 4-7 Ablation study for this dissertation proposed architecture.

Model	Architecture
General	**C1 + C2 + D + MP + FL + FC1 + FC2**
Ablation 1	C1 + D + MP + FL + FC1 + FC2
Ablation 2	C1 + D + MP + FL + FC2
Ablation 3	C1 + MP + FL + FC2
Ablation 4	C1 + FL + FC2
C- Convolution layer D – Dropout layer MP – Max pooling FL – Flatten layer FC – Fully connected layer	

4.4 Results

4.4.1 Fine-tuning Stage Performance

Table 4-8, presents the best training and validations results from grid search hyper-parameters optimization.

Table 4-8 Fine-Tuning stage optimal parameters found, and the results obtained with them.

Window timestep size	Parameters	Value	Dataset	F1 score (mean)	Standard deviation
240	Epochs	200	**Train**	0.93695	0.00423
	Overlapping	50 %			
	Number of filters	128			
	Number of neurons	60			
	Pooling size	2	**Validation**	0.91285	0.01084
	Data batch size	256			

	Dropout rate	0.5			
	Model learning rate	0.0001			

4.4.2 Model Training and Validation Performance

Through the optimization of the algorithm parameters by performing a grid search, it was found that the DL model obtained better performance using window samples containing 240 timesteps in batches of 256, 128 filters in each convolutional layer, a pooling size of 2, 60 neurons in each dense layer, dropout rate of 0.5 and the algorithm should update at 0.0001 rate. The best parameters for the proposed model are presented in Table 4-8. It was necessary to retrain the model with the new found parameters to understand if the model was performing well and to identify if it was underfitting or overfitting the data.

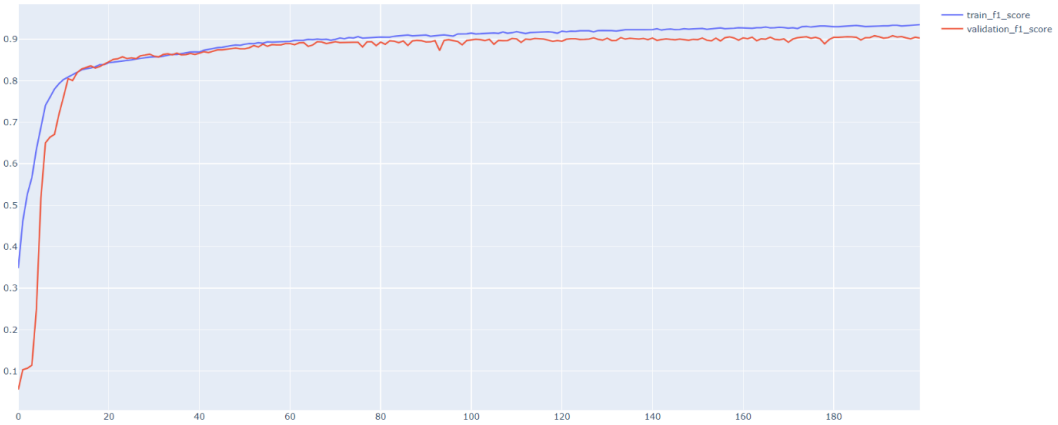


Figure 4-23 F1-score value evolution for train and validation dataset.

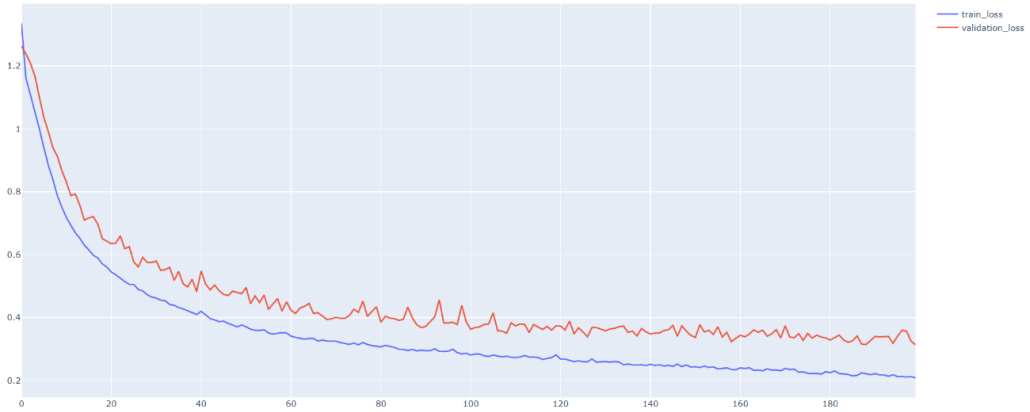


Figure 4-24 Loss value evolution for train and validation dataset.

After analyzing the evolution of the evaluation metric F1 score, and the Loss Function value, that evaluates how well specific algorithm models the given data, to retrain the final model with all data available it was concluded that to prevent a possible model overfitting the training data, the final algorithm should be retrained with 100 epochs and with all the optimized parameters found in grid search.

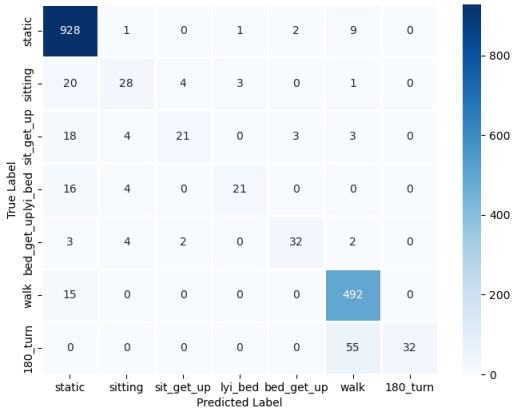


Figure 4-25 Confusion matrix results of the DL model trained for 100 epochs.

Table 4-9 Metrics evaluation for 100 epochs trained model.

Epochs	Dataset	Loss value	Accuracy	Precision	Recall	F1 score
100	**Train**	0.27854	0.91255	0.92799	0.89879	0.913127
	Validation	0.36319	0.90139	0.92848	0.88334	0.90518

4.4.3 Test Performance

Before evaluating this dissertation proposed solution, the algorithm was retrained with all training data available and with the optimal parameters discussed before, 4.4.1. To evaluate the performance of the DL model proposed, completely unknown data to the algorithm was used.

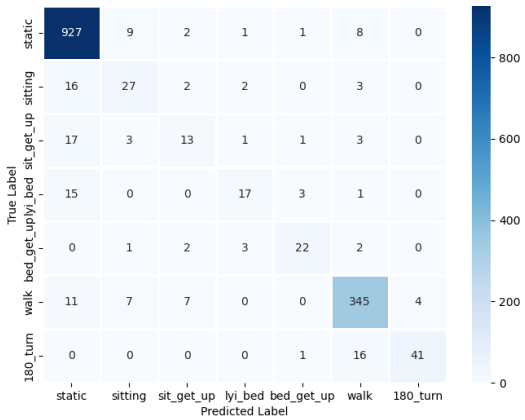


Figure 4-26 Confusion matrix test results, using Test dataset.

Table 4-10 DL model final evaluation using the Test dataset.

Dataset	Loss	Accuracy	Precision	recall	F1 score
Test	0.326218	0.907432	0.944849	0.87622	0.90892

Table 4-11 Accuracy values from HAR model final test, by each recognized activity.

	Activity						
Classification	**Standing still**	**Sitting on chair**	**Get up from chair**	**Lying on bed**	**Get up from bed**	**Walking**	**180º turn**
Well labeled	927	27	13	17	22	345	41
Wrong labeled	23	23	25	19	8	29	17
Total	950	50	38	36	30	374	58
Accuracy	97.78%	54.00%	34.42%	47.22%	73.33%	92.25%	70.49%

4.4.4 Ablation Index Performance

The following Table 4-12, present the different results obtained throughout layers removal from the complete DL HAR model architecture proposed in this dissertation. These obtained results were measured using the test dataset. The ablated models were trained with all training data available and with the same parameters as the complete DL HAR model suggested by this dissertation.

Table 4-12 Ablation study results using test data for this dissertation proposed DL model architecture.

	Models				
Metrics	**Complete**	**Ablation 1**	**Ablation 2**	**Ablation 3**	**Ablation 4**
Loss	0.326218	0.487938	0.491814	0.405462	0.361279
Accuracy	0.907432	0.898305	0.89309	0.887223	0.901565
Precision	0.944849	0.948564	0.939242	0.912337	0.924155
Recall	0.87622	0.848189	0.850122	0.866408	0.878824
F1 score	0.90892	0.895159	0.892253	0.888625	0.900728

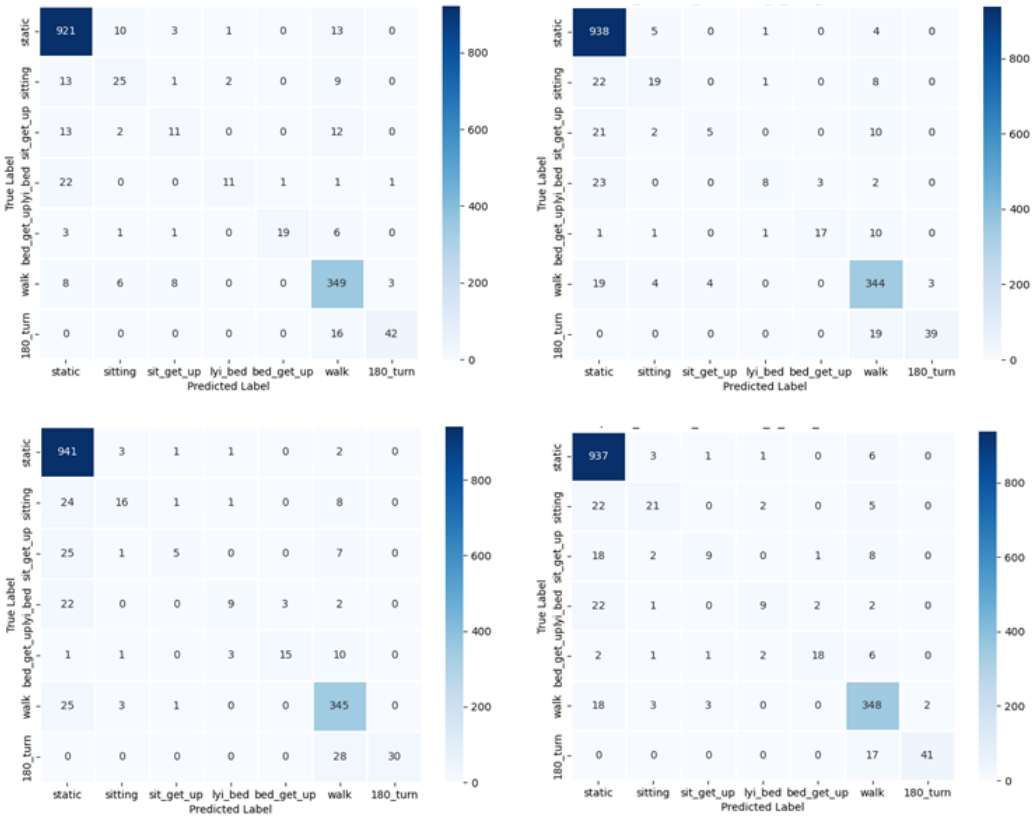


Figure 4-27 Confusion matrix result for architecture Ablation 1, Ablation 2 in upper row. Confusion matrix results for architectures Ablation 3 and 4 in lower row, respectively.

4.5 Discussion

4.5.1 Proposed DL HAR Model Performance

PD is a neurodegenerative disorder of the central nervous system which is characterized by the loss of non-motor and motor functions. PD impacts in a harmful way the patients' lives, and the life of people around them. Motor symptoms are the main trademark of PD, and physicians will benefit from clinical support tech-tools capable of supervise and gather data about patients' motor performance while performing daily physical activities. With this data physicians will be able to adopt a closed follow-up of the true motor stage of their patients, enhancing the power of prescribed treatments.

This dissertation aimed to develop a framework capable of recognizing basic daily living activities performed by PD patients through the application of DL algorithms.

Human activity recognition is a problem of classifying sequences of data. There are two main classes of an artificial neural network capable of performing this task: LSTM and CNN [40]. While both can perform the same task, they work in different ways. LSTM focus on learning and classify through temporal dependencies [41] while CNNs exploits spatial correlation in data to make the output [42]. In

literature there is no clear answer to which is the best of this neural network classes in performing human activity recognition. But is clear that they can outperform classic machine learning algorithms. In literature DL algorithms obtained around 25% higher performance than the classic ML algorithms.

A CNN model was implemented and during training stage obtained F1 score metric of 0.90518 and an Accuracy of 90.14%. This result was achieved by training the model during 100 epochs, using a window sample containing 240 timesteps in batches of 256, 128 filters in each convolutional layer, a pooling size of 2, 60 neurons in each dense layer, dropout rate of 0.5 and the algorithm should update at 0.0001 rate. In SoA , the best performing **LSTM architecture**, proposed by Murad et al [18], achieved an **97.8 % accuracy** in recognizing activities on WISDM dataset, while the better **CNN architecture** resulted in a **98,6% accuracy** and was proposed by Ravi et al [14] to assess his mode against USC HAR dataset. This dissertation used a **dataset** comprised of kinematic data collected from **PD patients while** the **SoA** found best models using **data** collected from **healthy persons** with ages between 21 – 49 years, which have significantly less data variability compared with the inherently impaired motion of PD patients with ages between 59 – 76 years.

The kinematic data collected to build this dissertation training, validation and test datasets is unbalanced due to the nature of activities performance duration, as seen in Table 4-3. Some activities take longer time to be performed (e.g., 10m *walk*) in comparison with others (e.g., *sitting on chair*). As such the main metric to **evaluate our model** was **F1 score**, according to Hossim et al [37] it is a more appropriate metric to be used because as discussed earlier gives a better measurement of the incorrectly classified cases than the, not so indicated metric for unbalanced dataset problems, Accuracy metric utilized by the SoA best performing models.

This **dissertation contributes to the SoA** with an algorithm based on **only one sensor instead of using several sensors** as in Murad et al [18], and **by standardizing the sensor placement at lower back instead of the assumption** made by Ravi et al [14] that the subject will always place the smartphone at front right hip, enabling to obtain more ergonomic, compact, and daily easy-to-use solution.

Final test results showed that the proposed CNN model in is able to recognize HAR with a F1 score of 0.90892 and an Accuracy of 90.7432%. The *standing still* activity, is the best recognized activity by the model with an accuracy of 97.78%, followed by the *walking* activity with 92.25% accuracy. The continuous and cyclic nature of the data from these two activities in conjugation with its length allowed the model to learn particularly well these actions. In contrast, the *sitting*, *getting up from chair*, with 54.00% and 34.42% accuracy each are where the model performed worst, as Table 4-10

indicates. These activities usually are short in timestep length, with a mean of 388 timestep length in performing *sitting* action and a mean of 303 timestep length in *getting up from chair*, as seen in Table 4-3. With 47.22%, 70.00% and 73.33% accuracy, *laying on bed*, 180° *turn and get up from bed* have overall better recognition than *sitting* and *get up from chair*, with 54.00% and 34.42% accuracy, respectively. *Laying on bed*, 180° *turn* and *get up from bed* have a longer duration in execution with a mean 512 timesteps to both *lie on bed* and *180° turn* and mean duration of 481 timesteps length to *get up from bed*. The longer the activity duration is, better results are obtained with this model, which might indicate that in Training stage more samples of small duration activities are needed to the model better learn the characteristics associated with them. By analysing the Test stage resultant confusion matrix, Figure 4-, it is also observable that the model has difficulty in distinguishing between *walking* activity and *turning 180°*. This difficulty is explained by observing the representing graphic of the collected tri-axial data, where they are extremely similar at naked eye, see Figure 4-5.

The **developed solution innovates by being able of classifying daily living activities performed by PD patients and adds to the SoA the recognition of two new activities: *get up from chair* and *180º turning*.**

4.5.2 Ablation Index Performance

At each layer remotion, the proposed **complete CNN solution maintained the best result**, with F1 score equals to 0.90892. The reduction in F1 score, accuracy, precision and recall were visible throughout the reduction layer process. The most trimmed model obtained a still good F1 score of 0.900728 for its performance and could also be a possibility if a lighter model should be needed.

Although it was not directly measured in this study, the training process of the ablated models decreased along the layers trimming. This is expected because less layer, less calculus the computer needs to be performed, thus less time computing.

4.6 Conclusion

Having obtained a **F1 Score of 0.90892,** this dissertation concludes that a satisfactory result was achieved, although this dissertation supports the idea that there is still room to improve since is a result lower than other SoA solutions.

This score was obtained by using data from PD patients, that as observed in dataset exploration have high variability. The data was subjectively and manually labelled. The dataset used in training stage

was highly unbalanced and have tri-axial data of acceleration and angular velocity collected from a single IMU sensor placed in lower back.

Having in mind the aspects mentioned before, as future work this dissertation suggests that to improve the learning capacity of the model more data should be collected, the data labelling should be performed and reviewed by 3 or more people with clinical experience. Techniques of data balancing like oversampling, under sampling and addition of gaussian noise should be tested out. Besides the raw tri-axial data, experimenting with features extracted from the dataset should also be considered.

Naturally, not all possible DL models were tested before reaching this dissertation proposed model. As such, experimenting with different models' architectures is also a viable option in search of a better result.

During model parameters optimization phase, different combination of values and parameters could also be tested as also the cross-validation method could also be changed for.

In validation training phase instead of dividing the dataset in twelve patients to train and three to validation, and although the computation time required is higher a Leave-one-out cross-validation should be considered.

To finalize, instead of testing the performance of the model with offline data, implementing the model proposed in +Sense project and subject it in a real-time evaluation during a real-world situation, would be even more interesting and would bring even more credibility to its real performance.

5 Deep-Learning Framework for Human Walk IC/FC Detection

5.1 Introduction

Walking impairments are among the most common and disabling symptoms of Parkinson's disease [43]. As PD it's a slow progressing disease, walking problems worsen as the disease progresses severely impacting the patient's autonomy [44]. Making the capability of continuously monitoring and assess patients walking patterns, like step time and stride length, an important diagnosis tool not only for PD specialized clinics as also for other scientific areas that are walking focused and are developing new assistive devices to bring more comfort, autonomy, and life quality to PD patients. Nowadays such assessments are performed by human clinics by following the standard scale UPDRS-III and the results might vary depending on the experience and subjective point of view of the professional. As a **complementary goal**, this dissertation aimed to **endow +Sense project with a walking IC/FC detection framework** using an **AI DL algorithm**.

Before starting to develop a solution, it was needed to know better this problematic. A brief SoA review was made. In this review this dissertation focused on knowing and understanding the different walking segments, data required, and it's needed transformations, which models are used and on how to evaluate them.

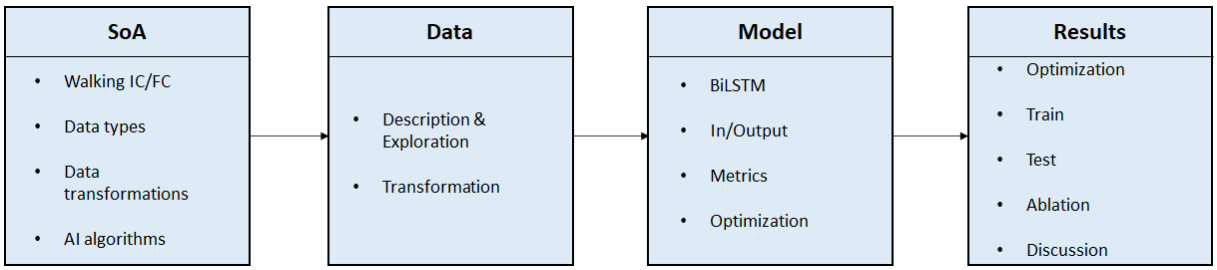


Figure 5-1 Human walk IC/FC detection framework schematic view.

5.2 Related Works: brief State of the Art

According to Michael W.Whittle [45] the human walking is defined as a locomotion method generated by the use of both legs in an alternated way, allowing the support and propulsion of the human body. Walking execution requires the presence of periodic movements of each foot toward a support position and reaction forces applied to the feet, enough to give support to the body even under distortions due to pathological conditions [46]. This periodic movement is called "walking cycle" and can be divided into eight different phases [47]:

1. Initial Contact: considered the starting and finishing event of a walking cycle. Corresponds to the first contact between the heel and the ground.
2. Foot flat: The plantar surface touches the ground.
3. Midstance: The contralateral foot passes the position of the stance foot.
4. Heel-off: The heel comes off the ground. Higher pressures are present in the metatarsal heads.
5. Final Contact: The toes lose their contact with the ground. The least amount of foot is in contact with the surface during this event.
6. Acceleration: quick acceleration of the swinging foot.
7. Mid swing: The swinging foot passes the position of the stance foot.
8. Deceleration: The swinging foot decelerates as it approaches the ground. The heel takes its position for a new Initial Contact.

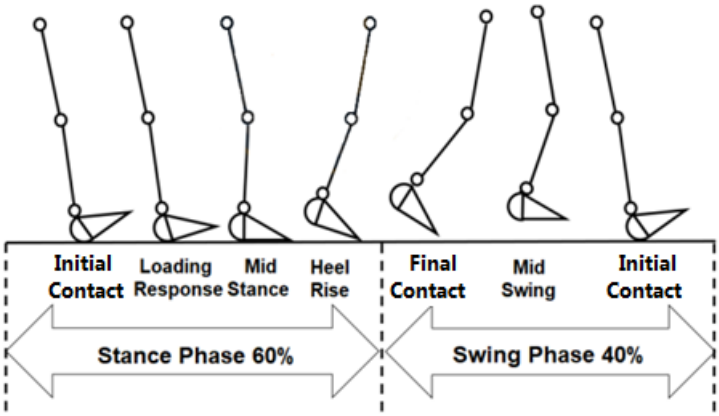


Figure 5-2 Walking cycle and walking events.

Since walking have a cyclical pattern, parameters that break the pattern norm can be good indicator that the individual is suffering walking impairment, thus the necessity and importance of a close monitoring of the walk, especially in non-healthy subjects.

Table 5-1 provides an overview of the different walk segmentations recognized by the selected studies during this brief SoA survey, as well as a synthesis of the studies objectives.

Table 5-1 Types of walking segmentation.

Paper	Goal		
	Target	**Participants number**	**Walking Segments**
[48]	Children with walking disorders	226	The 5 walking events (no event, left Foot Strike, right Foot Strike, left Foot Out and right Foot Out)
[49]	Healthy adults	20	Initial Contact and Final Contact
[50]	Health adults	4	Stance and Swing phase

Three different ways of segmenting the walking cycle were found. [49] segmented the walk by **two distinct walk events: Initial Contact and Final Contact**. While [48] divided the walk cycle by all of its walking events and [50] partitioned by walking phases: stance and wing phase.

Table 5-2 resumes the type of sensors used, data collected, and features needed by the different research to recognize the different walking segments.

Table 5-2 Sensors and features presented in SoA research.

Paper	Sensor				
	Sensor number	**Sensor Type**	**Sensor Location**	**Features collected**	**Features extracted**
[48]	8	Reflective markers	right and left anterior superior iliac spines, greater trochanter, midthigh, knee joint, mid shank, lateral malleolus, between the second and third metatarsal heads	X, Y, Z coordinates	-
	4	Force plates	floor	ground reactions forces	-
[49]	2	IMU	Right and left ankle	triaxial accelerometer data	four composite accelerations by calculating the root sum of squares
[50]	3	IMU	Foot, calf and thigh	triaxial accelerometer data	Standard deviation (SD), mean absolute value (MAV), maximum (MAX), minimum (MIN), and median (MED)

[49], [50] applied **IMU sensors** for data collection while [48] opted for Reflective Markers. **Tri-axial accelerometer data** was collected in [49], [50] and the relative positional coordinates were collected on [48]. Extra features were extracted and added to the different datasets, except on research paper [48].

In all research studies more than one sensor was used to collect data.

Table 5-3 is focused on the most common AI algorithm implemented to perform walking segmentation recognition

Table 5-3 AI algorithms presented in SoA research.

Paper	**Algorithm**			
	Model	**Data Split**	**Validation methods**	**Performance**
[48]	LSTM	60% trials for training, 10% trials for validation, 30% trials for testing	-	AUC of 0.9955 for the left FS, 0.9971 for the right FS, 0.9958 for the left FO and 0.9955 for the right FO
[49]	LSTM	First 70% of the total timesteps as training and the last 30% of timesteps as test data.	3-Fold cross-validation	F1 score of 0.98 for right/left Initial Contacts and 0.94 for right/left toe-offs
[50]	LSTM	3 volunteers for training and 1 for test	-	Stance and Swing phase; F1 score +/- 0.94

Research papers like [48], [49], [50] implementing AI **LSTM based algorithms were the most commonly found**. The **metric** usually defined to evaluate the performance of the different models are the **F1_score** and **Accuracy** with the combined application of k-Fold and Leave-one-out cross validation techniques.

5.2.1 Conclusion

From the brief **SoA review**, it was **concluded that IC and FC are crucial parameter in the quantification of PD patients walking disability**. Current approaches for evaluating PD patients walking disabilities are limited to presential observed-based methods. **IMUs sensors are the most chosen** instruments by authors, to collect and provide kinematic data input on their DL models.

There is **no standard methods** for kinematic data collection, minimum of sensors required and no consensus regarding their body placement. In SoA, it was found that two or more sensors are used for data collection and are often placed on lower body extremities of person body. These facts are **huge constraints** in a monitoring system that must be practical and easy to be worn in day-to-day and while performing the daily activities. It's important to have in mind the difficulty that PD patients have in executing motor plan movements.

An **absence of an open public dataset was constated**, since all SoA papers had to build and use their own dataset. These datasets kinematic **data were all collected on laboratory environment**.

In respect of DL models, **LSTM models are the most popular models implemented**, as also the usage of **F1 score metric to evaluate the performance** of the walking events recognition models.

In SoA research overview it was also noticed that **no solution proposed was tested in a real-life environment**. Table 5-4 summarizes the identified limitations regarding technological, adopted strategies, and validation methodology issues and it is also provided guidelines for their mitigation.

Table 5-4 Limitations and guidelines to overcome them.

Limitation	End user requirements	Guidelines
No clear body configuration of WS	Portability, comfort, easy set-up	Find a trade-off between the number/location of sensors without losing significant data
Poor data collection on home-based conditions	Personalized treatments	Perform experimental tests on home-based scenarios

No assessment of HAR usability	Acceptability of the device	Include the users' opinion in the development of the proposed solution and assess its acceptability and usability

5.3 Data Input

5.3.1 Data acquisition

Walking dataset with kinematic-driven data collected from **PD patients** was used to train, validate and test this dissertation proposed DL model. This data was collected and used as part of previous work for the +Sense project [28][51]. This custom dataset data acquisition was conducted under the ethical procedures of the Ethics Committee in Life and Health Sciences (CEICVS 147/2021), following the Helsinki Declaration and the Oviedo Convention. All participants gave their informed consent to be part of the study. Data was collected at the Hospital of Braga – 2CA Braga Academic Clinical Center.

5.3.2 Dataset description

The PDGD dataset is a collection of data from 40 different PD patients.

Two different trials were performed by PD patients, a ten meter strait line walk, Figure 5-3, and a more complex circuit descripted bellow on Figure 5-4.

Figure 5-3 Patients started the trial on (1), walked 10m and stopped at (2).

Figure 5-4 Patients started the trial on (1), walked 30m, turned right at (2), walk forward 2m (3), turned around (4), walking forward 2m (5), runed left at (6), made a 32m walk (7), tuner at left (8), walking 2m (9) until returning to the trial initial position.

Each trial datafile contains eight independent variables:

- time_sampling: internal time of the processing unit responsible for acquiring the IMU data;
- Sample: id number of the sample ;
- acc_x: acceleration value in x axis ;
- acc_y: acceleration value in y axis ;
- acc_z: acceleration value in z axis ;
- gyr_x: angular velocity value in x axis ;
- gyr_y: angular velocity value in y axis ;
- gyr_z: angular velocity value in z axis .

For each patient trial, the dataset has four distinct files. Each one of this four files have the timestamps of when the different walking events occurred, as seen in Figure 5-5. The four walking events registered are:

- IC left – Initial Contact left foot
- FC right – Final Contact right foot
- IC right – Initial Contact right foot
- FC left – Final Contact left foot

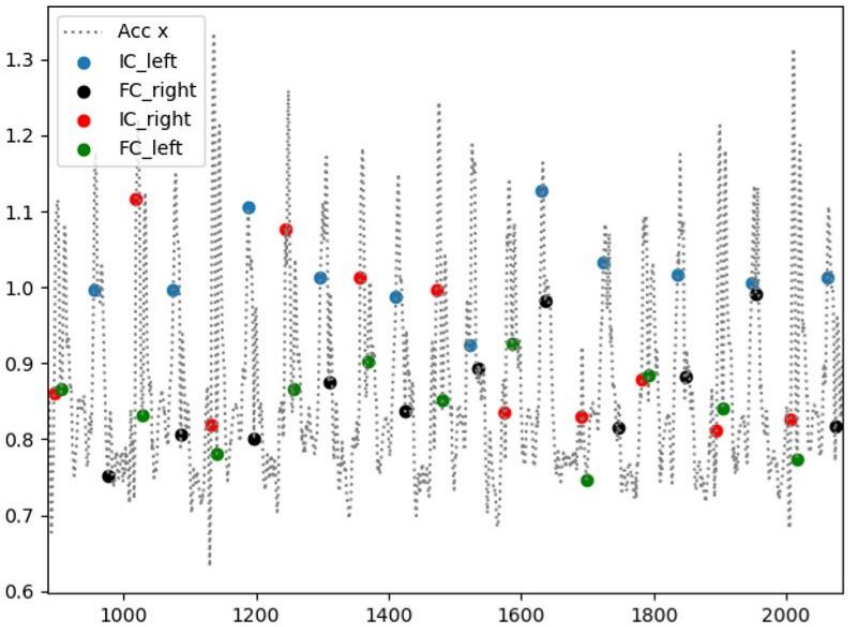


Figure 5-5 Walking events occurrence during the trial plotted over the x-axis values of the acceleration based on [28].

5.3.3 Data Input preparation

Decisions about the data labelling, what transformations should be made to the data and how the data should be split will be discussed and presented in this section. A schematic resume of this section is presented in Figure 5-6.

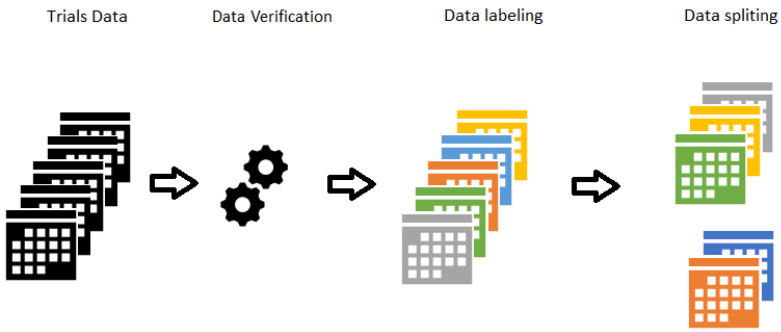


Figure 5-6 Human Walking IC/FC detection data input handling pipeline.

5.3.3.1 Data Verification

Before starting using the data in the AI DL model development, a manual revising and correction of the walking event timestamps was needed and performed.

A manual overhaul of all the walking events has been carried out. In this laborious processed, due to bad data collection three trials data files were eliminated and twenty-six wrong walking events timestamps were corrected based on previous literature [28]. Figure 5-7 exemplifies the wrong walking events timestamps and how they were corrected.

Figure 5-7 Correction to walking events timestamps. Green color circle representing a FC event timestamp was placed at green color dot timestamp. Orange color cross representing a IC event timestamp was placed at red color cross timestamp.

5.3.3.2 *Features extraction*

In SoA research was prevalent the utilization of handcrafted features [49], [50] and features like SD, MAV, MED or root sum square were the most frequently used. Although Lempereur et al [48] obtained good results without applying features extraction, the researcher have added to training dataset data from force plate sensors located at the floor. **This dissertation**, to add more value to the SoA, **chose not to add handcrafted features to the Training dataset**.

5.3.3.3 *Data Labelling*

Two prevalent types of labelling the data were found on literature research. Labelling the data and classifying it by **walking stance phase**, as in research [50]. Labelling data timesteps with **Initial Contact right foot, Initial Contact left foot, Final Contact right foot and Final Contact left** foot, as in [48] study. The **third most common** way found was as in research paper [49], where the timesteps where the walking events happened during trial were classified only as **Initial foot Contact and Final foot Contact**. In accordance with SoA research paper Zhen et al [50], **timesteps were classified as Initial foot Contact (IC) and Final foot Contact (FC).** With this chosen labelling style, to the walking events chosen to be recognized **'0'** will **represent IC** event and **'1' the FC event**, as presented on Table 5-5.

Table 5-5 Walking events to be recognized label ids'.

Label Id	Walking event
0	Initial foot Contact
1	Final foot Contact

An example on how the labelling for the kinematic data used in Training stage is visible in Figure 5-8, where the red colored timesteps are labelled as '0' which is the code identification of IC event and the green colored timesteps are labelled as '1' corresponding to FC event. When an IC event happens,

the following timesteps are labelled as '0' until a FC event appears. Whenever a FC event is performed the following timesteps are labeled as '1' until an IC event happens, and so on.

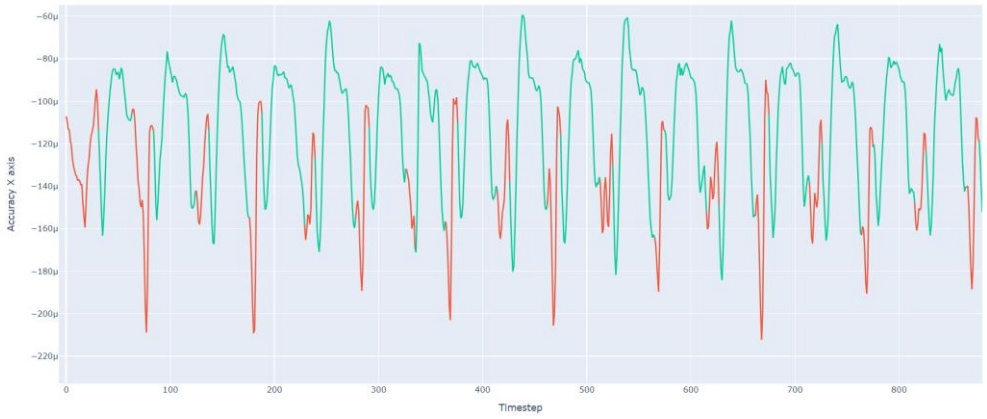


Figure 5-8 PD patient Trial labelled example. With red color timesteps labelled as IC. With green color timesteps labelled as FC.

5.3.3.4 Data Splitting

The whole dataset was divided into four different individual datasets. Specifically, 20% was used as test dataset (8 patients) for the proposed model final evaluation, 80% (32 patients) of the data was used in the fine-tuning stage of the proposed model. This 80% of data is further divided into 80% (26 patients) training and 20% (6 patients) validation datasets with the purpose of training and validating the final model using the best hyper-parameters found in fine-tunning stage phase. The testing dataset is only used once as unseen data for a final assessment of the performance the model.

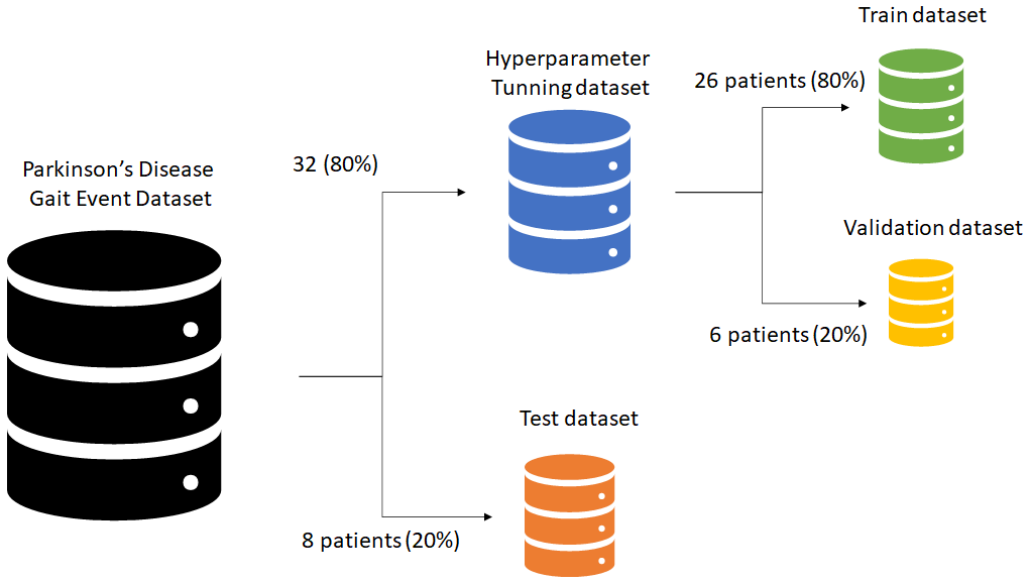


Figure 5-9 Schematic representation of the dataset splitting for the different Human walk IC/FC detection model development stages.

Although not mentioned on [48], [49], [50], a **MinMax normalization** was performed **feature-wise in an independent** way over the accelerometer and gyroscopic collected data**.** But as mentioned before**, variables measured** at **different scales do not contribute equally to the model fitting** and model learned function and might end up **creating a bias**, thus normalization is a needed procedure.

5.4 Model Pipeline

A **Bidirectional Long-Short Term Memory** (BiLSTM) AI DL algorithm **is proposed by this dissertation** for the recognition of walking events performed by PD patients while walking.

Several steps were needed to obtain the proposed model of this dissertation. These steps encompass decisions about data input and its format, model composition, how to evaluate its training performance and how to fine tune it to obtain its final version. In the following section a more detailed exposition will be made.

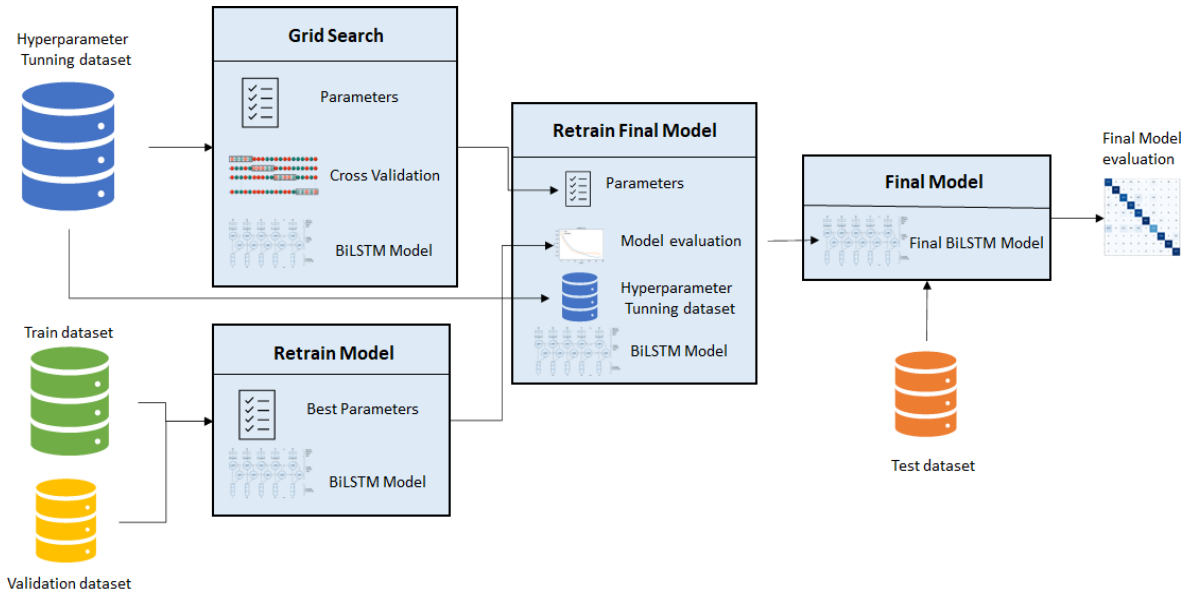


Figure 5-10 Model Pipeline Schema representing the development of the final Human walking IC/FC DL model for PD patients' solution.

5.4.1 Bidirectional Long Short-Term Memory

LSTM is a variety of RNN capable of learning long-term dependencies. LSTM due to its feedback connections is capable of processing and entire sequence of data.[52] By adding thresholds, called input

gate, forget gate, output gate and memory unit to a RNN, [52] solved the vanishing and exploding gradient problem present in the RNN. Figure 5-11 present a schematic view about the LSTM cell composition.

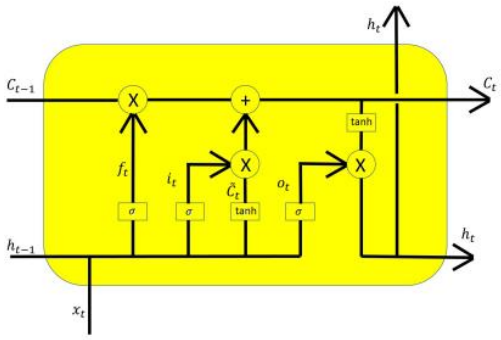


Figure 5-11 Architecture of a LSTM cell.

According to [53], LSTM by having the nodes between the hidden layers connected, allowed the input data of the hidden layers at the given time not only containing the output of prior hidden layer but also includes the output of its own hidden layer at a previous time. The historical information of the data sequence is stored in the hidden layer.

Bidirectional LSTM are an upgrade over LSTM. In BiLSTM, each training is presented forward and backwards while in LSTM the input sequence can only flow either forward or backward. Both sequences are connected to the same output layer, granting complete information about every point in a given sequence, everything before and after it [54]. A high-level schematic vision representation on how a LSTM and BiLSTM work is presented on Figure 5-12.

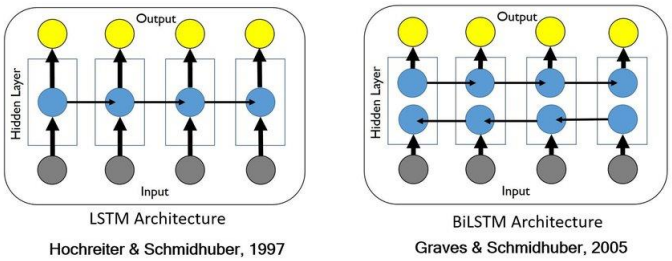


Figure 5-12 LSTM and BiLSTM architectures workflow.

BiLSTM architecture have multiple parameters, but five parameters can be considered key to its performance: batch size, timestep size, neurons number, dropout rate and learning rate. Batch size defines the number of samples to work on before the internal parameters of the model are updated. Timestep size is the number of observations that each sample window has. Neurons numbers defines

the quantity of nodes present in each layer. Dropout rate have values between 0 and 1, this rate helps avoid overfitting in training by bypassing randomly selected neurons, thereby reducing the sensitivity to specific weights of the individual neurons. Learning rate defines how quickly the network updates its parameters.

Setting a higher learning rate accelerates the learning but the model may not converge or even diverge. Conversely, a lower rate will slow down the learning drastically as steps towards the minimum of loss function will be smaller but will allow the model to converge smoothly.

5.4.2 Proposed Bidirectional Long Short-Term Memory Model

The proposed algorithm is based on Lempereur et al [48] algorithm and is able to classify walk IC/FC events from PD patients walking kinematic data produced by one IMU.

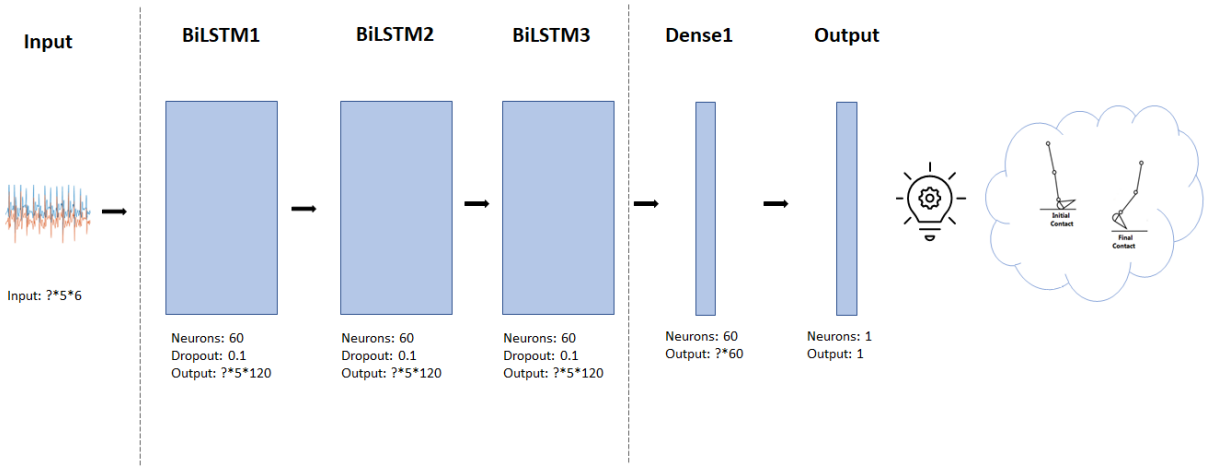


Figure 5-13 Proposed Human Walk IC/FC DL model architecture.

The model is composed of three Bidirectional LSTM layers, with 60 neurons each. Each of these layers implemented dropout at 0.1 rate. After the last Bidirectional LSTM layer, the outputs are flattened to one long vector and pass through a dropout layer, with 0.1 drop rate, before passing the outputs to a fully connected layer, with 60 neuros, using Rectified Linear Unit (Relu) activation function before the last fully connected layer, with 60 neurons, where a prediction is made through the application of Sigmoid activation function.

To optimize the network Adam version of stochastic gradient descent algorithm is used, and the binary cross entropy loss function is applied since we are learning a binary-class classification problem.

5.4.3 Input and Output Data Format

Similarly, to section 4.3.3 the model requires a three-dimensional input with [n° of samples, timesteps, n° of features]. Meaning, each input sample is one window of time that will have a pre-determined number of timestep, and a timesteps has a fixed number of variable or features.

The output for the model will be a two-element vector containing the probability of a given window belonging to one of the two possible walking events.

These input and output dimensions are required when fitting the model, and we can extract them from the provided training dataset.

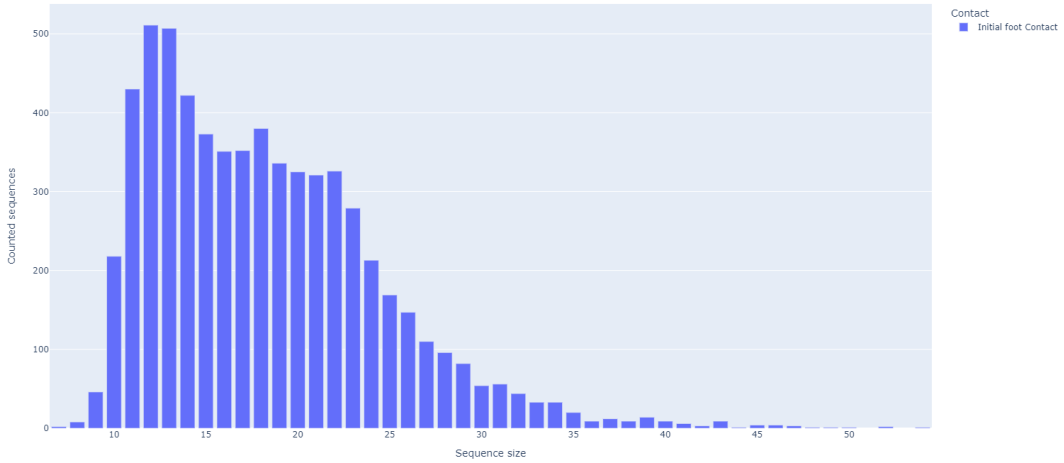


Figure 5-14 Counting the Initial foot Contact sequence size.

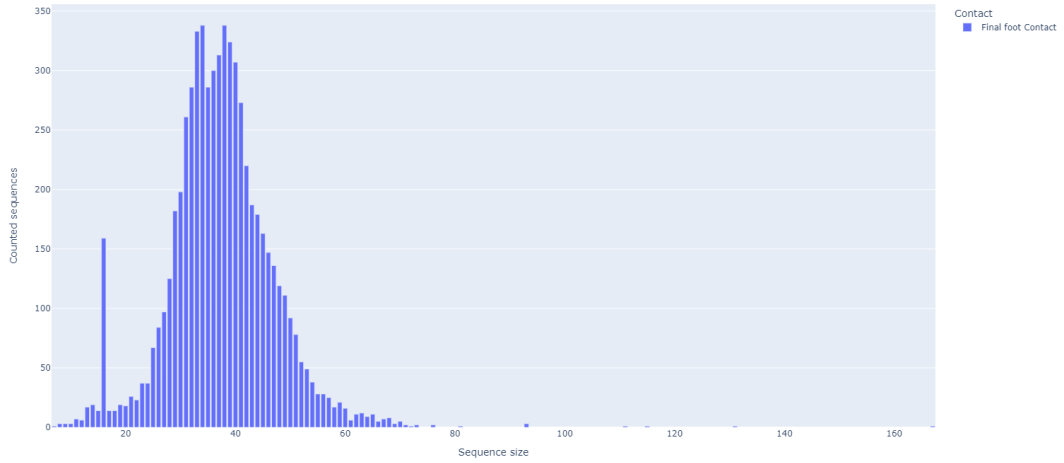


Figure 5-15 Counting the Final foot Contact sequence size.

Since the proposed algorithm training input works by feeding data samples with fixed timesteps size, a better understanding over the sizes of IC sequences, Figure 5-14, and FC sequences, Figure 5-15,

was studied. The most counted IC event sequence was the sequence of 12 timesteps with 511 occurrences, while in the FC walking event is the most represented sequence was divided between 34 and 38 sequences size with the corresponding counts of 338 for both.

In Table 5-6, it is possible to conclude that a minimum sequence of 7 timesteps for Initial Contact and Final Contact was performed by a subject in the trials, a mean sequence size of 18 and 37 ,respectively, for Initial Contact and Final Contact was collected and also a 54 sequence size for Initial Contact and 167 sequence size Final Contact was founded in all dataset.

This dissertation opted for a window input size of 5 timestep, and a **data overlapping of 80%** with the intention that the model would handle transitions more accurately.

Table 5-6 Initial Contact and Final Contact sequence size.

	Initial Contact	Final Contact
Minimum	7	7
Mean	18	37
Maximum	54	167

When fitting the model, it is also required a two-dimensional input [n° of samples, label]. Meaning, each sample have the corresponding label. The **label is in one-hot-encoding format**. The label of the correspondent sample input is defined by which label occurs the most in the selected sample window, this process works in the same way as stated in section 4.3.3.

5.4.4 Model Evaluation Metric

Although not mentioned in the literature review, Mathew Correlation Coefficient (MCC) is the metrics chosen to evaluate the quality of this dissertation proposed model. F1 score is the prevalent metric found in literature, Table 5-3, but according to [55] MCC produces a more informative and truthful score in evaluating binary classifications than accuracy and F1 score. MCC is a more reliable statistical rate which produces a high score only if the prediction obtained good results in all the four confusion matrix categories (true positives, false negatives, true negatives, and false positives), proportionally both to the size of positive elements and the size of negative elements in the dataset. MCC can be calculated as follow:

$$MCC = \frac{TN * TP - FN * FP}{\sqrt{((TP + FP) * (TP + FN) * (TN + FP) * (TN + FN))}}$$

Figure 5-16 Matthews Correlation Coefficient formula. TN: True Negative; TP: True Positive; FN: False Negative; FP: False Positive.

According to [56][57] MCC is the only binary classification rate that generates a high score only if the binary predictor was able to correctly predict the majority of positive data instances and the majority of negative data instances. It ranges in the interval [−1, +1], with extreme values – 1 and +1 reached in case of perfect misclassification and perfect classification, respectively, while MCC = 0 is the expected value for the coin tossing classifier

5.4.5 Fine-tunning Stage

A similar process to 4.3.5 was implemented to optimize the model performance. A Grid Search parameter grid while applying a 10-Fold cross validation technique to validate and evaluate the performance was executed. Batch size, number of neurons to be used, dropout rate and the learning rate were the tested hyper-parameters. In Table 5-7 are presented the values tested for each hyper-parameter.

Table 5-7 Optimization hyper-parameters.

Hyper-parameter	Tested values
Batch size	256, 512
Neurons number	30, 45, 60
Dropout rate	0.1, 0.15, 0.2
Learning rate	0.001, 0.0001

5.5 Results

5.5.1 Fine-tuning Stage Performance

The best combination of parameters found by grid search process are presented in Table 5-8.

Table 5-8 Fine-Tuning stage optimal parameters found, and the results obtained with them.

Window timestep size	Parameters	Value	Dataset	(mean) MCC score	Standard deviation (std dev)
5	Epochs	200	**Train**	0.95228	0.00179
	Overlapping	80 %			
	Batch size	512			
	Number of neurons	60	**Validation**	0.88104	0.00385
	Dropout rate	0.1			

	Model learning rate	0.001			

5.5.2 Model Training and Validation Performance

By performing a grid search process several parameters were tested to find the more indicated to be used in the final model. It was revealed that the BiLSTM architecture proposed by this dissertation performed better by using batches containing 512 window samples with 5 registered timesteps each. BiLSTM layers should have 60 neuros, dropout rate of 0.1 and the algorithm should update at 0.001 rate. With this a satisfactory MCC value of 0.88104 (std dev = 0.00385) was achieved, see Table 5-8.

Similarly, to what was done in the Chapter 4, the model was retrained with the new set of parameters to evaluate how the model was performing and to avoid underfitting or overfitting situations.

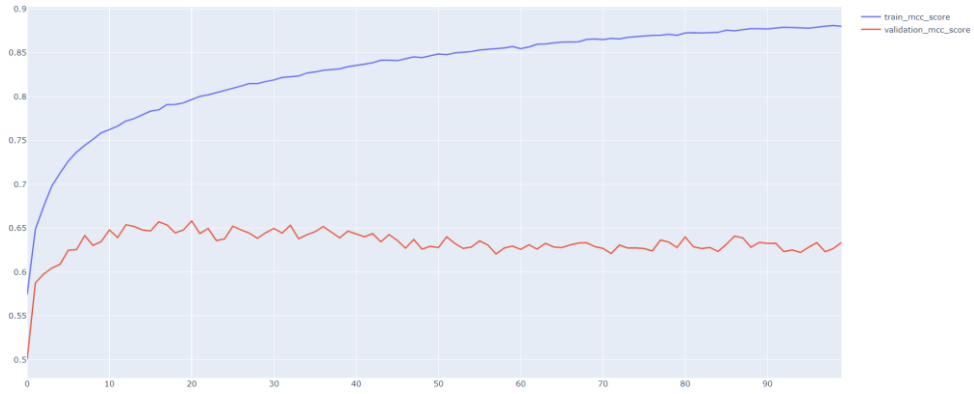


Figure 5-17 MCC Score value evolution for train and validation dataset.

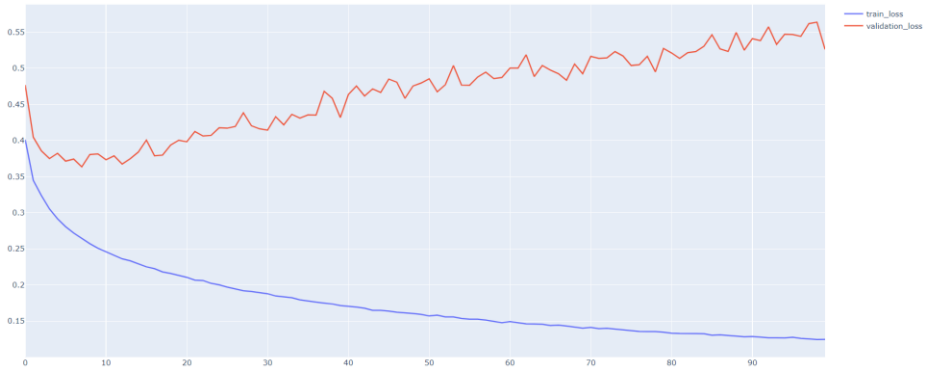


Figure 5-18 Loss value evolution for train and validation dataset.

By analyzing the evolution of the algorithm learning curve, Figure 5-18, and MCC Score value, Figure 5-17, it is possible to avoid overfitting and underfitting situations. In this case, it is clearly possible to see in Figure 5-18, that an overfitting is starting to occur after the tenth epochs of training. As such it

is advised to stop the training process there. The final model was retrained with 10 epochs and with the best parameters found in the optimization model stage, section 5.5.1 .

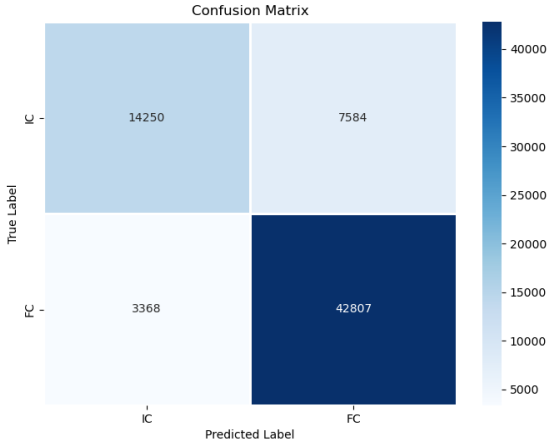


Figure 5-19 Confusion matrix results of the DL model trained for 10 epochs.

Table 5-9 Metrics evaluation for 10 epochs trained model.

Epochs	Dataset	Loss	Accuracy	Precision	Recall	F1 score	AUC	MCC
10	**Train**	0.250750	0.89726	0.906184	0.94777	0.926378	0.951711	0.75861
	Validation	0.381386	0.838962	0.852137	0.928297	0.885351	0.899318	0.63466

5.5.3 Test Performance

To do the final evaluation of this dissertation model proposition, the model was retrained with all data available for training, best parameters found previously in section 5.5.1, during the epochs necessary to avoid overfitting discussed formerly in 5.5.2 . After this process, the model performance was tested with completely new and unknown data. This data was isolated from the training dataset as discussed in section 5.3.3.4 .

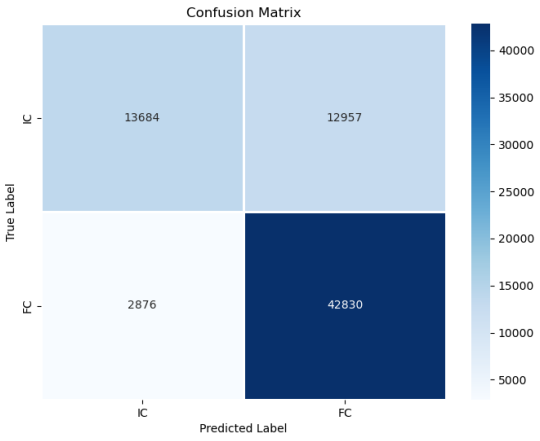


Figure 5-20 Confusion matrix test results, using Test dataset.

Table 5-10 DL model final evaluation using the Test dataset.

Dataset	Loss	Accuracy	Precision	Recall	F1 score	AUC	MCC
Test	0.540085	0.781152	0.774337	0.93841	0.841554	0.854381	0.538386

Table 5-11 Accuracy values for Human Walk IC/FC Detection model final test, for each walking event recognized.

	Walking event	
Classification	**Initial Contact**	**Final Contact**
Well classified	13684	42830
Wrongly classified	12957	2876
Total	26641	45706
Accuracy	51.36%	93.70%

5.6 DISCUSSION

Walking impairments are among the most common and disabling symptoms of PD. Walking difficulties worsen as the disease progresses severely impacting the patient's autonomy. Through the detection of IC and FC key walking patterns, like step time and stride length, can be determined and be of use to perform disease diagnosis by PD clinical experts or the development of walking assistance devices. The existence of a tool capable of continuously monitoring and assess patients walking patterns is vital.

This dissertation complementary goal aimed to develop a framework capable of detecting IC an FC events during PD patients walking through the application of DL algorithms.

Walking event detection, just like in Human activity recognition problematic, is a problem of classifying sequence of data. Human Walking IC/FC Detection assessment is a crucial aspect in identifying and interpret a user's walking abnormalities and in designing assistive tools [49].

During the brief literature research donned by this dissertation, DL models presented superior results compared to classical machine learning algorithms. Although CNNs are also capable of performing data sequence recognition, architectures based on LSTMs were found to be more common to perform this task in the literature. In fact, this dissertation model proposition is BiDirectional Long Short-Term Memory architecture base on Lempereur et al [48] but with a slight alteration to the input layer. While Lempereur et al [48] injects the input directly to the BiDirectional layers, this dissertation proposed that the first layer should be a Batch Normalization Layer to reduce the number of training epochs required to train deep neural networks. This alteration was needed due to the lack of computational resources at

time, and possibly justify why the validation Loss value at the tenth epoch of training starts to enter in overfitting, see Figure 5-17 and Figure 5-18.

This dissertation DL model proposition during training stage obtained a MCC score of 0.63466, F1 score of 0.885351, AUC of 0.899318 in the recognition of walking IC and FC events. This result was achieved by using a 80% overlapped window sample containing 5 timesteps in batches of 512, 60 neurons in each BiDirectional layer and Dense layer, dropout rate of 0.1 and the algorithm should update at 0.001 rate. This dissertation base line LSTM architecture proposed by **Lempereue et al** [48] **obtained AUC values of 0.9955 for the left IC, 0.9971 for the right IC, 0.9958 for the left FC and 0.9955 for the right FC**, and **Xing Tan et al** [49] achieved with its' LSTM architecture a **mean F1 score** of **0.98 for the left and right IC, 0.94 for the right and left FC.**

To build the necessary dataset to train, validate and test this dissertation proposed framework for human walk IC and FC detection, **one IMU sensor positioned at lower back was used to collect kinematic data from PD patients**.

This dissertation dataset used to train, validate, and test the proposed framework for human walk IC and FC detection, utilizes kinematic data from PD patients while compared with the best SoA performing models, **Lempereur et al** [48] and **Xing Tan et al** [49] **utilizes data from** children's' with gait disorders and **healthy people with age ranged** between 26.6 – 40.4 years**, respectively**. Another key factor for the obtained results is the number of wearable sensors utilized in the process of data collection. This dissertation utilized kinematic data collected through a single IMU sensor while Lempereur et al [48] utilized data from 12 different sensors, 8 reflexive markers placed on patient's body and 4 force plates on the floor, and in Xing Tan et al [49] 2 IMU sensors placed on each ankle was used.

The kinematic data utilized in the different building stages of the DL model proposed by this dissertation was also different from Lempereur et al [48] and Xing Tan et al [49]. This **dissertation only utilizes raw kinematic data** comprised of tri-axial values for acceleration and angular velocity, while **Xing Tan et al** [49] **extracted 4 new features** from the tri-axial acceleration values. Due to the nature of the data collected from reflexive markers and force plate sensors, **Lempereur et al** [48] **utilized positional 3D data and ground force reaction data**.

5.7 Conclusion

Having obtained a **MCC score of 0.538386,** this dissertation concludes that a not so satisfactory result was obtained. Although by observing the resulting confusion matrix test data, Figure 5-20, a satisfactory accuracy 93.70% in the FC walking event classification was obtained, the same cannot be said about de 51.36% of accuracy when classifying the Initial Contact walking event.

This result was obtained through tri-axial accelerometer and gyroscope raw data from **one IMU placed at the lower back**, while performing walking activities. It was observed that the dataset collected was highly unbalanced because Initial Contact walking event timestep sequence is shorter than the Final Contact walking event timestep sequence. As a future suggestion, this dissertation proposed the implementation of data oversampling while preparing the sample window to feed the DL model.

Experimenting with different window timestep sizes, model parameters, architecture layers and composition is also a viable suggestion. Another suggestion is the addition of new features extracted from the data.

In future work by experimenting with these suggestions, it might be possible for this dissertation proposed solution to have more space to grow and to be improve.

6 Conclusion

The main objective of this work was the development of a DL framework capable of automatic detection of daily activities in people with Parkinson's Disease using kinematic-driven data.

In this work a CNN based architecture was implemented and can recognize daily activities such as *walking*, *standing still*, *sit on chair*, *get up from chair*, *lay on bed*, *get up from bed* and can identify *180º turns* performed by PD patients.

The developed work was based and inspired in an extensive literature review on ML classification algorithms and DL recognition algorithms applied to HAR problematic. This research helped to comprehend the so far implemented techniques potential and inherent limitations, providing insights about the promising approaches that could be implemented and the areas about HAR in PD more adept to enhancement.

A custom dataset was necessary to acquire, and a protocol to collect data was made. The dataset comprised of 18 PD patients (10 males, 8 females) who performed 9 different daily activities and repeating each trial 3 times. Tri-axial accelerometer and gyroscope data collected was using the single IMU sensor present in the waistband of +Sense Project. The data was posteriori manually labelled, resulting in an unbalanced dataset data which was used to train, evaluate, and test the CNN based algorithm proposed by this dissertation. Data from 13 patients was used as training, 2 patients for validation and 3 patients' data was used to test the final model composition.

To find the best parameters of model architecture a grid search process testing different number of window sample timesteps, number of filters, neurons quantity, data batch size, algorithm learning rate and layer dropout rate.

During this optimization stage, 10-Fold cross-validation was implemented to assess the model validity. In validation stage, the model was subjected to overfitting, underfitting study where Holdout cross-validation was used in the assessment. Finally, the model was retrained with all training data and was assessed by using the unknown data of the 3 remaining patients.

CNN based architecture proposed by this dissertation obtained F1 Score of 0.90892. Although satisfactory result was achieved, this dissertation supports the idea that there is still room to improve. Data balancing, improvements to the manual data labelling and experimentation of different parameter should be considered to further improve this dissertation proposed in future work.

As a complementary objective of this work a DL framework capable of automatic detection of walking events in people with Parkinson's Disease using kinematic-driven data was developed.

BiLSTM based architecture was implemented and is capable of recognizing walking events such as Initial Contact and Final Contact performed by PD patients during walking activity.

The developed work was based in a brief literature review on DL recognition algorithms applied to walking IC/FC detection problematic. This research helped to comprehend the so far implemented techniques potential and inherent limitations, providing insights about the promising approaches that could be implemented in walking IC/FC detection in PD.

An already existing dataset was used from +sense project. The dataset comprised of 40 PD patients who performed 2 different walking activities. Data from 26 patients was used as training, 6 patients for validation and 8 patients' data was used to test the final model composition. To find the best parameters of model architecture a grid search process testing different number of neurons quantity, data batch size, algorithm learning rate and layer dropout rate.

During this optimization stage, 10-Fold cross-validation was implemented to assess the model validity. In validation stage, the model was subjected to overfitting, underfitting study where Holdout cross-validation was used in the assessment. Finally, the model was retrained with all training data and was assessed by using the unknown data of the 18 remaining patients.

BiLSTM based architecture proposed by this dissertation obtained MCC Score of 0.538386. Although a satisfactory result was achieved while recognizing Final Contact walking events (93.70% Accuracy) a not so good result was obtained while classifying Initial Contact events (51.36% Accuracy). Data balancing, experimentation of different parameter should be considered to further improve this dissertation proposal in future work.

The presented work allowed to obtain answers for the addressed research Question (RQs):

RQ1: What is the importance of daily activities recognition for motor assessment in PD? How have human activities been detected? And of IC/FCs?

PD symptoms restricts the quality of life of the affected person. The more severe the symptoms are the more daily activities like walking, sitting or lay on bed are affected. Measuring the correlation between the disease symptoms and daily activities performance can be a good indicator on how disease is evolving or how well the patient is responding to medicine or therapy. As such, it is important the existence of a tool capable of continuously monitoring the PD patient, enabling a personalized and accurate support to the patient from the clinical expert.

Due to its low cost, low power consumption, small size, light weight, easy usability, and accessibility, and not intrusive to a person, IMUs are frequently used as monitoring tools. IMUs integrates

accelerometer and gyroscopes that produces kinematic driven data. This produced kinematic data have been explored and utilized, with success, by researchers to perform not only HAR but also, in a more specific way, the recognition of IC/FC walking events.

With the ability to recognize IC/FC events, walking associated metric such as velocity, step length/time, cadence, walking variability or walking asymmetry can be calculated. This data can then be used by clinical experts to improve a patient diagnosis or by clinical supportive tools capable of stimulating the patient walking capability.

RQ2: Which DL model produces best result in recognizing daily human activities? And in IC/FC walking events recognition?

A CNN model architecture fed with raw kinematic data with a sliding window approach of length 240, attained the most promising results (0.90518 F1 Score and 0.90139 Accuracy) in the recognition of human activities performed by PD patients.

For the Human walk IC/FC recognition a BiLSTM model architecture fed with raw kinematic data and with a sliding window length of 5, achieved the better results (0.8455 F1 Score, 0.78115 Accuracy and 0.53839 MCC score).

RQ3: How effective and robust are the proposed DL solutions toward new patients?

The CNN model for HAR of PD patients was tested with unknown data from three different PD patients. The testing dataset contained data of three trials for each of the following activities: walking, 180º turning, sit on chair, get up chair, lay on bed and get up from bed. The CNN model obtained a performance of 0.90892 F1 Score and an Accuracy of 0.90743.

The BiLSTM model for HW IC/FC detection was tested with unknown data from six different PD patients. The dataset contained two trial of walking activity for each different patient. The BiLSTM model attained a 0.53839 MCC Score, F1 Score of 0.84155 and an Accuracy of 0.78115.

Promising results were attained for both DL frameworks, however this dissertation suggests that there is still space for further improvements. Enrich the dataset with more data from different patient, data balancing and feature extraction techniques, experimenting new models' architectures should be considered in future works.

7 References

[1] M. Kazemimoghadam and N. P. Fey, “Continuous Classification of Locomotion in Response to Task Complexity and Anticipatory State,” *Front. Bioeng. Biotechnol.*, vol. 9, no. April, pp. 1–13, 2021, doi: 10.3389/fbioe.2021.628050.

[2] J. M. Beitz, “School of Nursing-Camden, Rutgers University, 311 N. 5,” *Front. Biosci.*, vol. 6, no. 3, pp. 65–74, 2014.

[3] Associação Portuguesa de Doentes de Parkinson, “Manual Para Pessoas Com Parkinson,” *Msd*, pp. 6, 7, 8, 2014, [Online]. Available: http://msd.pt/wp-content/uploads/2015/10/Parkinson-Manual_XXXX_v7_pt.pdf

[4] W. H. Organization, “Parkinson disease,” 2022. https://www.who.int/news-room/fact-sheets/detail/parkinson-disease

[5] J. J. Ferreira *et al.*, “Prevalence of Parkinson’s disease: a population-based study in Portugal,” *Eur. J. Neurol.*, vol. 24, no. 5, pp. 748–750, 2017, doi: 10.1111/ene.13273.

[6] C. Jobanputra, J. Bavishi, and N. Doshi, “Human activity recognition: A survey,” *Procedia Comput. Sci.*, vol. 155, no. January 2019, pp. 698–703, 2019, doi: 10.1016/j.procs.2019.08.100.

[7] S. Nasiri and M. R. Khosravani, “Progress and challenges in fabrication of wearable sensors for health monitoring,” *Sensors Actuators, A Phys.*, vol. 312, p. 112105, 2020, doi: 10.1016/j.sna.2020.112105.

[8] C. Fadillioglu, B. J. Stetter, S. Ringhof, F. C. Krafft, S. Sell, and T. Stein, “Automated gait event detection for a variety of locomotion tasks using a novel gyroscope-based algorithm,” *Gait Posture*, vol. 81, no. November 2019, pp. 102–108, 2020, doi: 10.1016/j.gaitpost.2020.06.019.

[9] S. Baskar, P. Mohamed Shakeel, R. Kumar, M. A. Burhanuddin, and R. Sampath, “A dynamic and interoperable communication framework for controlling the operations of wearable sensors in smart healthcare applications,” *Comput. Commun.*, vol. 149, no. October 2019, pp. 17–26, 2020, doi: 10.1016/j.comcom.2019.10.004.

[10] M. Kazemimoghadam and N. P. Fey, “An Activity Recognition Framework for Monitoring Non-Steady-State Locomotion of Individuals with Parkinson’s Disease,” pp. 1–4, 2021.

[11] D. G. M. Zwartjes, T. Heida, J. P. P. Van Vugt, J. A. G. Geelen, and P. H. Veltink, “Ambulatory monitoring of activities and motor symptoms in Parkinsons disease,” *IEEE Trans. Biomed. Eng.*, vol. 57, no. 11, pp. 2778–2786, 2010, doi: 10.1109/TBME.2010.2049573.

[12] S. Pardoel, J. Kofman, J. Nantel, and E. D. Lemaire, “Wearable-sensor-based detection and prediction of freezing of gait in parkinson’s disease: A review,” *Sensors (Switzerland)*, vol. 19, no. 23, pp. 1–37, 2019, doi: 10.3390/s19235141.

[13] N. Shawen *et al.*, "Role of data measurement characteristics in the accurate detection of Parkinson's disease symptoms using wearable sensors," *J. Neuroeng. Rehabil.*, vol. 17, no. 1, pp. 1–14, 2020, doi: 10.1186/s12984-020-00684-4.

[14] D. Ravi, C. Wong, B. Lo, and G. Z. Yang, "A Deep Learning Approach to on-Node Sensor Data Analytics for Mobile or Wearable Devices," *IEEE J. Biomed. Heal. Informatics*, vol. 21, no. 1, pp. 56–64, 2017, doi: 10.1109/JBHI.2016.2633287.

[15] M. N. S. Zainudin, M. N. Sulaiman, N. Mustapha, and T. Perumal, "Activity recognition based on accelerometer sensor using combinational classifiers," *ICOS 2015 - 2015 IEEE Conf. Open Syst.*, pp. 68–73, 2016, doi: 10.1109/ICOS.2015.7377280.

[16] D. Ravi, C. Wong, B. Lo, and G. Z. Yang, "Deep learning for human activity recognition: A resource efficient implementation on low-power devices," *BSN 2016 - 13th Annu. Body Sens. Networks Conf.*, pp. 71–76, 2016, doi: 10.1109/BSN.2016.7516235.

[17] M. Inoue, S. Inoue, and T. Nishida, "Deep recurrent neural network for mobile human activity recognition with high throughput," *Artif. Life Robot.*, vol. 23, no. 2, pp. 173–185, 2018, doi: 10.1007/s10015-017-0422-x.

[18] A. Murad and J. Y. Pyun, "Deep recurrent neural networks for human activity recognition," *Sensors (Switzerland)*, vol. 17, no. 11, 2017, doi: 10.3390/s17112556.

[19] Y. Guan and T. Plötz, "Ensembles of Deep LSTM Learners for Activity Recognition using Wearables," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 2, pp. 1–28, 2017, doi: 10.1145/3090076.

[20] C. Xu, D. Chai, J. He, X. Zhang, and S. Duan, "InnoHAR: A deep neural network for complex human activity recognition," *IEEE Access*, vol. 7, pp. 9893–9902, 2019, doi: 10.1109/ACCESS.2018.2890675.

[21] M. Batool, A. Jalal, and K. Kim, "Sensors Technologies for Human Activity Analysis Based on SVM Optimized by PSO Algorithm," *2019 Int. Conf. Appl. Eng. Math. ICAEM 2019 - Proc.*, pp. 145–150, 2019, doi: 10.1109/ICAEM.2019.8853770.

[22] F. M. Rueda, R. Grzeszick, G. A. Fink, S. Feldhorst, and M. Ten Hompel, "Convolutional neural networks for human activity recognition using body-worn sensors," *Informatics*, vol. 5, no. 2, pp. 1–17, 2018, doi: 10.3390/informatics5020026.

[23] Y. Kwon, K. Kang, and C. Bae, "Unsupervised learning for human activity recognition using smartphone sensors," *Expert Syst. Appl.*, vol. 41, no. 14, pp. 6067–6074, 2014, doi: 10.1016/j.eswa.2014.04.037.

[24] F. Tahavori *et al.*, "Physical activity recognition of elderly people and people with Parkinson's (PwP) during standard mobility tests using wearable sensors," *2017 Int. Smart Cities Conf. ISC2 2017*, 2017, doi: 10.1109/ISC2.2017.8090858.

[25] K. D. Garcia *et al.*, "An ensemble of autonomous auto-encoders for human activity recognition," *Neurocomputing*, vol. 439, pp. 271–280, 2021, doi: 10.1016/j.neucom.2020.01.125.

[26] M. Straczkiewicz, P. James, and J. P. Onnela, "A systematic review of smartphone-based human activity recognition methods for health research," *npj Digit. Med.*, vol. 4, no. 1, pp. 1–15, 2021, doi: 10.1038/s41746-021-00514-4.

[27] M. Fusca, F. Negrini, P. Perego, L. Magoni, F. Molteni, and G. Andreoni, "Validation of a wearable IMU system for gait analysis: Protocol and application to a new system," *Appl. Sci.*, vol. 8, no. 7, pp. 1–16, 2018, doi: 10.3390/app8071167.

[28] H. R. Gonçalves, A. Rodrigues, and C. P. Santos, "Gait monitoring system for patients with Parkinson's disease," *Expert Syst. Appl.*, vol. 185, no. June, p. 115653, 2021, doi: 10.1016/j.eswa.2021.115653.

[29] C. F. Caiafa, Z. Sun, T. Tanaka, P. Marti-Puig, and J. Solé-Casals, "Machine learning methods with noisy, incomplete or small datasets," *Appl. Sci.*, vol. 11, no. 9, pp. 0–3, 2021, doi: 10.3390/app11094132.

[30] M. Kazemimoghadam and N. P. Fey, "An Activity Recognition Framework for Monitoring Non-Steady-State Locomotion of Individuals with Parkinson's Disease".

[31] B. Nguyen, Y. Coelho, T. Bastos, and S. Krishnan, "Trends in human activity recognition with focus on machine learning and power requirements," *Mach. Learn. with Appl.*, vol. 5, no. April, p. 100072, 2021, doi: 10.1016/j.mlwa.2021.100072.

[32] German Research Centre for Artificial Intelligence and A. Reiss, "PAMAP2 Physical Activity Monitoring Data Set," 2012. https://medium.com/@arifwicaksanaa/pengertian-use-case-a7e576e1b6bf

[33] D. Roggen, A. Calatroni, L.-V. Nguyen-Dinh, R. Chavarriaga, H. Sagha, and S. Tejaswi Digumarti, "OPPORTUNITY Activity Recognition Data Set," *Systems, Man and Cybernetics Workshop on "Robust machine learning techniques for human activity recognition,"* 2011. https://archive.ics.uci.edu/ml/datasets/opportunity+activity+recognition

[34] J. L.Reyes, D. Anguita, A. Ghio, L. Oneto, and X. Parra, "Human Activity Recognition Using Smartphones Data Set," *A Public Domain Dataset for Human Activity Recognition Using Smartphones*.

https://archive.ics.uci.edu/ml/datasets/human+activity+recognition+using+smartphones

[35] J. Brownlee, "1D Convolutional Neural Network Models for Human Activity Recognition," *Deep Learning for Time Series*, 2018. https://machinelearningmastery.com/cnn-models-for-human-activity-recognition-time-series-classification/

[36] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," pp. 1–11, 2015, [Online]. Available: http://arxiv.org/abs/1511.08458

[37] M. Hossin and M. . Sulaiman, "A Review on Evaluation Metrics for Data Classification Evaluations," *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, no. 2, pp. 1–11, 2015.

[38] Z. C. Lipton, C. Elkan, and B. Narayanaswamy, "Thresholding Classifiers to Maximize F1 Score," 2014, [Online]. Available: http://arxiv.org/abs/1402.1892

[39] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-Validation," in *Encyclopedia of Database Systems*, Boston, MA: Springer US, 2009, pp. 532–538. doi: 10.1007/978-0-387-39940-9_565.

[40] A. Emam, M. Shalaby, M. A. Aboelazm, H. E. A. Bakr, and H. A. A. Mansour, "A Comparative Study between CNN, LSTM, and CLDNN Models in the Context of Radio Modulation Classification," *2020 12th Int. Conf. Electr. Eng. ICEENG 2020*, pp. 190–195, 2020, doi: 10.1109/ICEENG45378.2020.9171706.

[41] T. Zebin, M. Sperrin, N. Peek, and A. J. Casson, "Human activity recognition from inertial sensor time-series using batch normalized deep LSTM recurrent networks," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2018-July, no. September, pp. 1–4, 2018, doi: 10.1109/EMBC.2018.8513115.

[42] F. Cruciani *et al.*, "Feature learning for Human Activity Recognition using Convolutional Neural Networks: A case study for Inertial Measurement Unit and audio data," *CCF Trans. Pervasive Comput. Interact.*, vol. 2, no. 1, pp. 18–32, 2020, doi: 10.1007/s42486-020-00026-2.

[43] A. Mirelman *et al.*, "Gait impairments in Parkinson's disease," *Lancet Neurol.*, vol. 18, no. 7, pp. 697–708, 2019, doi: 10.1016/S1474-4422(19)30044-4.

[44] M. Politis, K. Wu, S. Molloy, P. G. Bain, K. R. Chaudhuri, and P. Piccini, "Parkinson's disease symptoms: The patient's perspective," *Mov. Disord.*, vol. 25, no. 11, pp. 1646–1651, 2010, doi: 10.1002/mds.23135.

[45] M. W. Whittle, "Gait Analysis An Introduction," *Elsevier*, vol. 4th Editio, p. https://news.ge/anakliis-porti-aris-qveynis-momava, 2007.

[46] C. . Vaughan, B. . Davis, and J. . O'Connor, *Dynamics of Human Gait*, 2nd ed. Cape Town, South Africa: Kiboho Publishers, 1992.

[47] P. Aqueveque, E. Germany, R. Osorio, and F. Pastene, "Gait segmentation method using a plantar pressure measurement system with custom-made capacitive sensors," *Sensors (Switzerland)*, vol. 20, no. 3, 2020, doi: 10.3390/s20030656.

[48] M. Lempereur *et al.*, "A new deep learning-based method for the detection of gait events in children with gait disorders: Proof-of-concept and concurrent validity," *J. Biomech.*, vol. 98, p. 109490, 2020, doi: 10.1016/j.jbiomech.2019.109490.

[49] H. X. Tan, N. N. Aung, J. Tian, M. C. H. Chua, and Y. O. Yang, "Time series classification using a modified LSTM approach from accelerometer-based data: A comparative study for gait cycle detection," *Gait Posture*, vol. 74, no. August, pp. 128–134, 2019, doi: 10.1016/j.gaitpost.2019.09.007.

[50] T. Zhen, L. Yan, and P. Yuan, "Walking gait phase detection based on acceleration signals using LSTM-DNN algorithm," *Algorithms*, vol. 12, no. 2, 2019, doi: 10.3390/A12120253.

[51] A. Branquinho, H. R. Goncalves, J. F. Pinto, A. M. Rodrigues, and C. P. Santos, "Wearable gait Analysis LAB as a biomarker of Parkinson's disease motor stages and Quality of life: A preliminary study," *2021 IEEE Int. Conf. Auton. Robot Syst. Compet. ICARSC 2021*, no. i, pp. 234–239, 2021, doi: 10.1109/ICARSC52212.2021.9429770.

[52] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[53] J. Donahue *et al.*, "Long-Term Recurrent Convolutional Networks for Visual Recognition and Description," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 677–691, 2017, doi: 10.1109/TPAMI.2016.2599174.

[54] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, 1997, doi: 10.1109/78.650093.

[55] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, pp. 1–13, 2020, doi: 10.1186/s12864-019-6413-7.

[56] G. Jurman, S. Riccadonna, and C. Furlanello, "A comparison of MCC and CEN error measures in multi-class prediction," *PLoS One*, vol. 7, no. 8, pp. 1–8, 2012, doi: 10.1371/journal.pone.0041882.

[57] D. Chicco, "Ten quick tips for machine learning in computational biology," *BioData Min.*, vol. 10, no. 1, pp. 1–17, 2017, doi: 10.1186/s13040-017-0155-3.

APPENDIX A

Experimental Protocol

Postural Assessment in Parkinson's Disease

Purpose:

OB1: Create an open-source multimodal dataset of the pull test and physical activities in Parkinson's Disease based on 3D motion data and kinematic-driven walking parameters acquisitions through wearable miniaturized inertial sensors.

OB2: Assess dynamic postural instability.

OB3: Automatic estimation of pull test score based on artificial intelligence models.

Study design:

- Cross-sectional study.

Local:

- Hospital of Braga – 2CA Braga Academic Clinical Center.

Study chronology:

- **T0:** Patients' selection and recruitment
- **T1:** Experimental procedure
- **T2:** Data analysis
- **T3:** Dissemination

Figure A-1 Study chronology (W: week).

Participants:

- **Number of participants**: 10 participants with PD

Table A-1 Inclusion and exclusion criteria.

Inclusion Criteria	Exclusion Criteria
- Diagnosis of PD according to the UK Parkinson's Disease Society Brain Bank criteria - Present Freezing of Walking (FoG) - Hoen & Yahr scale ≤ 3 - Age between 50-85 years old	- Co-morbid disorders likely to affect walking, including stroke, orthopedic disease, rheumatologic disease, neurological and musculoskeletal disorders, cardiovascular disease and pulmonary disorders and normal flexibility and muscle strength - Cognitive impairment - Obvious motor impairments - Cognitive impairments - Visual acuity deficits

	- Audiometric deficits - Other neurological disease - Pain that may affect walking - Be evaluated by the same neurologist

Material:

- Hoehn & Yar scale (H&Y)
- Unified Parkinson's Disease Rating Scale (UPDRS-III)
- Pull test score
- Participants demographic registration document
- +sense waistband: +senseMotion
- Xsens:
    - 4 sensors: right/left trunk, and front/back lower trunk (L3-L5 level) of participant
    - 2 sensors: right/left hand of evaluator
    - MTManager SW

Data acquisition and outcomes:

Table A-2 Acquired variables and respective necessary material.

Type	Variables	Material
Clinic	Disease stage	H&Y
	Motor disability	UPDRS-III
Demographic	Age [years]	Participants demographic registration document
	Gender [F/M/Non-binary]	
	Weight [Kg]	
	Height [cm]	
Motion	1. 3D motion data 2. Kinematic-driven walking parameters: ▪ Rhythm: step/stride time and stance/swing/double-support phase. ▪ Pace: step/stride length, velocity, and cadence. ▪ Variability: step length/time, velocity, and stance/swing phase standard deviation. ▪ Asymmetry: step length/time, velocity, and stance/swing phase asymmetry. 3. Postural related metrics: ▪ Trunk pitch and roll. ▪ Range of motion. ▪ Root mean square ▪ JERK.	+senseMotion Xsens

Experimental procedure:

Step 1. Participation informed consent signature.

Step 2. Record participants demographic data: age, gender, weight and heigh.

Step 3. Wear Xsens in participant and evaluator.

Step 4. Record and assess participants clinic scales (UPDRS-III except point 3.12).

Step 5. Wear Xsens.

Step 6. Wear +sense.

Step 7. Turn on +sense and pair with +S APP (**Error! Reference source not found.**).

Step 8. Explain the first task (point 3.12 of UPDRS-III – pull test) to perform on data acquisition discrete type (**Error! Reference source not found.** and Figure A-2) and, if necessary, demonstrate it.

Step 9. In +S APP configure +sense to execute a motion monitoring acquisition as described in **Error! Reference source not found.**.

Step 10. Connect +sense with Xsens base station via wire.

Step 11. **Start data acquisition:**

- **a.** Press start button in +S APP (**Error! Reference source not found.**).
- **b.** Confirm in Xsens desktop if started data acquisition and disconnect +sense to Xsens base station.
- **c.** Indicate to the participant to perform the explained task.
- **d.** Record the moment where participants start a new activity transition.

Step 12. **Finish data acquisition:**

- **a.** Indicate to the participant to finish the first motion trial.
- **b.** Confirm in Xsens desktop if stopped data acquisition.
- **c.** Press stop button in +S APP (**Error! Reference source not found.**).

Step 13. Repeat the procedure from **step 8** until subsequently complete the tasks indicated on **Error! Reference source not found.** referred to data acquisition discrete type.

Step 14. In +S APP, plot acquired data to confirm that no losses have occurred during trials acquisition (**Error! Reference source not found.**).

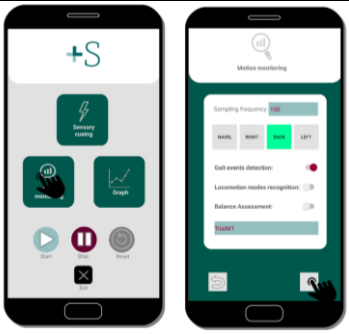
Step 15. Exit +S APP, turn off +sense and remove it from the participant (**Error! Reference source not found.**).

Step 16. Ask participant to fill out SUS questionnaire.

Step 17. Confirm acquired data in +S Desktop APP.

Estimated time per subject: ~30min

Table A-3 +S APP configuration for respective experimental step and +sense strategy.

+S APP configuration	Experimental procedure	+sense command
	Step 7	Bluetooth pairing
	Step 9	Motion monitoring
	Step 11	Start trial
	Step 12	Stop trial

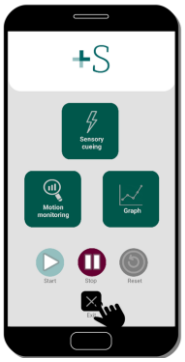
+S APP configuration	Experimental procedure	+sense command
	Step 14	Plot acquired data
	Step 15	Exit +SAPP

Table A-4 Data acquisition type (discrete and continuous), physical activity and corresponding tasks, and estimated and total estimate time.

Activity		Tasks	Estimated time Total estimated time Per trial	 Total
Pull-test (3.12 point of UPDRS-III)			15sec	~1min
Human motor-related activities	(1) Sit in a chair	10sec stand + sit in a chair + 10sec sitting	25sec	~12min
	(2) Get up from a chair	10sec sitting + get up + 10sec stand	25sec	
	(3) Sit in a chair	10sec stand + sit in a chair + 10sec sitting	25sec	
	(4) Get up from a chair	10sec sitting + get up + 10sec stand	25sec	
	(5) Sit in a chair	10sec stand + sit in a chair + 10sec sitting	25sec	
	(6) Get up from a chair	10sec sitting + get up + 10sec stand	25sec	
	(7) Lie on a bed	10sec stand + lie on a bed + 10sec lying	25sec	
	(8) Get up from a bed	10sec lying + get up + 10sec stand	25 sec	
	(9) Lie on a bed	10sec stand + lie on a bed + 10sec lying	25sec	
	(10) Get up from a bed	10sec lying + get up + 10sec stand	25 sec	
	(11) Lie on a bed	10sec stand + lie on a bed + 10sec lying	25sec	
	(12) Get up from a bed	10sec lying + get up + 10sec stand	25 sec	
	(13) Walk	10sec stand + 10m walk + 10sec stand	10sec	
	(14) Walk	10sec stand + 10m walk + 10sec stand	10sec	
	(15) Walk	10sec stand + 10m walk + 10sec stand	10sec	
	(16) 180° turning	5sec lying + 10m walk + right 180° turning +	75sec	

		10m walk + left 180° turning + 10m walk + right 180° turning + 10m walk + left 180° turning + 10m walk + right 180° turning + 10m walk + left 180° turning + 10m walk + 5sec lying		

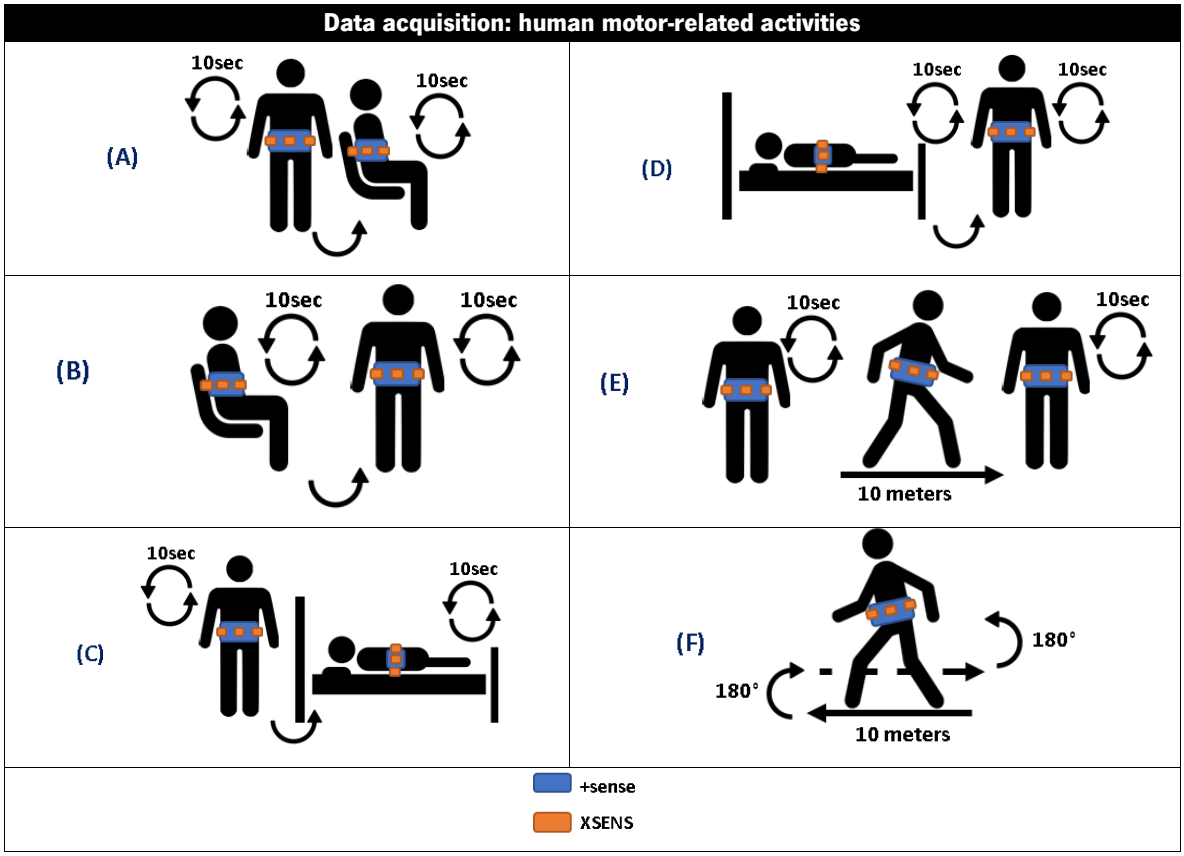


Figure A-2 - Experimental procedure tasks for human motor-related activities: (A) Sit in a chair, (B) Get up from a char, (C) Lie on a bed, (D) Get up from a bed, (E) Walk/Stand, (F) Right/Left 180° turning.

Data analysis:

Table A-5 Variables and methods designation to achieve respective study purposes.

Purpose	Variables	Method
Create an open-source multimodal dataset of physical activities in Parkinson's Disease based on 3D motion data and kinematic-driven walking parameters acquisitions through on wearable miniaturized inertial sensors.	Motion data	Open-source database guidelines (https://www.dbta.com/Editorial/Trends-and-Applications/A-Practical-Guide-to-Adopting-an-Open-Source-Database-for-Enterprise-IT-Use-66368.aspx)
Assess posture instability and estimate the pull test score using the created dataset.		AI-based models and statistical analysis
Descriptive and visual analysis of clinic and demographic data.	Clinic and demographic data	Statistical descriptive (mean ± standard deviation) *

*SPSS will be used to accomplish the statistical analysis.

APPENDIX B

1. Framework for HAR model Fine-Tunning Stage different Size Window input experimentation results

Table B-1 Fine-tunning best result using an input window size of 5.

Window size	Parameters	Value	Dataset	F1 score (mean)	Standard deviation
5	Epochs	200	Train	0.89136	0.00185
	Overlapping	50 %			
	Number of filters	128			
	Number of neurons	100			
	Pooling size	2	Validation	0.88845	0.00225
	Data batch size	512			
	Dropout rate	0.3			
	Model learning rate	0.001			

Table B-2 Fine-tunning best result using an input window size of 56.

Window timestep size	Parameters	Value	Dataset	F1 score (mean)	Standard deviation
56	Epochs	200	Train	0.91341	0.00204
	Overlapping	50 %			
	Number of filters	128			
	Number of neurons	100			
	Pooling size	2	Validation	0.90182	0.00504
	Data batch size	256			
	Dropout rate	0.5			
	Model learning rate	0.0001			

Table B-3 Fine-tunning best result using an input window size of 112.

Window timestep size	Parameters	Value	Dataset	F1 score (mean)	Standard deviation
112	Epochs	200	Train	0.92215	0.00243
	Overlapping	50 %			
	Number of filters	128			
	Number of neurons	60			
	Pooling size	2	Validation	0.90793	0.00533
	Data batch size	256			
	Dropout rate	0.5			
	Model learning rate	0.0001			