DIFFERENTIAL DIAGNOSIS OF NEURODEGENERATIVE DISEASES USING TIMED PETRI NETS

Universidad del Cauca ®

Master of Science research dissertation

CRISTIAN DAVID TOBAR MONTILLA

Director
PhD Carlos Felipe Rengifo Rodas
Co-director
PhD Mariela Muñoz Añasco

UNIVERSITY OF CAUCA
ELECTRONIC AND TELECOMMUNICATIONS ENGINEERING FACULTY
ELECTRONICS, INSTRUMENTATION AND CONTROL DEPARTMENT
RESEARCH GROUP ON INDUSTRIAL AUTOMATION
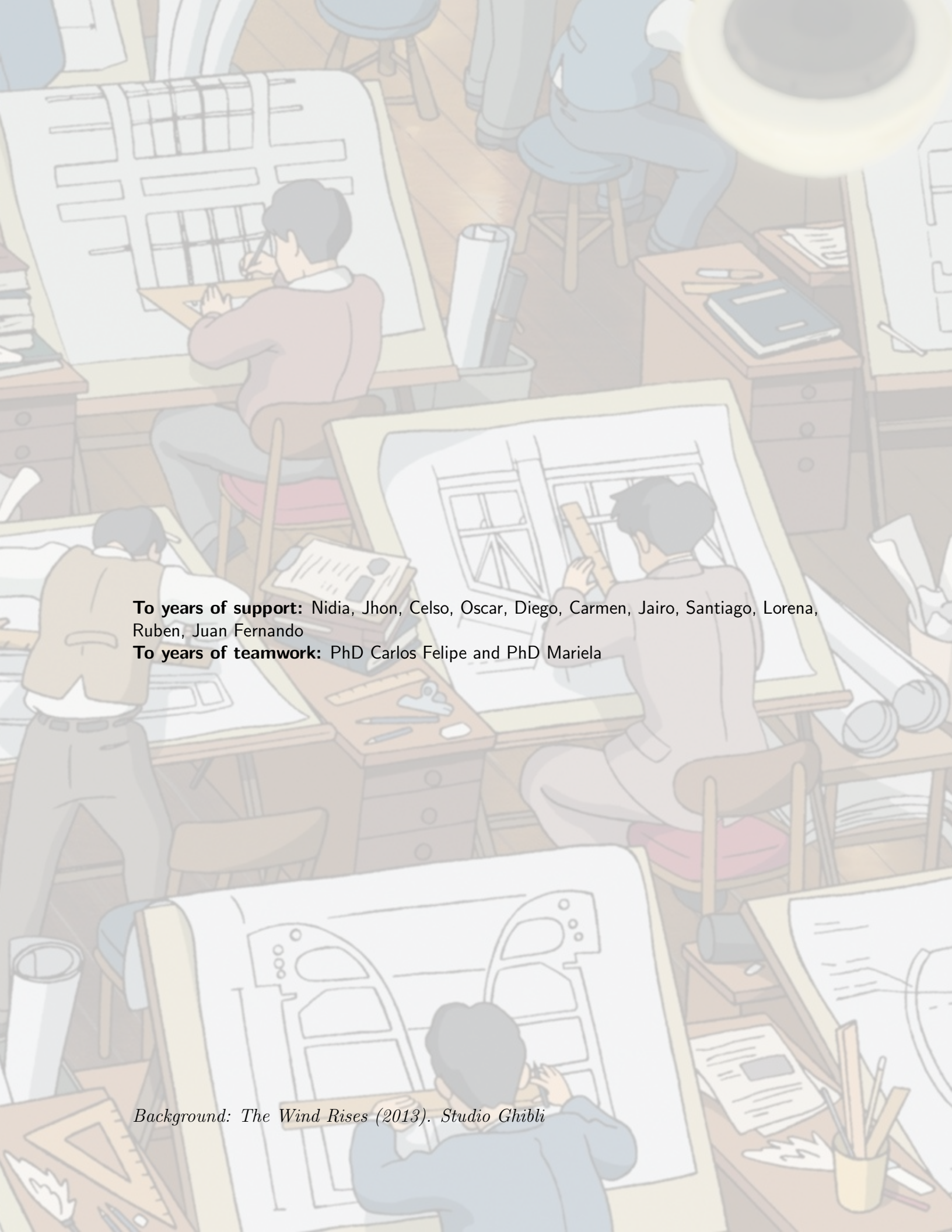MASTER OF SCIENCE IN AUTOMATICS
POPAYÁN, 2022

Cristian David Tobar Montilla

DIFFERENTIAL DIAGNOSIS OF NEURODEGENERATIVE DISEASES USING TIMED PETRI NETS

Thesis presented to the Faculty of Electronic Engineering and Telecommunications of the University of Cauca to obtain the title of

Master of Science in Automatics

Popayan,
2022

Background: The Wind Rises (2013). Studio Ghibli

Contents

List of Figures

List of Tables

Acronyms

ALS Amyotrophic Lateral Sclerosis 3, 5–8, 21, 25, 43

CG Control Group 3, 6, 7, 21

CNN Convolutional Neural Network 7

ConvLSTM Long-Short Term Memory Convolutional Neural Network 7

DES Discrete Event Systems 14

HD Huntington's Disease 3, 5–8, 21, 25, 43

HS Heel Strike 12, 16, 23, 25

IC Initial Contact 11, 27

IS Initial swing 11, 13

kNN K-nearest Neighbors Algorithm 6, 7

LR Loading Response 11, 12

MS Mid-swing 11, 13

MSE Multi-Scale Entropy 5

MSt Mid-Stance 11, 12

NNLS Non-negative Least Square 6

PD Parkinson's Disease 3, 5–8, 21, 25, 43

PS Pre Swing 11, 13

QR Quick Response 6

st-IPN Stochastic Timed Interpreted Petri Net 16, 17, 23, 25, 27

SVM Support Vector Machines 6

TO Toe-Off 12, 13, 16, 23, 25, 27

TS Terminal swing 11, 13

TSt Terminal Stance 11, 13

UPDRS Unified Parkinson's Disease Rating Scale 6

VGRF Vertical Ground Reaction Force vii, 6, 10, 13, 16, 19, 21–23, 27

Chapter 1

Preliminaries

1.1 Introduction

In neurodegenerative diseases (NDD), e.g., Parkinson's disease (PD), Huntington's disease (HD) and Amyotrophic Lateral Sclerosis (ALS), the dysfunction or death of nerve cells lead to the progressive deterioration of the nervous system. PD is related to the abnormal activity of dopamine-producing neurons, which affects the basal ganglia [6]. PD affects between one and two people in every 1,000 [7] with symptoms including irregular gait pace, rhythmic tremors at rest, and slow handwriting, walking, and blinking [6, 8]. HD, which affects 2.71 in every $100,000$ [9], is caused by the loss of nerve cells in the subcortical ganglia [10], resulting in muscle aches, learning difficulties, and feelings of irritability, sadness, apathy [11]. ALS affects five in $100,000$ individuals and results from damage to the motor neurons of the cerebral cortex, the brain stem, and the spinal cord [12]. Among its manifestations are gradual muscular atrophy, mobility difficulties, and cognitive impairment. Although these diseases affect different areas of the brain, the similarity in symptoms makes the diagnosis of PD, HD, and ALS problematic. Consequently, symptoms associated with PD may appear in advanced stages of both HD and ALS [13], [10], leading to 25% of misdiagnosis in people with HD or ALS [14].

The traditional diagnosis of neurodegenerative diseases involves the use of questionnaires and biomarkers. The downside of questionnaires, such as the Modified Parkinson Activity Scale [15], Unified Huntington's disease rating scale [16], Amyotrophic Lateral Sclerosis Functional Rating Scale [17] is their dependence on the expertise of the person applying the test [18, 19]. The use of biomarkers involves blood and genetic tests, functional neuroimaging, and biopsies of nerves and muscles [20, 21, 22], which are invasive and expensive and may involve long waits for results. In addition, the effectiveness of the approach remains a topic of study [23, 24, 25].

To overcome these issues, different alternatives have been proposed in the last few decades based on the predominant role of the brain regions affected by ALS, HD, and PD on the human gait [20, 26, 21, 22]. These approaches are based on the analysis of time series describing gait variables such as stride time, step length, and walking speed [27, 28].

Example of such techniques include those of [29, 30], who show that stride time variability significantly differs between groups with different neurodegenerative diseases. Other examples from [31, 32, 33] use trunk accelerations, arm movement asymmetry, and VGRFs, respectively, instead of stride time. The results reported by [34, 30, 31, 32, 33] should be understood not as a proposal for the diagnosis of neurodegenerative diseases but as a contribution to clarify the effect of these diseases on the human locomotor system.

The diagnosis of neurodegenerative diseases using pattern classification algorithms has been a major research topic in recent years. Such algorithms are commonly based on artificial neural networks, vector support machines, and Bayesian classifiers, among others. The following works focus on binary classifications that occur between healthy and pathological gait [35], between CG and PD [36, 37, 38, 39, 40, 41], or between CG and ALS groups [42]. Other approaches, such as [43, 44, 45, 46, 47, 48, 49], make multiple binary comparisons (CG versus PD, CG versus HD, and CG versus ALS). [50] proposed a signal processing approach named shifted one-dimensional local binary patterns to extract gait features and then train a multilayer perceptron and a random forest. These classifiers were used to differentiate between participants belonging to the CG and to PD groups. [51] trained four adaptive neuro fuzzy inference systems to perform binary classifications between CG and ALS, CG and HD, CG and PD, and normal and abnormal gait. [52] applies a radial basis function neural network to conduct the same binary classifications as [51]. [53] employ the gait features given by a recurrence quantification analysis to train six binary classifiers for the following pairs of classes: (CG, ALS), (CG, HD), (CG, PD), (ALS, HD), (ALS, PD), and (HD, PD). [54] trained the same six classifiers as [53] but using statistical and entropic features as inputs to a support vector machine and a multilayer feedforward neural network. [55] also use recurrence quantification analysis, but unlike [53], these features are used to differentiate between CG and PD groups rather than classifying subjects. [56], as [55], does not train classifiers but posits that long-range dependence, chaos, and information content in gait records are significantly different between healthy young, healthy older adults, and PD. In a clinical context, the previous binary classifiers are of little use since their output is uncertain when the evaluated individual falls outside the two categories for which the classifier was trained. For example, if a vector support machine learns to differentiate between CG and PD, it is not possible to know in advance how someone with HD or ALS will be classified.

A notable exception to the predominance of binary classifiers is the work of [5], who proposed a multilevel classification approach comprising a master classifier whose output is CG or NDD. When the output of this classifier is NDD, the gait features are simultaneously processed by three binary subclassifiers: ALS versus PD, ALS versus HD, and HD versus PD. The final class is defined by the voting algorithm described in [57]. The main inconvenient of a pipeline of classifiers is that the overall accuracy is the product of accuracies of each step. Hence, the accuracy of a sequence of two classifiers with 90% of accuracy each is only 81%.

Given these limitations, this study focuses on the differential diagnosis using a RF

and a CT with decision rules based on the transition times of a timed Petri net [58, 59, 60] model of human gait. The latter is used to represent systems whose evolution is defined by deterministic or stochastic transition times. These allow the representation of systems composed of sequential and parallel events, whose evolution is defined exclusively in function of transition times that can be either deterministic or stochastic. In From the above, the research question arises: *What should be the constituent components of a method of diagnosis of the neurodegenerative diseases PD, HD and ALS, based on timed Petri nets that model the human gait?*

1.2 Objectives

1.2.1 Main objective

To propose a differential diagnosis method for the neurodegenerative diseases Parkinson's Disease (PD), Huntington's Disease (HD) and Amyotrophic Lateral Sclerosis (ALS), based on a timed Petri net model of human gait.

1.2.2 Specific objectives

- To identify a timed Petri net model of the duration of the balancing period and of the subphases that compose the support period of the human gait cycle of four groups of people (Control Group (CG), PD, HD and ALS), from a database of reaction forces between the foot and the ground.

- To contrast both the language generated and the statistical moments of the transition times of each of the four identified timed Petri nets, with a view to establishing whether there are significant differences between these parameters.

- To assess the accuracy, sensitivity, and specificity of a classifier based on the agreement between the probability density functions of the transition times of an individual's gait model and that of the four groups under study (CG, PD, HD and ALS).

1.3 Structure of the document

This dissertation is organised as follows: Chapter 1 develops the introduction, context, research problem and research question, in addition, presents the objectives for the research development; Chapter 2 presents the literature review and the background for this study; Chapter 3 presents the database and develops the data preprocessing, the method for event generation, Petri net modelling and statistical procedures; Chapter 4 presents the results and Chapter 5 presents the discussion and future studies.

Chapter 2

State of the Art and Background

2.1 State of the Art

This state of the art comprises four sections. First, studies aimed to characterise gait parameters as indicators of the presence of a NDD. The second section describes binary classifiers aimed at determining whether an individual has certain NDD. The third section presents the works that address the duration of the gait phases as a NDD diagnosis marker. The fourth section presents the research gaps detected.

2.1.1 Characteristic parameters of groups with neurodegenerative diseases

In [29] and [30] it was determined that the correlation function between elements of the time series obtained from the variability of the stride time varies significantly between groups with different neurodegenerative diseases [29]. In [31] was observed that people with sleep behavior disorder during the phase of rapid movement of the eyes, a condition that in 50 % of cases precedes PD, presented a greater variability in accelerations of the trunk in comparison with the subjects of the control group. In [32], it was observed that those with PD had greater asymmetry in arm movement and greater variability in stride time. In [27], it was observed that the duration of the single support, double support and swing periods are different among groups of people with PD, HD and ALS. [56] reported differences of gait due to age and pathology by measuring complexity parameters such as self-similarity, chaos, and randomness in the comparisons CG vs PD subjects, and healthy elderly subjects vs. PD. In addition, since these indicators are employed to identify non-linear patterns in human gait stride time series, said patterns proved statistical difference in the same set of comparisons. A balance and gait assessment was carried out in [28], from which it was concluded that patients diagnosed with PD, compared to control subjects, showed a significant reduction in gait speed and step length. In [33] a measure of the amount of information contained in a signal, called Multi-Scale Entropy (MSE) , was applied to the reaction forces between the foot and the ground, and it was deduced that the MSE was different between the groups with PD, HD and ALS.

2.1.2 Binary diagnosis of neurodegenerative diseases

In [35], joint positions, stride length and gait speed are estimated. Subsequently, these variables are used to train different variants of the K-nearest Neighbors Algorithm (kNN), and it is reported that when kNN is applied to the joint positions, using as a similarity criterion Dynamic Time Warping, a 100% classification accuracy is reached between healthy subjects and neurodegenerative diseases. In [43], the accuracy is 89.33%, but by training Support Vector Machines (SVM) with the reaction forces between the foot and the ground.

Binary classifiers have also been applied to the differentiation among individuals from a CG and from participants with PD. In [36], a SVM trained with stride time and duration of swing and simple support periods obtains 90.32% of successful classifications. In [37], this percentage is 94.1% for a Bayesian classifier that uses stride length and gait speed as input signals. Alike the previous research, in [41], the PD diagnosis is performed through the extraction of VGRF features during the stance period, such as VGRF peak value, VGRF peak delay, area under the curve, and the kurtosis of the signal. This information is fed to a SVM classifier, which reports an accuracy of 90.82% for PD discrimination. In [38], five machine learning methods are compared through the Cartesian positions of 15 reflective markers, and the following percentages of accuracy are obtained: Fisher linear discriminant (87.4%), Naive Bayesian classifier (82.0%), kNN (84.4%), support vector machines (86.0%) and random forests (92.6%). In [40], the authors detect the severity of PD, according to the Unified Parkinson's Disease Rating Scale (UPDRS), by using an SVM that processes spatiotemporal gait characteristics extracted from a dataset of VGRF. The VGRF classifies a subject with 98.65% accuracy into one of four categories: Healthy, PD severity 2, PD severity 2.5, and PD severity 3. In [39], an extension of the principal components method to tensors (multi-dimensional extensions of matrices) is applied to the reaction forces between the foot and the ground reaching a specificity and sensitivity of 100%.

The most recent works in the diagnosis of neurodegenerative diseases report the application of multiple binary classifiers. In [44], an algorithm named empirical signal decomposition mode is applied to the reaction forces between the foot and the ground of the four groups of participants analyzed (CG, PD, HD and ALS). The accuracy of the classifier was: CG-PD 94.9%, CG-HD 90.0%, and CG-ALS 93.4%. In [48], an automated method for detecting PD, HD and ALS based in localized time-frecuency information of gait signals is presented. This dual information, leads to the creation of a set of linear and non-linear features for the training of a multicategorical Non-negative Least Square (NNLS) classifier. The mentioned approach achieved accuracy rates of 93%, 94%, and 97% for PD, ALS, and HD detection, respectively. In [45], the authors used the same signals as in [44]; however, these are transformed into grayscale images applying the technique of diffuse analysis of recurrence. Subsequently, these images are used to train SVM, where accuracy percentages of 100% are obtained in the classifications: CG-PD, CG-HD, and CG-ALS.A proposal similar to [45] is presented in [49], where force signals were transformed into a Quick Response (QR) code. This 2D feature allows the employment of deep

learning techniques for classification such as the Long-Short Term Memory Convolutional Neural Network (ConvLSTM) and a 3D Convolutional Neural Network (CNN). Again, the classification problems are CG-PD, CG-HD and CG-ALS providing an accuracy value of 89.44% using the ConvLSTM approach and a value of 86.05% using the 3D CNN approach. [55] discriminated with a clear clustering and high separability between groups CG vs. PD through gait features obtained from the RQA technique applied in gait rhythm timeseries from five healthy elderly subjects and five PD individuals. [53] conducted six binary classification tasks ALS vs. CG, PD vs. CG, HD vs. CG, ALS vs. PD, ALS vs. HD, and PD vs. HD using Support Vector Machine (SVM) and Probabilistic Neural Network (PNN) techniques, with a set of twelve improved gait features extracted with Recurrence Quantification Analysis (RQA) and statistical measures, to raise the classifiers performances. Accuracies reported in the binary classifications activities ranges from 96% (ALS vs. CG) to 100% (remainder comparisons). In [54], from original VGRF signals, basic time-domain features such as mean, standard deviation, skewness, kurtosis, approximate entropy were obtained and fed a classifier based on Non-Negative Least Squares (NNLS) performing six binary classification tasks ALS vs. CG, PD vs. CG, HD vs. CG, ALS vs. PD, ALS vs. HD and PD vs. HD with accuracies ranging from 99% to 100% and a multiclass approach CG vs. ALS vs. PD vs. HD of 99.57% accuracy. [51] proposed a classifier based on the Adaptive Neuro-fuzzy Inference System (ANFIS) model. This classifier received features obtained from five gait cadence timeseries such as left stride interval, right stride interval, left stance interval, right stance interval, and double support interval. The three binary classifications tasks ALS vs. CO, PD vs. CO, and HD vs. CO reported accuracies of 93.10%, 90.22%, and 94.44% respectively. [52] presented a classifier based on radial basis function neural networks and performed the same classification tasks as in [51] with accuracies of 93.1%, 100% and 100% respectively. The left and right swing and stance interval timeseries served as input. In [47], a topological data analysis was applied to the duration of the stride time in order to form the features vector that will be used as input to four algorithms: Decision Trees, Random Forests, Naive Bayesian classifier and kNN, which are compared in terms of their ability to perform the same four binary classifications as in [44]. For each of the 16 results (4 algorithms × 4 classifications) accuracy, sensitivity and specificity are presented. The superiority of one algorithm over another depends on the pathology being compared. Thus, for example, the random forests show better results than the decision trees for the CG versus PD classifier; however, this relationship is opposite for the CG versus HD classifier.

2.1.3 Multiclass Classifier

In [5], unlike the works previously mentioned, a classifier with four possible outputs such as CG, PD, HD or ALS is proposed. This proposal initially applied a binary classifier to determine whether a person belongs to the CG or has a neurodegenerative disease. In case of not belonging to the CG, the data is entered into three binary classifiers: PD versus HD, HD versus ALS, and ALS versus PD, and based on an algorithm named majority voting, the category PD, HD or ALS is selected. In this work, the same database of reaction forces

between the foot and the ground is used as in the works cited in the preceding paragraph. The authors report the accuracy, sensitivity, and specificity of the binary classifiers, but not the multilevel classifier, making it impossible to determine the usefulness of the tool.

2.1.4   Duration of the gait phases

In [61], the Lyapunov exponents for the load response phase were compared between youth and adults with no records of pathologies with effects on gait, and it was concluded that the older group presented greater instability. In [62], the effect of speed on the duration of the phases during the support period is studied, and it was concluded that these times decrease linearly as the walking speed of the subject increases. In [63], the conclusion obtained in [62] is endorsed, but this time alternating 12 gait speeds instead of 3.

2.1.5   Research gaps found

From the present state of the art, the absence of works that address the diagnosis between PD, HD and ALS from human gait signals can be evidenced. Additionally, the research regarding the incidence of neurodegenerative pathologies in the duration of gait phases [1] is scarce, as is the use of Petri nets for the analysis of human gait.

2.2   Background

2.2.1   Human gait

Gait is defined as *a method of locomotion that involves the use of two legs, alternately, to provide support and propulsion* [2] allowing human beings to move from one place to another. When a subject walks, he moves from one position to another, and, while he develops this action, the human body can be assumed as a mass subjected to translations generated by the movement of the lower extremities. In biomechanical terms, the human gait is composed by a set of interrelated movements, moments, powers, forces, pressures, coordination and balance [64, 65, 66]. In fact, the latter are some of the variables that can be processed by a specific gait analysis method for characterisation and diagnosis purposes. [64] [65] [66]

2.2.2   Gait cycle

Step and stride

As can be seen in Figure 2.1, when a lower limb goes forward and has contact with the ground, it has made a *step*. For example, when the right limb moves forward and touches the support surface, a right step has taken place. This can be perceived symmetrically in

---

[1]The phases in which the single and double support periods are divided
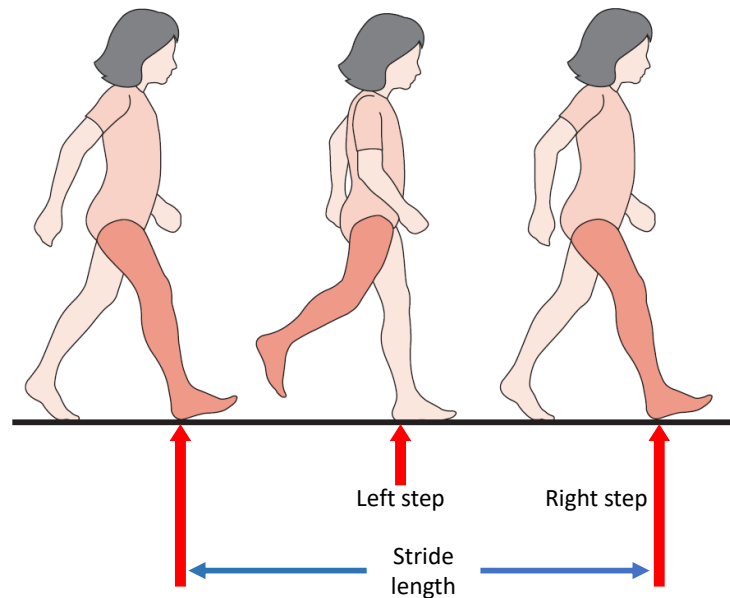
Figure 2.1: Occurring of a step (left or right) and a stride within the gait cycle. Source from [1].

the left foot. In fact, the *step length* is defined as the distance from the heel of the lagging member to the heel of the leading member.

The gait (or stride) cycle is defined as the time interval that occurs, starting with the contact of the heel of one foot with the ground, until the new contact with the ground with the heel of the same foot. This cycle is repeated on both feet, but it happens out of phase with each other. For this work, it is decided to take as a reference the contact of the heel of the right foot as the beginning and end of the gait cycle. Consequently, the gait cycle of the left foot is analogous but half cycle out of phase [2].

The beginning of the cycle (0%) occurs with the first contact (initial contact or heel strike in normal gait) of the right foot, and the end of the cycle (100%) occurs with the next contact of the same foot, which it will correspond to the start of the next cycle. As can be seen in Figure 2.2, in a normal symmetrical gait, toe-off occurs approximately 60% or 62 % of the cycle [2], dividing this in two: First, the *Support* period that lasts between 60% to 62% of the total cycle, and, the *Swing* period in which one foot swings and lasts about 38% to 40% [1] [2].

As can be seen in Figure 2.3, the right foot (*ipsilateral*) and the left foot (*contralateral*) are in a phase shift of around 50% [1]. For this reason, when one foot swings, the other rests on the ground. Also, there is a period when both feet contact on the ground. This period is known as *double support* and lasts about 20 % of the gait cycle [2]. However,
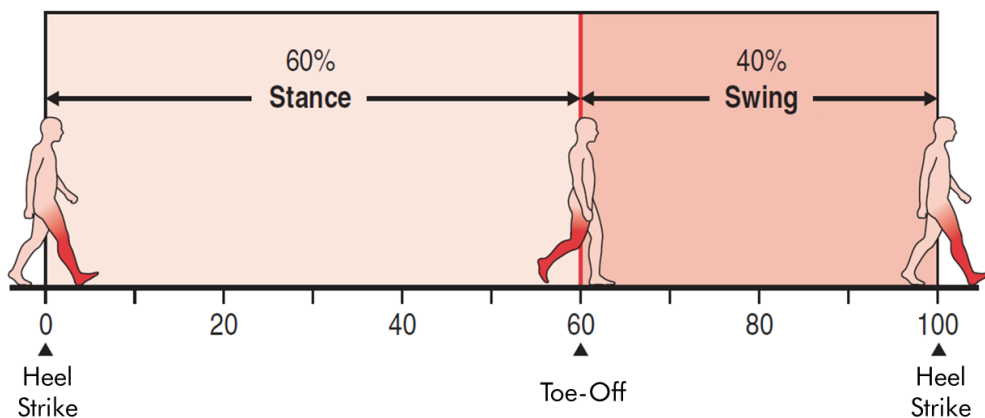
Figure 2.2: Division of the gait cycle into 2 periods: Stance and swing. Source [1].

all these times can change due to the walking speed. The swing period becomes proportionally longer, and the support and double support periods shorten as speed increases [67].

Due to the high complexity involved in doing an analysis on the gait, it can generally be subdivided into different phases (states) [68] [3] and events [65]. Table 2.1 shows the phase division of the gait cycle according to [3].

In summary, the right foot gait cycle is composed of two periods and eight phases as follows: The period of support comprises the phases: (For clarification purposes, the subindices l and r will refer to the left and right foot respectively). initial contact (IC_r), loading response (LR_r), mid-stance (MSt_r), terminal stance (TSt_r) and pre-swing (PS_r). The swing period comprises the phases: initial swing (IS_r), mid swing (MS_r), and terminal swing (TS_r). Similarly, the gait cycle of the left foot is also composed of a period of swing and stance, but with the succession of phases in the following order: PS_l IS_l MS_l TS_l IC_l LR_l MSt_l TSt_l.

2.3   Vertical Ground Reaction Forces

When the time comes to make a diagnosis about a patient, or it is desired to describe in depth the qualitative and quantitative variables of a gait cycle, it is necessary to define a technique that guarantees these objectives. For this reason, with the help of technological advances and from physical principles, a variety of measurement systems have been developed. This is the case of the VGRF.

As the weight of the body falls and moves across the surface of the foot, forces are

N°	Phase	Ocurrence		Purpose
		Temporal	Description	
1	**Initial Contact (IC)**	0%	This phase occurs the moment the ipsilateral foot touches the ground.	The lower limb is positioned to start the period of support with a plantar flexion of the ankle.
2	**Loading Response (LR)**	2%	This phase begins with the heel strike of the ipsilateral foot on the surface and continues until the contralateral foot is lifted for swing.	Shock absorption, weight bearing stability and preservation of gait progression
3	**Mid-Stance (MSt)**	10%	This phase is the first half of the single support interval. It begins as soon as the contralateral foot is lifted and continues until the weight of the body is aligned on the ipsilateral forefoot.	Progression on stationary foot, stability of limbs and trunk
4	**Terminal Stance (TSt)**	30%	This phase completes the single support interval. Begin with raising the heel of the ipsilateral foot and continue until the contralateral foot touches the ground. During this entire phase, the weight of the body is shifted in front of the forefoot.	Progression of the body beyond the supporting foot.
5	**Pre Swing (PS)**	50%	This is the final phase of stance and is the beginning of the double support interval of the cycle. It begins with the heel-strike of the contralateral foot and ends with the detachment of the toes of the ipsilateral lower limb.	Positioning of the lower limb for swinging.
6	**Initial swing (IS)**	60%	This phase is approximately one third of the swing period. It begins with lifting the ipsilateral foot off the ground and ends when the swinging foot is opposite the contralateral foot in stance.	Clearance of the foot with respect to the ground and advance of the limb from its lagging position.
7	**Mid-swing (MS)**	73%	The second phase of the swing period begins when the swinging limb is opposite the lower limb is in stance. The phase ends when the swinging limb is forward and the tibia is vertical (i.e. the hip and knee flexion postures are equal).	Advance the limb and clear the foot with respect to the ground.
8	**Terminal swing (TS)**	87%	The final phase of the swing period begins with the tibia vertical and ends when the ipsilateral foot hits the floor. The advancement of the limb is completed when the leg advances the thigh.	Complete advancement of the limb and preparation of the lower limb for the stance.

Table 2.1: Division of phases of the gait cycle of the right foot and its temporal occurrence from [3]. In addition, its description and objective are presented within the cycle.
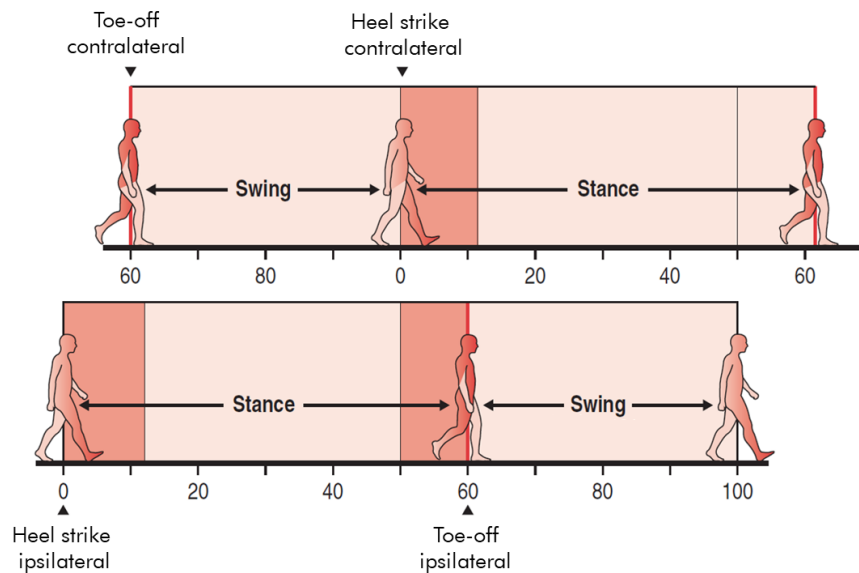
Figure 2.3: Gait cycle for the right foot (ipsilateral) and left foot (contralateral). The periods of double support of the cycle are highlighted in intense red. Source [1].

generated on the floor that can be measured using force platforms, which are fixed to ground, or forces insoles, which are fixed to feet.

The individual load can be measured in laboratories with force platforms. These devices use strain gauges or piezoelectric quartz crystals that convert force into electrical signals. The data are processed to obtain information on the related rotational moments, pressure centers and reaction force vectors. The main inconvenient of these devices is that subjects tend to modify their gait patterns according to the location of the force plates (Figure 2.4), and as a consequence, the resulting data are not representative of a typical gait. However, technology has advanced to the point that force sensors can be inserted into a shoe like a traditional insole. Thus, a natural gait process is guaranteed throughout the experiment.

Vertical load

The classic gait model proposed by [3] (Figure 2.5), manifest that the normal gait during the stance period contains two peaks separated by a local minimum. The value of the peaks is approximately 110 % of the body weight, while in the valley it is about 80 % of the weight [3]. In the same Figure, in blue, the active phases of the gait begin and end with the occurrence of critical points in the force signal (maximum and minimum); which are called Heel Strike (HS) , F1, F2, F3 and Toe-Off (TO) . At the beginning of the gait profile, the point HS takes place as indication of the first contact of the foot with the ground during the LR phase. Due to gravity, the body is falling which increases in force towards a first peak (F1) that initiates the MSt phase in response to weight acceptance
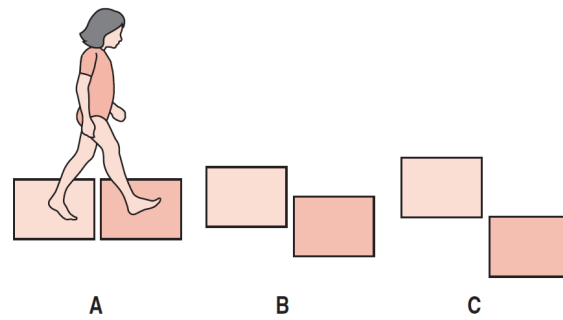
Figure 2.4: A, B and C, represent the usual arrangements of force platforms in laboratories. Source [1].

events during the loading response. At the beginning of the TSt, the valley (F2) appears, which is generated due to the translation the body's center of gravity over the plantigrade foot. The second peak (F3) occurs at the beginning of the PS and indicates an increase in acceleration in order to lift the weight of the body. Finally, the TO corresponds to the total take-off of the foot from the ground, starting the swing period. In addition, it can be notice the 50% phase shift in the occurrence of the events and the gait phases of the left foot compared to the right foot. Finally, given the absence of force during the swing period, in this research the IS, MS and TS phases are combined into a single swing phase.
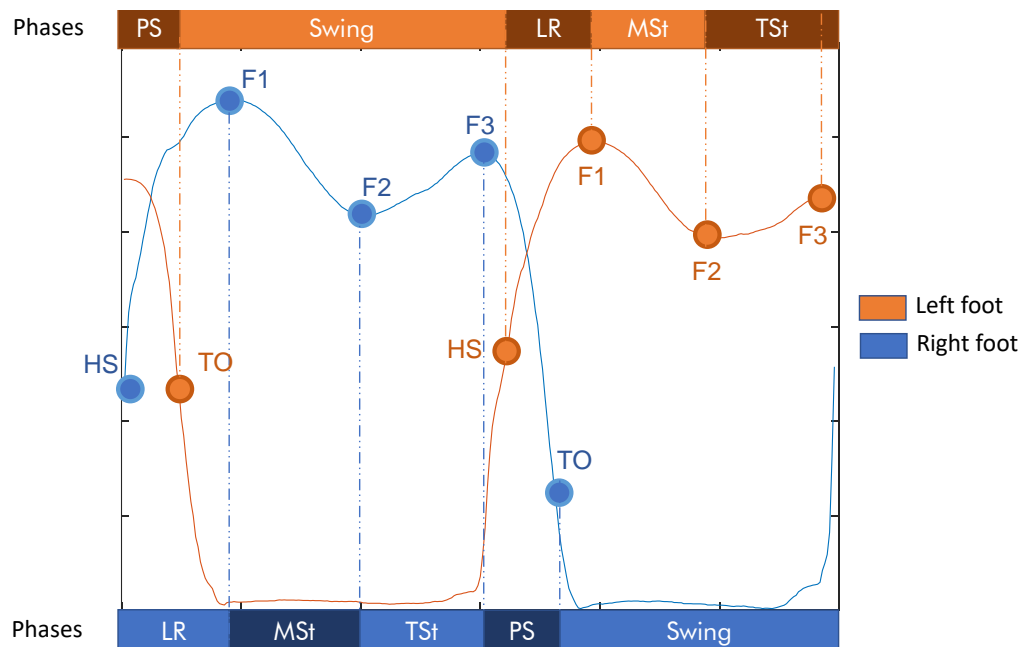
Figure 2.5: Phases of gait evidenced through variations in the left and right VGRF signals. Sources [2] [3].

2.4   Discrete Event Systems

2.4.1   Systems, State and Event

A system is a combination of components that work together to perform a function that cannot be performed by the individual parts [69].

Systems can be divided according to their behaviour over time into:

- Continuous Time Systems: The system receives as input a continuous signal, which is a signal defined for a continuous sequence of values, and transforms it into continuous output signals.

- Discrete Time Systems: The system receives discrete time signals as inputs, which is a signal defined in discrete times (periodic or aperiodic), and transforms it into discrete time output signals [70].

- Discrete Event Systems (DES): It is a dynamic system that varies its behaviour or state according to the abrupt occurrence, in possibly unknown intervals, of physical events [71].

- Hybrid Systems: It is a dynamic system that cannot be represented and analyzed with sufficient precision by the theory of continuous systems or the theories of discrete systems [72].

A DES is a dynamic system that evolves over time and under a discrete state space [71]. The change of states is due to the occurring of a transition which in general, its occurrence will be unpredictable.

2.4.2   Petri nets

A Petri net N is a bipartite directed graph represented by the tuple $N = (P, T, Pre, Post, M_0)$, in which:

- $P = \{p_1, p_2, \ldots, p_n\}$ is the finite set of places.

- $T = \{tr_1, tr_2, \ldots, tr_m\}$ is the finite set of transitions.

- Pre : It is a matrix of n rows and m columns, where the element $i, j \in \mathbf{Z}^+$ of the pre-occurrence matrix indicates the weight of the arc from place p_i to the transition tr_j. The element in row i and column j is equal to zero if there is no connection between the place p_i and the transition tr_j.

- $Post$: It is a matrix of n rows and m columns, where the element $i, j \in \mathbf{Z}^+$ of the post-incidence matrix indicates the weight of the arc from the transition tr_j to the place p_i. The element in row i and column j is equal to zero if there is no connection between the transition tr_j and the place p_i.

- $m_k : P \longrightarrow \mathbf{Z}^+$ is a function that assigns an integer number of marks to the place p_i after k transitions. $M_k = [m_k(p_1), m_k(p_2), \ldots m_k(p_n)]$ is a vector with the number of marks of each place in the network, after k transitions have occurred.

Graphically, places are represented by circles, transitions by rectangles, and arcs by arrows. The incidence matrix for N is $C = Post - Pre$.

In a Petri net, a tr_j transition is enabled to a marking M_k at time t_k if all the places p_i that precede it have a marking greater than or equal to the weight of their respective arc. Mathematically, this means that the condition $m_k(p_i) \geq Pre(i,j)$ must be fulfilled for $i = 1, 2, \ldots, n$. If a transition tr_j is enabled and fired, a new marking at a new instant of time M_{k+1} is reached. The new marking is calculated by $M_{k+1} = M_k + C.\overrightarrow{tr_j}$ where $\overrightarrow{tr_j}$ is a vector of zeros of size m, except in row j since tr_j was fired; this expression is known as the equation of state of a Petri net.

Figure 2.6 shows a Petri net with three places and four transitions. I.e., the sets P and T are $P = \{p_1, p_2, p_3\}$ $T = \{tr_1, tr_2, tr_3, tr_4\}$. The Pre and $Post$ matrices of the network are displayed in the Equations 2.1 and 2.2, which show the connections through arcs between places and transitions.
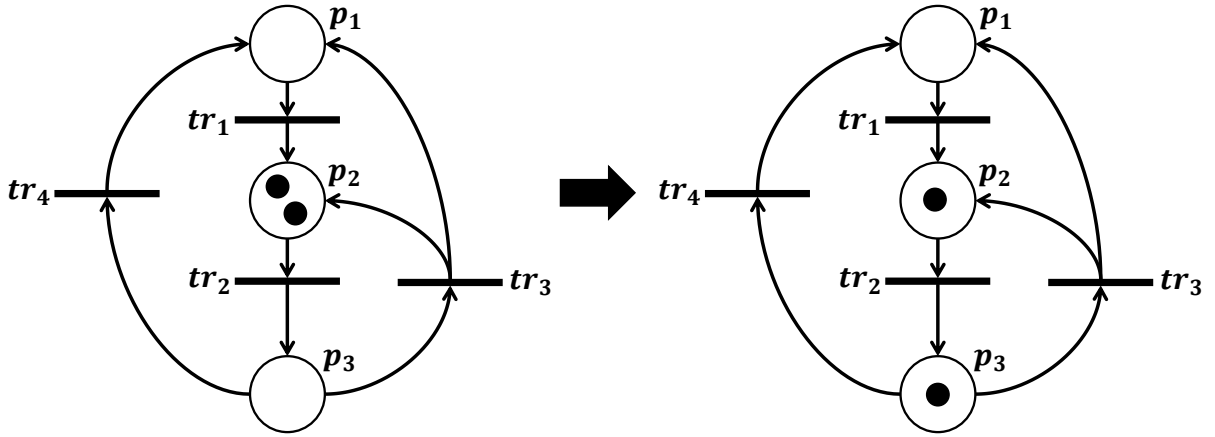
Figure 2.6: Petri net example.

$$Pre = \begin{array}{c} \\ p_1 \\ p_2 \\ p_3 \end{array} \begin{array}{cccc} tr_1 & tr_2 & tr_3 & tr_4 \\ \left(\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{array} \right) \end{array} \qquad (2.1)$$

$$Post = \begin{array}{c} \\ p_1 \\ p_2 \\ p_3 \end{array} \begin{array}{cccc} tr_1 & tr_2 & tr_3 & tr_4 \\ \left(\begin{array}{cccc} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{array} \right) \end{array} \qquad (2.2)$$

The evolution of this network is evidenced by the marking. The current marking (left side of the figure) is described as $M_0 = [0\ 2\ 0]^T$. As tr_2 is enabled since it meets the previously mentioned enabling marking condition it can be fired allowing the flow of a single mark which evolves to marking $M_1 = [0\ 1\ 1]^T$ as seen in Equation 2.3.

$$M_1 = M_0 + C\, tr_2$$

$$M_1 = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} + \left(\begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \right) \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \tag{2.3}$$

$$M_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

2.4.3   Petri net approach to model human gait

The dynamic system to be modeled is autonomous, i.e., it is a system whose evolution depends on time and its behavior is modeled from a sequence of events. Under this premise, in this research a Timed Interpreted Petri net is proposed, where first, the interpretation of the net is given in the assignment of a label on the net places that represent the phases of the gait cycle, and second, the events occurring time is associated to the transitions.

A Stochastic Timed Interpreted Petri Net (st-IPN) is composed by the tuple: $stQ = (N, X, \varphi, \tau, \lambda)$ where:

- N is a Petri net as previously defined.

- $X = \{x_1, x_2, \ldots x_5\}$ is the set of phases that comprise the running where x_1 : LR, x_2 : MSt, x_3 : TSt, x_4 : PS, x_5 : S as shown in Figure 2.5. In this work, the gait phases are obtained from the generation of events over the VGRF signals. For this reason, the Mid and Terminal Swing phases, seen in the figure, are not part of the X set since they are late phases of the balancing period.

- $\varphi : P \longrightarrow X$ is a labeling function that associates an item of X to each place, that is $\varphi(p_i) = x_i, i = 1, \ldots, 5$.

- $\tau = \{t_1, t_2, \ldots t_5\}$ is the set of random variables with normalized time from 0 to 1 where t_1 represents the instant of time in which the maximum of the derivative of the VGRF signal occurs, taking place the event HS. t_2 represents the first maximum of the VGRF signal which fires the $F1$ event. t_3 represents the global minimum of the VGRF signal where the event $F2$ occurs. t_4 is registered with the next maximum value of the force signal where the event $F3$ takes place, t_5 is detected when the minimum value of the derivative of the force signal occurs and indicates the occurrence of the event TO.

- $\lambda : T \longrightarrow \tau$ is a function that assigns a time element τ to each transition, that is $\lambda(tr_i) = t_i, i = 1, \ldots, 5$.

A $tr_j \in T$ transition is enabled if, in addition to fulfilling the marking condition described in section 2.4.2, the random variable t_j is within two standard deviations around the mean value of time in which it occurs tr_j.

If tr_m is enabled, then tr_m fires and reaches a new markup M_{k+1} represented by the equation of state 2.4 that represents said evolution.

$$M_{k+1} = M_k + C.\overrightarrow{tr_m} \tag{2.4}$$

Figure 2.7 shows the st-IPN that represents the behavior of human gait for both feet. The events are differentiated according to the sub-index: On the right foot is the sequence of events HS_r, $F1_r$, $F2_r$, $F3_r$ and TO_r and on the left foot is the sequence TO_l, HS_l, $F1_l$, $F2_l$ and $F3_l$. Each of the phases of the gait cycle is represented by the places in the net: On the right foot are LR_r, MSt_r, TSt_r, PS_r, S_r and on the left foot are PS_l, S_l, LR_l, MSt_l, TSt_l. The transitions are associated with both the label and the time intervals of the timed events that generates the phases change: tr_1 is fired at 10% when the event TO_l occurs, tr_2 fires at 10% when the event $F1_r$ occurs, tr_3 fires at 30% when the event $F2_r$ occurs, tr_4 fires at 50% when the $F3_r$ event occurs, tr_5 fires at 50% when the HS_l event occurs, tr_6 fires at 60% when the TO_r event occurs, tr_7 fires at 60% when the event $F1_l$ occurs, tr_8 fires at 80% when the event $F2_l$ occurs, tr_9 fires at 100% when the event HS_r occurs, tr_{10} is triggered at 100% when the event $F3_l$ occurs. The time intervals are obtained from Table 2.1. Table 2.2 shows the Pre and Post locations of each of the transitions as a function of the gait phases. In this net, it has been assumed that the current states, due to the presence of marks in the places, are the initial contact in the right foot and the pre-swing in the left foot.

2.5 Supervised Learning

2.5.1 Classification Trees

The classification tree (CT) is a supervised machine learning technique that concludes about a set of observations into into a discrete series of values (categories), employing a hierarchical structure composed of nodes and directed edges connected. The structure has an initial node called *root* with no incoming edges. All other nodes have a unique incoming edge. Now, if the node has outcoming edges then is called an internal or test node. Otherwise, it is a terminal node or *leaf* [4].

The test nodes splits into two or more nodes according to certain conditions of the data attributes labelled in the outcoming edges. These conditions may be nominal attributes values or, in the case of numeric values, the conditions are a range. The leaf nodes are

Left foot
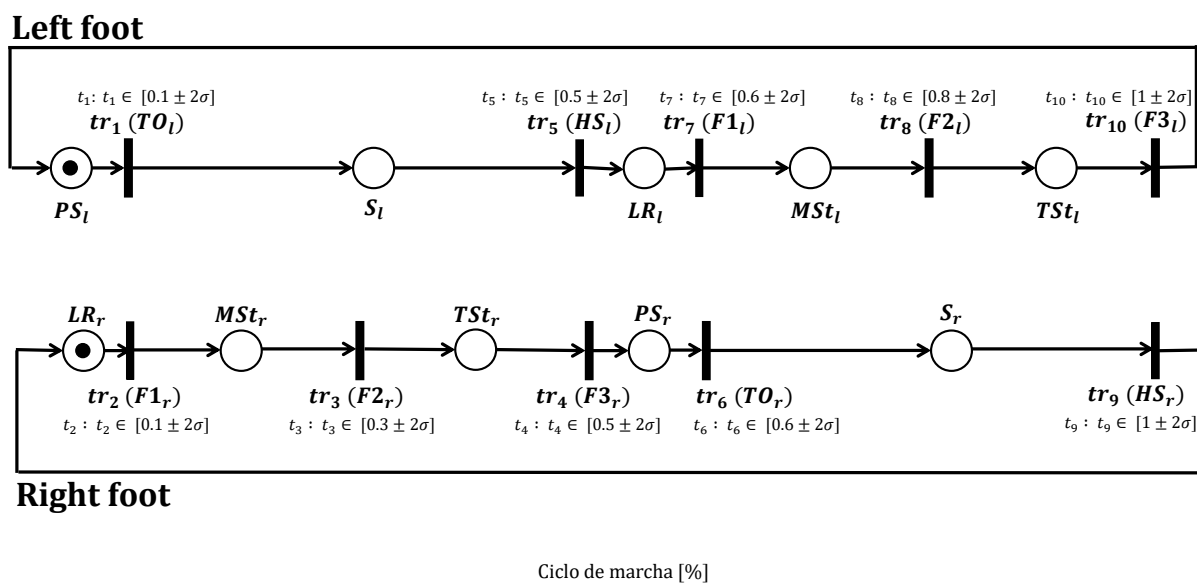
Figure 2.7: Gait st-IPN generated from the phases segmentation seen in the chapter.

assigned a category and represent graphically the convergence of classification of input data. Thus, the set of observations are classified by navigating from the tree's root to a leaf, according to the conditions met along the path.

For instance, Figure 2.8 presents the classification tree that asses demographics attributes of a customer in order to decide their potential response to a direct mail. The root node is the *Age* attribute, the test nodes are *Gender* and *Last R* (Response) attributes. The leafs have the values "Yes" and "No". Notice that the classification tree involves both type of conditions in the edges: nominal ("Yes") and numerical (≤ 30).

2.5.2   Random Forests

The random forests is an ensemble method for classification since constructs a set of classification trees in the training phase. The classification output is the class selected by most trees [73].

The random forests can be applied to a wide range of prediction problems and the few parameters for tuning is advantageous. In addition, its simplicity and accuracy, even with small sample sizes, have increased its popularity [74].

In this work, the R package `randomForest` was employed for the processing of gait timed features. This package is available in `http://www.r-project.org`.

T	Places	
	Pre	**Post**
tr_1	PS_l	S_l
tr_2	LR_r	MSt_r
tr_3	MSt_r	TSt_r
tr_4	TSt_r	PS_r
tr_5	S_l	LR_l
tr_6	PS_r	S_r
tr_7	LR_l	MSt_l
tr_8	MSt_l	TSt_r
tr_9	S_r	LR_r
tr_10	TSt_l	PS_l

Table 2.2: Pre and Post places of each transition seen in Figure 2.7.

2.6   Chapter Summary

In this chapter, we presented the theoretical conceptualisation necessary for the development of the research. First, the concepts of human gait, stride, timing of the gait cycle, and phases of it were addressed. Second, the measurement system based on VGRF and the events that occur in the force profiles that indicate phase changes were described. Third, the concepts of Systems, State, Discrete Events and Petri Nets were exposed. Then, the Petri net proposal that defined to model the behaviour of the gait cycle for both feet. Finally, the supervised learning techniques: Classification Trees and Random Forests were addressed.
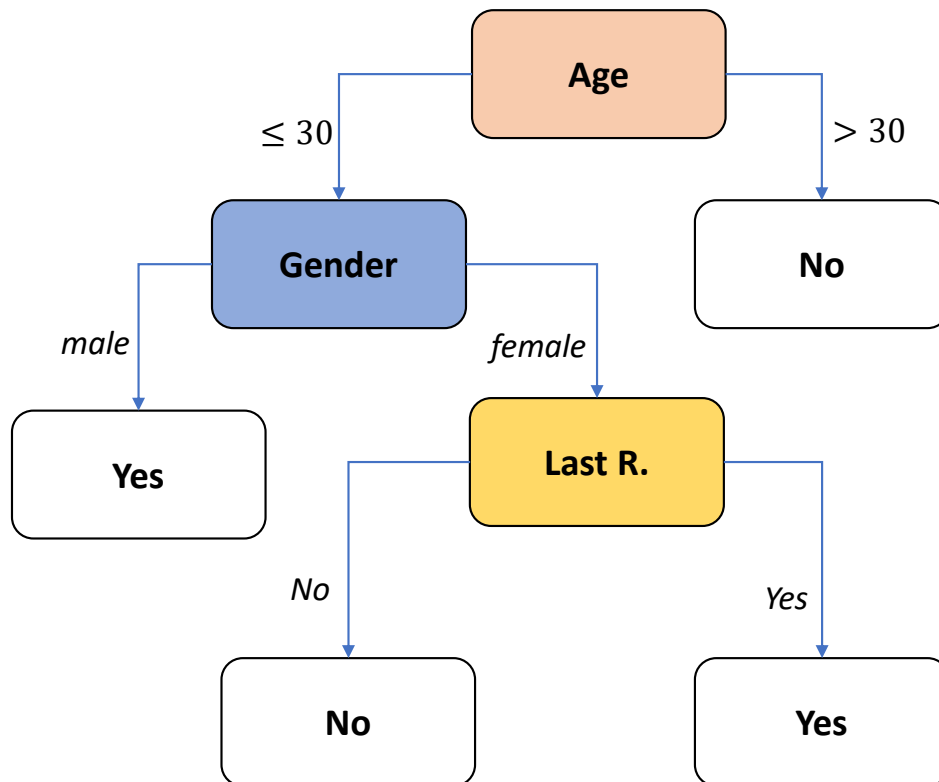
Figure 2.8: Classification tree of response to direct mailing. From [4]

Chapter 3

Method

3.1 Database Presentation

The database employed in this research is freely available at `https://physionet.org/content/gaitndd/1.0.0/`. This database storages the VGRF of 64 participants: 16 belonging to the healthy subjects (CG), 15 diagnosed with PD, 20 affected with HD and 13 with ALS. The data were acquired while the users walked at their preferred speed in a hallway of 77 meters length during five minutes. The VGRF were measured through a pair of insoles with resistive force sensors, whose digital output is a voltage approximately proportional to the contact force between the foot and the floor. A wearable device located in the ankle performed the analog to digital conversion (12 bits) at a sampling rate of 300 Hz [75]. Table 3.1 presents the mean and standard deviation for the height, weight, and age of the participants of each group. The data set for each participant comprises two columns with the voltages (mV) for the left and right foot.

3.2 Data Preprocessing

The data were verified to detect discontinuities due to instrument deficiencies or measurement saturations. As a result, the participant 13 of the control group, the participants 10 and 14 of the Parkinson group, and the participant 13 of the Huntington group were removed (see Figure 3.1).

	Control	PD	ALS	HD
Height [m]	1.833 ± 0.081	1.870 ± 0.152	1.745 ± 0.095	1.830 ± 0.106
Weight [kg]	66.813 ± 11.077	75.067 ± 16.897	77.115 ± 21.150	72.050 ± 17.046
Age [Yrs.]	39.313 ± 18.514	66.800 ± 10.851	55.615 ± 12.829	46.650 ± 12.596

Table 3.1: Demographic information of the Control, PD, ALS, and HD groups.
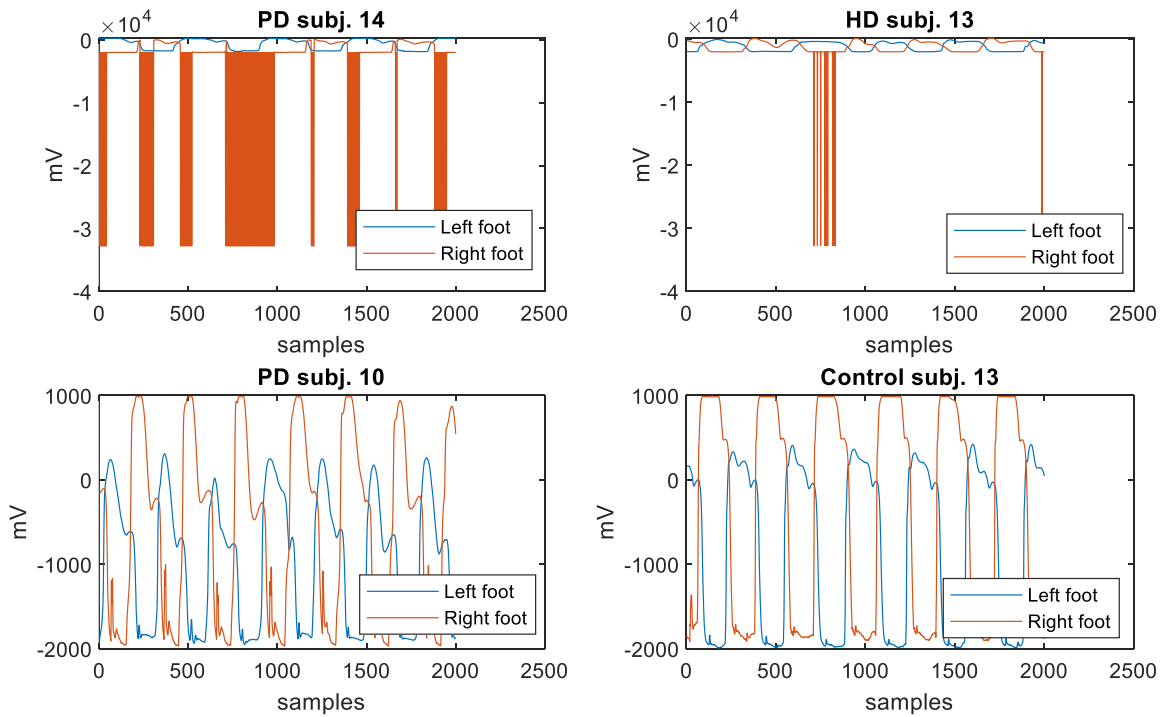
Figure 3.1: VGRF signal saturations and discontinuities of the subjects removed from the present study.

3.2.1   Human Gait Cycle Timing

The duration of each stride was normalized to 0%-100% (101 points) to compare the occurrence of events on same time scale. The 0% of each stride corresponds to the heel strike and it is detected by finding the peaks of the time derivative of the VGRF (Figure 3.2). To avoid false heel strikes an amplitude restriction was imposed, so only peaks greater than the 40% of the maximum of the force derivative of each participant were considered as heel strikes. The left VGRF is timed according to the heel strike of the right foot. Table 3.2 presents the number of strides per participant in each group.

Table 3.2: Computed number of strides per participant in the analysis groups. The lowest number of strides (participant 3) and the highest number of strides (participant 12) are both observed in the ALS group.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
CG	272	236	246	244	194	245	254	235	250	262	227	265	255	256	269
PD	213	238	245	232	238	268	222	218	256	262	215	198	216	-	-
ALS	240	224	111	177	187	236	207	123	194	171	125	326	207	-	-
HD	215	232	243	246	210	301	221	226	259	257	254	226	250	264	-
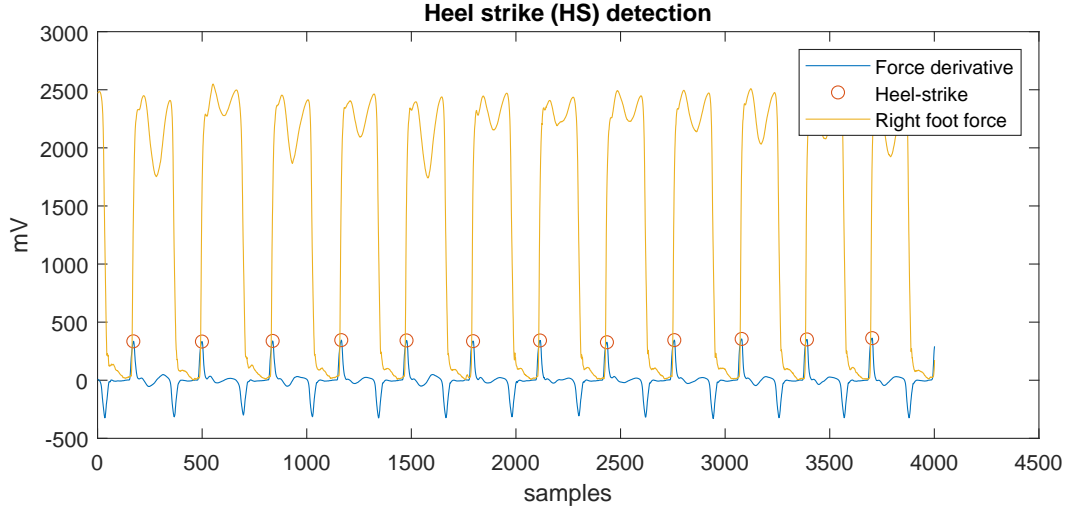
Figure 3.2: Start of each stride according to the total force data (red) and its derivative (blue).

3.2.2 Timed Events Generation

As seen in Figure 3.3, the gait cycle of one foot is delayed 50% with respect to the other; therefore, the events detected for each foot remain the same in number although not in order. The first event detected is HS, that occurs as the VGRF derivative reaches its maximum value. The second event is $F1$, that takes place when the VGRF reaches its first maximum value. The third event is $F2$, which occurs as soon as the VGRF reaches a local minimum value. The fourth event is $F3$, that takes place when the VGRF reaches its second maximum value. Finally, the fifth event is TO that occurs when the time derivative of the VGRF reaches its minimum value.

3.3 Human gait model using a st-IPN

The st-IPN modelling algorithm 1 was adapted to model the human gait cycle [76]. The input to the algorithm is the sequence of events for a given foot, and the output is a Petri net. With this algorithm, we obtained one st-IPN for the right foot and another for the left foot. Figure 3.6 presents the st-IPN of the first participant of the control group. The circles represent the phases of the human gait, the rectangles the transitions between these phases, and the arrows define the sequence of phases. A black dot in a circle indicates that it is the current phase of the gait cycle.

3.3.1 Net Trimming

The nominal sequence of transitions for the right foot is HS, $F1$, $F2$, $F3$, and TO, and for the left foot is TO, HS, $F1$, $F2$, $F3$. However, for the participants of the groups

---

Algorithm 1 st-IPN construction

Input: $(HS, t_1), (F1, t_2), (F2, t_3), (F3, t_4), (TO, t_5)$
Output: Pre, $Post$ and δ.
  Initial conditions: number of places $= 0$, number of $tr=0$;
  {An event occurs}
  {A place is created}
  {A gait phase label is assigned to the place}
  if the place with the assigned label does not exist **then**
    {A transition is created}
    {The matrices Pre y $Post$ for the new tr are created. In addition, the event occurring time
    is registered in δ}
  else
    {None additional places are created}
    {A new transition is created}
    for the first transition **to** the existing number of transitions **do**
      if the created transition already exists connecting two already created places **then**
        true{The transition already exists and connects two existing places. The matrices Pre
        y $Post$ remain the same and the event occurring time is registered in δ} **and break**
        the for loop
        false{This is actually a new transition. The matrices Pre y $Post$ for the new tr are
        created and the event occurring time is registered in δ} ;
      end if
    end for
  end if
  {Wait for a new event}

Transition times for the strides 1 to n						Mean firing times
HS_r	0%	0%	$\cdots$	$\cdots$	t_n	$\overline{HS_r}$
TO_l	t_2	t_{11}	$\cdots$	$\cdots$	t_n	$\overline{TO_l}$
$F1_r$	t_3	t_{12}	$\cdots$	$\cdots$	t_n	$\overline{F1_r}$
$F2_r$	t_4	t_{13}	$\cdots$	$\cdots$	t_n	$\overline{F2_r}$
$F3_r$	t_5	t_{14}	$\cdots$	$\cdots$	t_n	$\overline{F3_r}$
HS_l	t_6	t_{15}	$\cdots$	$\cdots$	t_n	$\overline{HS_l}$
TO_r	t_7	t_{16}	$\cdots$	$\cdots$	t_n	$\overline{TO_r}$
$F1_l$	t_8	t_{17}	$\cdots$	$\cdots$	t_n	$\overline{F1_l}$
$F2_l$	t_9	t_{18}	$\cdots$	$\cdots$	t_n	$\overline{F2_l}$
$F3_l$	t_{10}	t_{19}	$\cdots$	$\cdots$	t_n	$\overline{F3_l}$

Table 3.3: Mean firing times arrangement for a both feet in the $i-th$ subject

PD, HD and ALS these sequences are not always obtained. In such cases, the strides with non-nominal sequences of events are eliminated. For example, Figure 3.5 shows an example where only HS and TO transitions are detected, while $F1$, $F2$, and $F3$ are not. In such a case, the modelling algorithm adds a direct transition from LR_l to S_l (left side of Figure 3.6). Since this transition only occurs once, the trimming algorithm considers it spurious. In general, these non nominal transitions are removed from the st-IPN (right side of Figure 3.6).

3.4 Timed arrangements for group comparison

The st-IPN of each participant comprises nine numbers, which are the average times of occurrence of the transitions $F1_r$, $F2_r$, $F3_r$, TO_r, HS_l, $F1_l$, $F2_l$, $F3_l$ and TO_l. Since HS_r is always 0%, it is not considered for statistical analyses.

3.5 Discrete Events Sequence Analysis

The transitions were chronologically ordered. Consequently, the gait of each participant is represented by sequences of nine transitions, one sequence by stride. Table 3.4 shows the most common sequences for the control group. The integers in the columns two to ten indicate the order in which the transitions occur. The last column contains the percentage of strides whose sequence is the indicated by the preceding columns.

Table 3.5 shows the sequences of transitions that took place for the 15 participants of the control group (s_1 to s_{15}). For example, for the participant s_1, 254 sequences correspond to the first row of Table 3.4 and 14 to the second row of the same table. Among all participants of the control group, the most common sequence is the described by the first row of Table 3.4. Since this sequence occurred 2.173 times, we had 2.173 samples that

Seq. Id	TO_l	$F1_r$	$F2_r$	$F3_r$	HS_l	TO_r	$F1_l$	$F2_l$	$F3_l$	% Occur.
1	1	2	3	4	5	6	7	8	9	82.66
2	1	2	3	**5**	**4**	6	7	8	9	7.76
3	1	2	3	4	5	**7**	**6**	8	9	2.32
4	1	2	3	4	5	6	7	**9**	**8**	1.56
5	1	2	3	**5**	**4**	6	7	**9**	**8**	0.99
Other										4.72

Table 3.4: Sequences of events of the control group.

Seq. Id.	Sequences participation among control group subjects															% ocurr.
	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}	s_{13}	s_{14}	s_{15}	
1	254	202	150	192	145	112	10	217	22	197	17	218	254	1	182	82.66
2	14	8	1	3	31	75	1	1	0	41	6	6	0	17	0	7.76
3	0	0	36	1	4	0	0	0	0	0	19	0	0	0	1	2.32
4	0	5	6	7	2	4	0	2	8	0	0	3	0	4	0	1.56
5	0	0	0	0	1	1	0	0	2	0	0	0	0	22	0	0.99

Table 3.5: Participation of the sequences in the subjects of the Control group.

were used to establish the 95% confidence interval of each transition (Table 3.6).

For each participant in the ALS, HD, and PD groups, we found the most common sequence of transitions, which we call the dominant sequence. Subsequently, we eliminated the strides whose sequence was different from the dominant and plotted the time of occurrence of each transition as a function of the stride (Figure 3.7). For each of the nine graphs, we calculate the percentage of values within the 95% confidence interval.

Event	Lower (%)	Upper (%)
TO_l	6	11
$F1_r$	9	19
$F2_r$	18	44
$F3_r$	32	51
HS_l	48	53
TO_r	54	61
$F1_l$	59	74
$F2_l$	78	96
$F3_l$	87	100

Table 3.6: 95% confidence intervals for the transitions of the control group.

No	Feature	Unit
p_1	Left Stride Interval	Seconds
p_2	Right Stride Interval	Seconds
p_3	Left Swing Interval	Seconds
p_4	Right Swing Interval	Seconds
p_5	Left Swing Interval	% of stride
p_6	Right Swing Interval	% of stride
p_7	Left Stance Interval	Seconds
p_8	Right Stance Interval	Seconds
p_9	Left Stance Interval	% of stride
p_{10}	Right Stance Interval	% of stride
p_{11}	Double Support Interval	Seconds
p_{12}	Double Support Interval	% of stride

Table 3.7: Features obtained from the data set of PhysioNet.

3.6 Supervised learning approach

This section presents the results obtained with RFs and CTs, which are by default multi-class classifiers based on supervised learning. Both RFs and CTs were trained using two different sets of features. First are the nine means and nine standard deviations of Petri net's transition times (18 features in total), while second are the twelve means and twelve standard deviations of the features presented in Table 3.7 (24 features in total). Most machine-learning approaches presented in the introduction section use these last features, which are available from PhysioNet VGRF database. The data in all the experiments reported in this section were divided randomly between training (70%) and validation (30%).

3.7 Chapter Summary

In this chapter, we presented the method of a neurodegenerative disease diagnoser based on st-IPN via VGRF processing. Firstly, the VGRF database and its preprocessing was presented, which involved the removal of some participants, since their data were unfitting for this study. Secondly, the set of events (IC, F1, F2, F3 and TO) were gathered from every stance period of all subjects. Per individual, these events were fed to the st-IPN modelling algorithm that created a st-IPN for each foot; the net was *trimmed* in case of any inconsistency in terms of events order. Thirdly, a nominal sequence of events from the control group was determined and a 95% interval of time, per event, was calculated. Then, for each pathological participant the percentage of inclusion within the mentioned interval was calculated. Finally, the supervised learning approach is addressed.
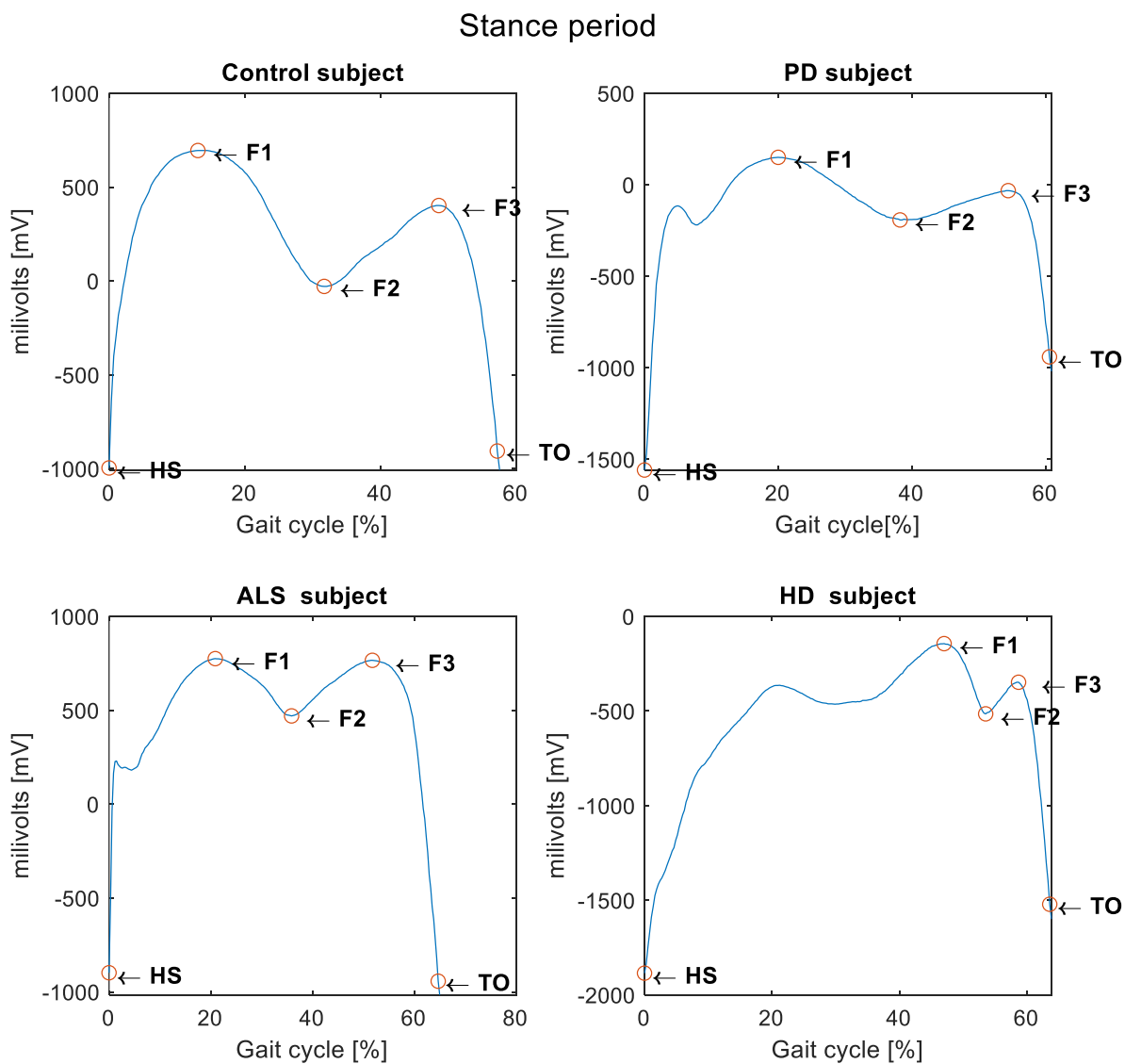
Figure 3.3: Events detection across the GVRF of a healthy and pathological subjects.
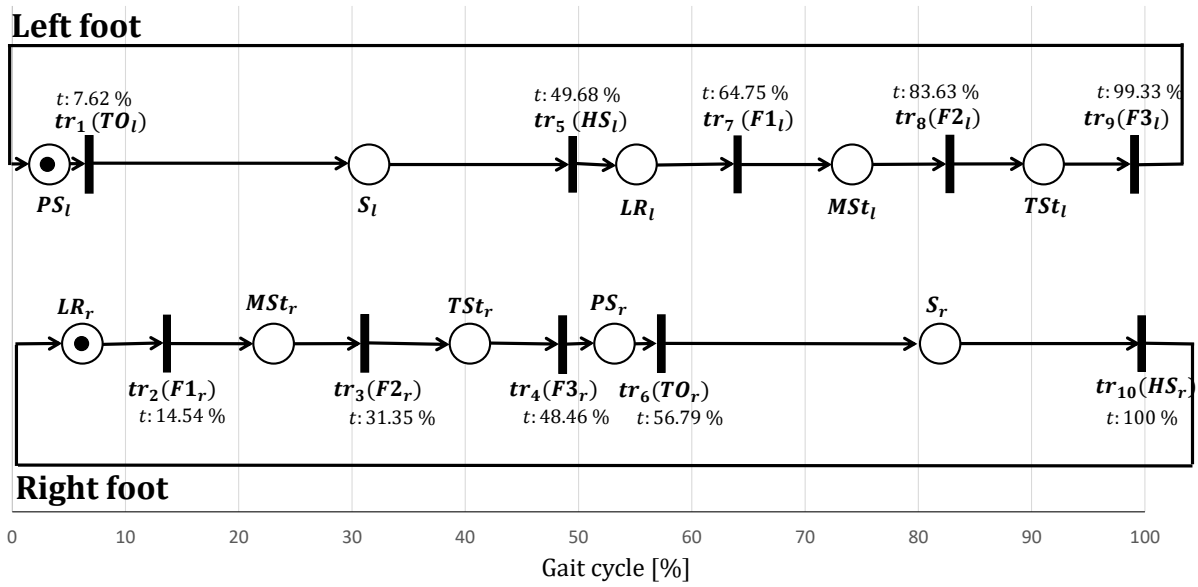
Figure 3.4: Gait st-IPN generated for the Control subject 1.
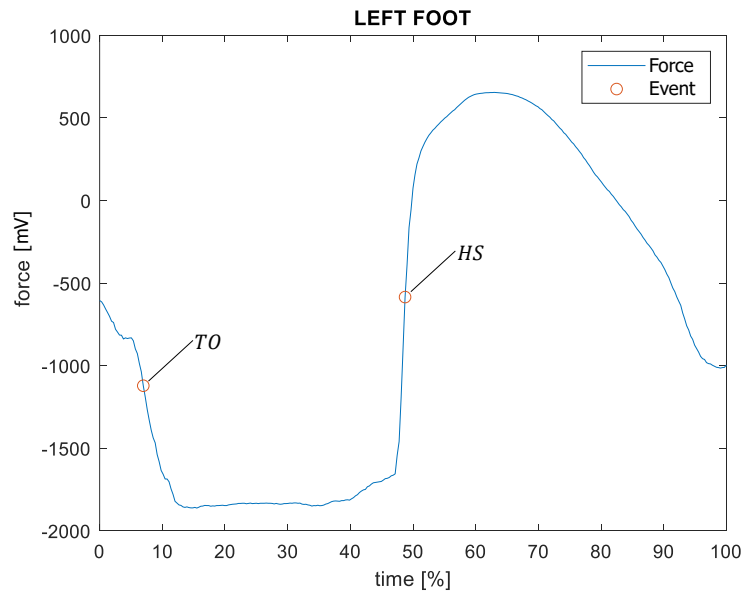
Figure 3.5: Non nominal events generated
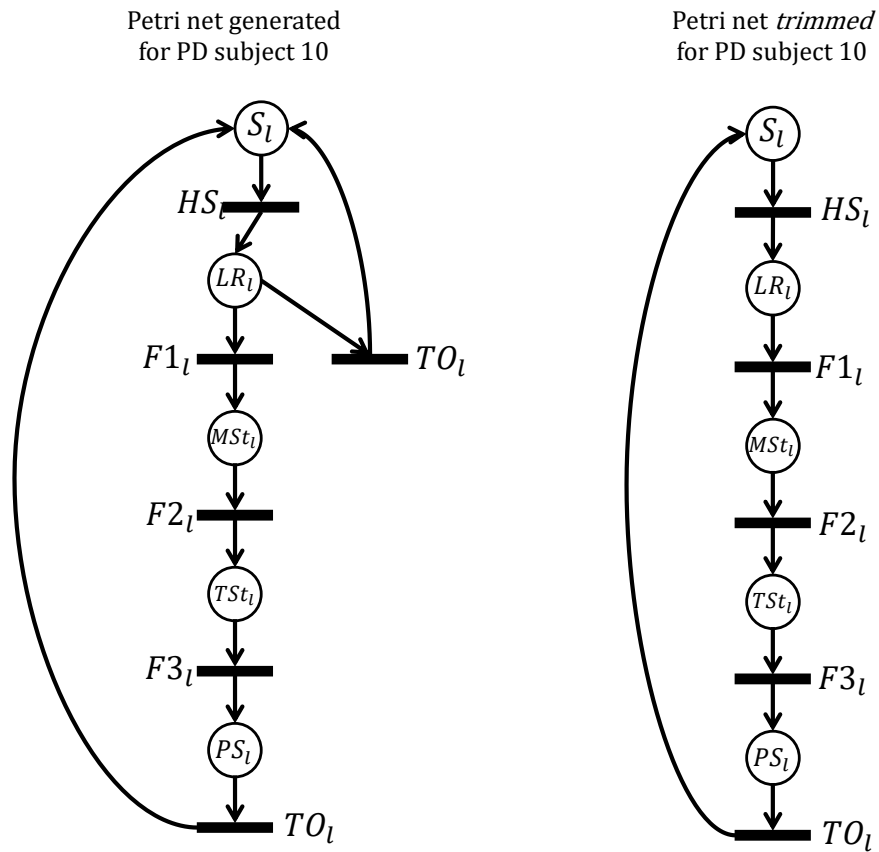
Figure 3.6: Comparison between the Petri net generated and the *trimmed* Petri net from the PD subject 10 left foot.
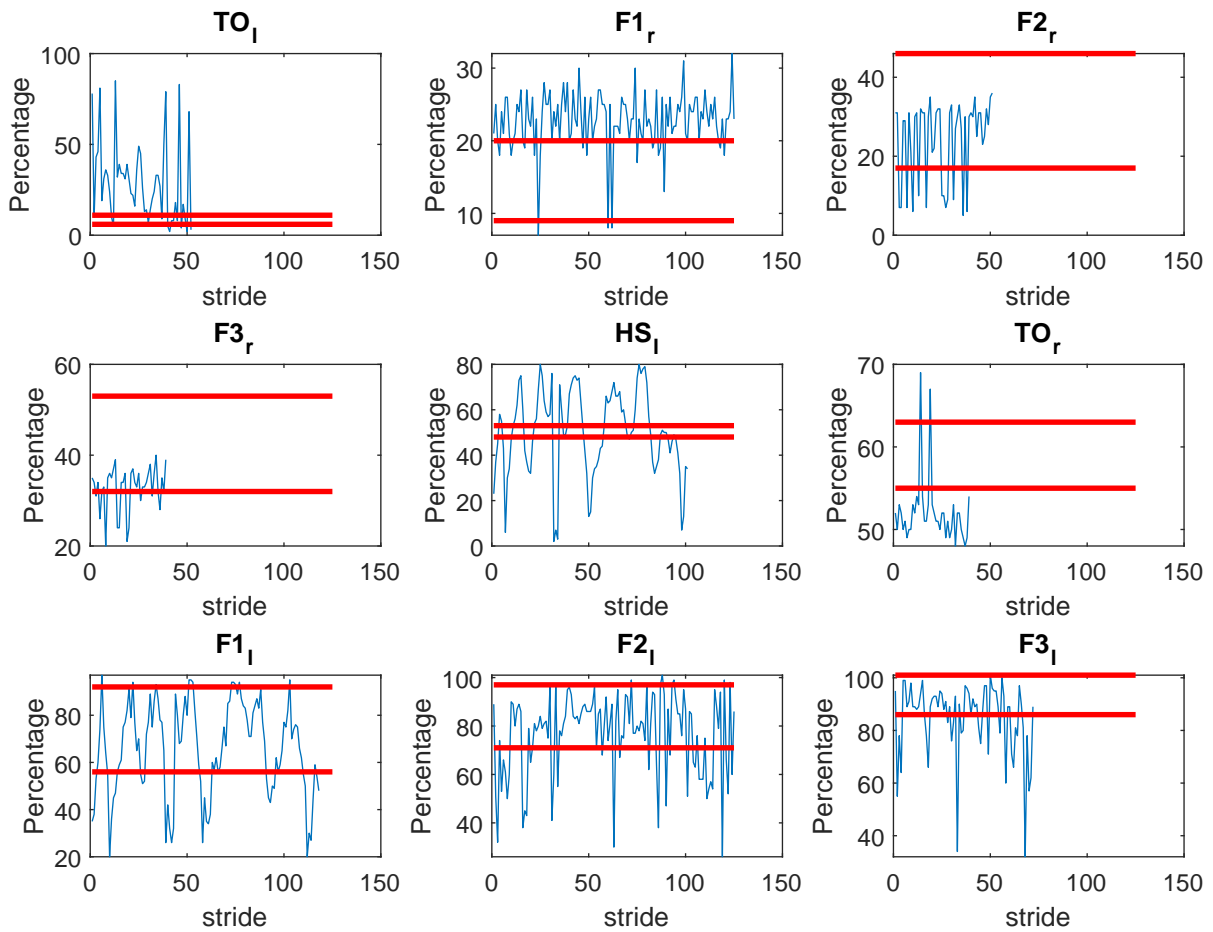
Figure 3.7: Comparison of the occurrence times of each event with the upper and lower bounds generated for subject 11 of ALS.

Chapter 4

Results

This chapter presents the following results i) the events in each subject that present a reduced percentage of values within the 95% interval ii) per group, the multiple comparison test is performed among all the set of gait events in order to verify which is the less contained within bounds for each group, iii) the CT and RF classification performance with the set of features obtained from both the Petri net model and PhysioNet's database features.

4.1 Sequence analysis

Tables 4.1, 4.3, 4.2, and 4.4 present the percentage of transitions that belong to the confidence intervals given by Table 3.6. For the participant three of the Control group (Table 4.1), the transition with the lowest percentage of values within the confidence interval is $F2_l$ (45.21%). For the Parkinson group, this value is 0% and correspond to the transition TO_l of the subject four. For the participant number one of the ALS group, transition TO_l is always outside the confidence interval. For the first participant of the Huntington group, the transition TO_r belongs to its corresponding confidence interval with the lowest value: 2.92%.

Figure 4.1 shows that for the ALS group, the percentage of TO_l within its confidence interval is significantly lower than for the transition $F1_l$. Figure 4.2 shows that for the Huntington group, the percentage of TO_l within its confidence interval is only significantly lower than HS_l transition. Figure 4.3 shows that for the Parkinson group, the percentage of TO_l belonging to its respective confidence intervals is significantly lower than HS_l, $F1_r$ and $F3_r$. In summary, for each group TO_l is the transition with a significantly lower percentage of values within its confidence interval.

Group	Subject	TO_l	$F1_r$	$F2_r$	$F3_r$	IC_l	TO_r	$F1_l$	$F2_l$	$F3_l$
CG	1	99.63	99.63	98.52	99.26	98.16	99.63	100.00	99.63	100.00
CG	2	98.61	93.64	99.15	97.88	97.88	98.31	99.58	100.00	100.00
CG	3	99.50	99.59	99.59	80.41	97.56	99.18	91.06	99.52	89.66
CG	4	98.20	100.00	99.57	90.75	97.95	98.68	100.00	96.22	97.31
CG	5	98.95	92.27	99.48	98.94	96.91	98.94	96.91	98.45	98.96
CG	6	98.97	98.37	98.77	95.08	98.78	96.72	100.00	99.18	97.44
CG	7	72.73	98.03	88.89	90.08	94.09	89.29	100.00	99.60	90.91
CG	8	97.40	81.28	93.59	98.72	94.04	77.78	99.57	99.57	98.71
CG	9	100.00	99.60	99.20	98.80	98.80	99.20	100.00	98.39	96.88
CG	10	99.58	99.24	100.00	100.00	98.47	99.62	100.00	100.00	100.00
CG	11	96.47	67.40	84.93	97.86	94.71	96.26	55.95	45.21	98.82
CG	12	97.80	98.11	99.62	100.00	100.00	100.00	100.00	100.00	99.12
CG	13	92.13	100.00	100.00	98.82	99.22	98.43	100.00	100.00	98.82
CG	14	95.45	99.22	92.65	93.06	99.61	97.14	100.00	73.47	86.36
CG	15	99.25	95.91	91.03	99.53	99.63	99.53	100.00	99.63	99.63

Table 4.1: Percentage of transitions of the CG group that belong to the confidence intervals defined by the control group.

Group	Subject	TO_l	$F1_r$	$F2_r$	$F3_r$	IC_l	TO_r	$F1_l$	$F2_l$	$F3_l$
ALS	1	0.00	67.08	78.21	99.45	72.50	38.12	100.00	99.17	97.92
ALS	2	30.00	92.41	97.30	16.22	95.98	88.74	100.00	97.32	96.67
ALS	3	100.00	92.79	92.79	89.19	24.32	82.88	97.30	32.99	91.67
ALS	4	1.75	71.19	38.64	31.85	82.49	31.85	100.00	79.45	72.17
ALS	5	81.82	4.28	95.70	95.14	44.92	85.41	22.46	95.70	66.67
ALS	6	100.00	92.80	96.96	98.69	99.58	98.69	100.00	99.15	97.02
ALS	7	81.33	97.10	99.51	97.07	18.84	51.22	77.78	83.17	85.33
ALS	8	1.75	15.45	82.09	56.92	54.47	16.92	98.37	63.10	28.07
ALS	9	40.79	89.69	90.91	75.00	13.68	25.00	82.47	72.32	80.38
ALS	10	2.50	6.43	91.18	93.53	70.76	71.76	89.82	90.12	96.25
ALS	11	15.38	21.60	74.51	74.36	14.85	0.00	61.02	67.20	65.28
ALS	12	31.65	88.04	54.50	35.00	22.57	53.64	47.86	68.10	60.71
ALS	13	90.00	52.17	97.56	97.56	97.58	98.54	99.52	99.52	70.00

Table 4.2: Percentage of transitions of the ALS group that belong to the confidence intervals defined by the control group.

Group	Subject	TO_l	$F1_r$	$F2_r$	$F3_r$	IC_l	TO_r	$F1_l$	$F2_l$	$F3_l$
PD	1	18.42	48.36	28.70	79.41	49.30	29.41	84.06	30.40	100.00
PD	2	80.00	95.80	97.84	97.38	95.80	96.94	98.74	93.28	100.00
PD	3	86.00	93.06	100.00	72.60	89.39	94.06	98.78	73.04	88.00
PD	4	0.00	67.67	85.90	78.67	95.69	52.00	99.13	92.58	100.00
PD	5	74.64	92.86	85.17	86.02	86.55	92.37	99.16	86.73	87.77
PD	6	99.21	98.88	99.25	98.88	98.88	98.88	99.25	100.00	100.00
PD	7	12.50	4.95	99.10	83.64	95.50	53.18	99.10	21.81	50.00
PD	8	100.00	81.65	84.74	64.71	81.65	95.19	99.54	58.85	100.00
PD	9	5.56	97.27	99.21	76.77	89.06	46.06	100.00	98.43	83.33
PD	10	94.59	95.80	32.00	19.78	93.51	98.90	99.62	98.85	98.65
PD	11	38.10	97.67	52.09	30.70	80.84	88.37	99.53	98.60	90.91
PD	12	28.57	75.25	88.14	75.26	44.44	48.95	98.48	50.00	61.90
PD	13	98.57	76.85	99.53	98.60	97.22	98.60	77.78	98.15	99.53

Table 4.3: Percentage of transitions of the PD group that belong to the confidence intervals defined by the control group.

Group	Subject	TO_l	$F1_r$	$F2_r$	$F3_r$	IC_l	TO_r	$F1_l$	$F2_l$	$F3_l$
HD	1	94.22	85.58	97.13	91.23	22.79	2.92	98.60	87.02	87.36
HD	2	74.67	97.84	86.84	96.33	90.95	93.58	99.14	97.84	93.38
HD	3	62.10	62.14	87.50	88.74	64.20	84.77	98.35	92.31	79.09
HD	4	88.24	56.91	99.59	97.15	95.12	98.37	98.78	90.04	91.89
HD	5	40.98	66.19	59.26	54.55	26.67	12.73	66.14	55.37	87.10
HD	6	5.43	15.95	91.84	98.59	70.43	91.55	63.46	88.04	97.68
HD	7	73.40	99.10	99.53	87.38	83.71	96.26	97.74	98.18	100.00
HD	8	84.71	92.04	95.07	81.08	66.37	79.28	97.35	88.58	62.57
HD	9	85.17	63.71	81.82	92.42	67.57	79.15	93.82	97.28	97.47
HD	10	94.93	98.83	98.32	97.46	52.53	96.61	98.83	97.67	98.17
HD	11	93.42	82.68	98.02	94.42	79.13	93.23	98.43	97.61	99.12
HD	12	12.86	60.62	43.20	21.46	23.01	48.78	100.00	79.10	92.86
HD	13	98.74	95.20	100.00	99.20	97.60	99.20	99.60	99.17	89.12
HD	14	98.25	96.59	100.00	98.48	95.83	97.73	100.00	99.24	100.00

Table 4.4: Percentage of transitions of the HD group that belong to the confidence intervals defined by the control group.
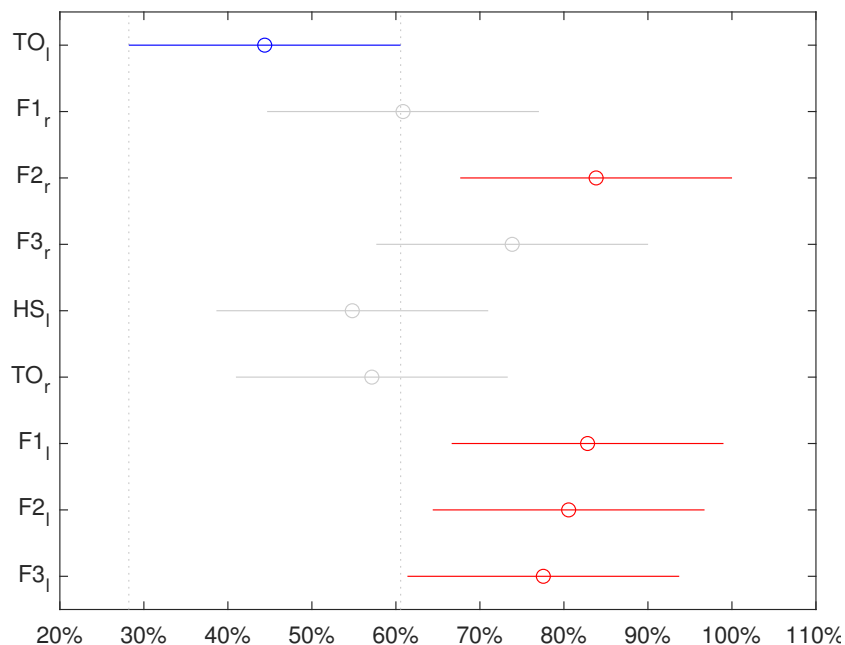
Figure 4.1: ALS group. Percentage of the transitions belonging to the confidence intervals defined by the control group.

Class	Petri net features			PhysioNet features		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
ALS	91.7	83.3	100	95.2	92.3	98.0
Control	94.7	100	89.5	89.4	93.8	85.1
Huntington	95.1	92.9	97.4	85.7	73.7	97.7
Parkinson	91.7	83.3	100	89.0	80.0	97.9

Table 4.5: The percentage of accuracy, sensitivity, and specificity of the RFs according to the training features.

4.2 Unsupervised Learning results

4.2.1 Random forest

The two RFs presented in this section were obtained from the function `train` of the `caret` package of the R programming language [77]. The sensitivities and the specificities are provided in Table 4.5. The overall accuracy of the model trained with the transition times of the Petri net was 90.6%, while the model trained with the PhysioNet features had an accuracy of 84.1%. One of the sensitivities and two of specificities of the proposed features reached as high as 100%. In addition, the lower indicator for the Petri net features was 83.3%, while for the other features this value was 73.7%. Figure 4.4 compares the two RFs using a radar chart.
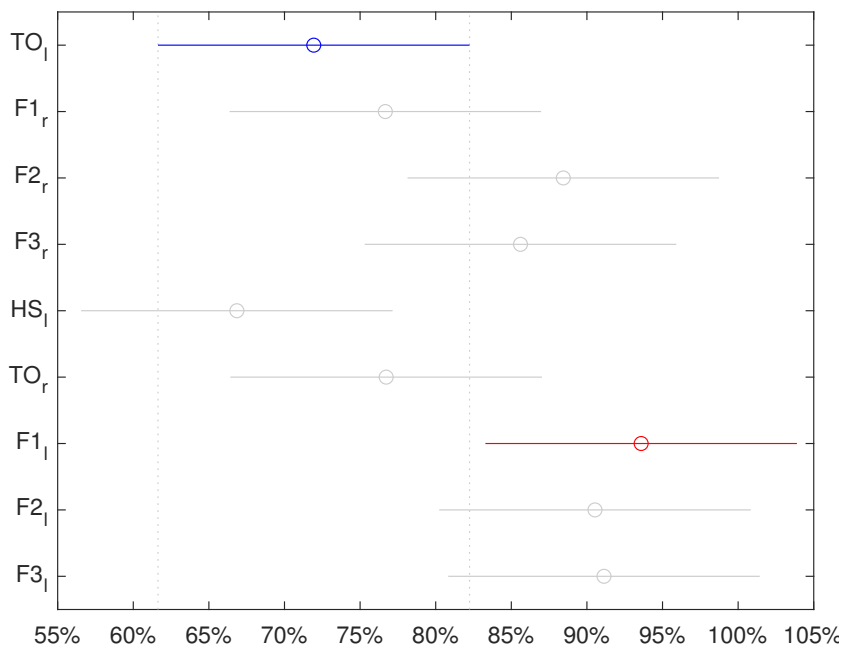
Figure 4.2: Huntington group. Percentage of the transitions belonging to the confidence intervals defined by the control group.

4.2.2 Classification trees

Since our aim is to provide a screening tool to support differential diagnosis rather than a black box for pattern classification, we used CTs, which enable easy clinical interpretation. The CTs were generated by the function `rpart` of the R language's `rpart` library.

Features defined by the Petri model

Figure 4.5 presents the CT obtained from the transition times of the Petri net. Of the 18 features used to train the model, the CT uses only the following four: (1) the standard deviation of the HS for the left foot, (2) the average of the TO for the right foot, (3) the average of the TO for the left foot, and (4) the average for the time at which the highest local minimum of the ground reaction occurs ($F2$).

PhysioNet's Features

Figure 4.6 presents the CT obtained from the features of Table 3.7. It uses four of the 24 features used to train the model: (1) the standard deviation of the predictor p_7 (left stance interval in seconds), (2) the average of the predictor p_2 (right stride interval in seconds), (3) the standard deviation of the predictor p_8 (right stance interval in seconds), and (4) the standard deviation of the predictor p_{12} (double support interval as a percentage of the stride duration).
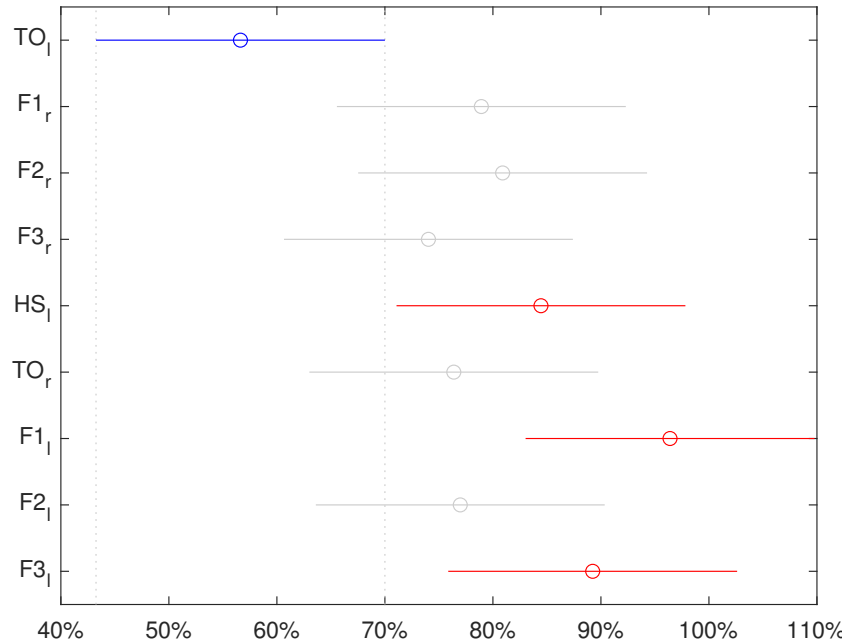
Figure 4.3: Parkinson group. Percentage of the transitions belonging to the confidence intervals defined by the control group.

Class	Petri net features			PhysioNet features		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
ALS	69.6	41.7	97.6	66.2	38.5	94.0
Control	79.4	66.7	92.1	79.0	75.0	83.0
Huntington	77.3	57.1	97.4	66.1	36.8	95.5
Parkinson	73.4	83.3	63.4	71.0	73.3	68.8

Table 4.6: The percentage of accuracy, sensitivity and specificity of the CT according to the training features.

Comparison of the classification trees

The Table 4.6 presents the accuracy, sensitivity and specificity of each CT. The overall accuracy of the model trained with the features obtained from the Petri net was 62.3%, while the accuracy of the model trained with the PhysioNet features was 55.6%. According to Table 4.6, the first three classes of both models have high specificities and moderate sensitivities, which implies that for ALS, CG and HD, the two CTs tend to produce more false negatives than false positives. However, for PD, the opposite is true; since the sensitivity is greater than the specificity, the model is more likely to generate false positives than false negatives. A radar chart is used to compare the two CTs in Figure 4.7.

Figure 4.4: RF comparison in terms of the sensitivity and specificity of each class. ALS.Se and ALS.Sp are the sensitivity and the specificity of the ALS class, respectively. An equivalent notation is used for the other three classes.
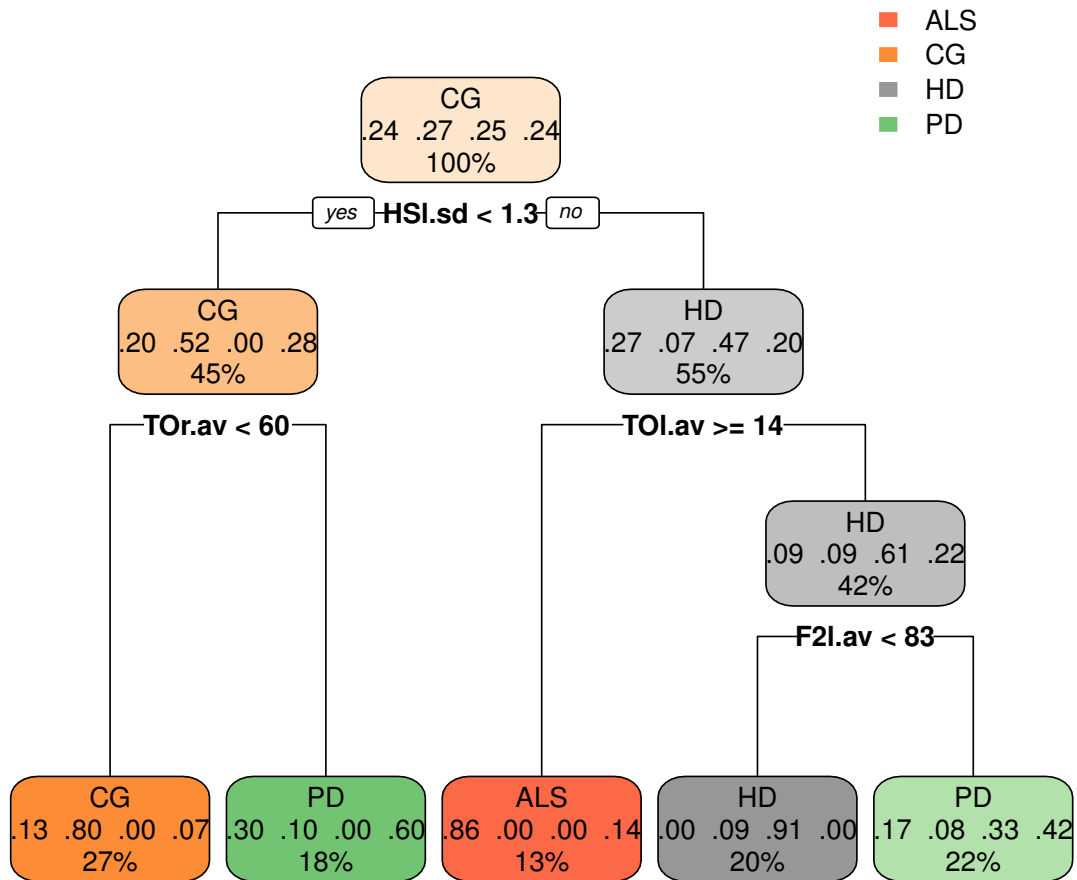
Figure 4.5: CT trained with the nine means and the nine standard deviations of the transitions times of each participant. Each node comprises three lines: the predicted class, the probability for each of the four classes, and the percentage of observations in the node. The position of the highest number in the second row indicates the predicted class (1 for ALS, 2 for Control, 3 for Huntington, and 4 for Parkinson). The average and the standard deviation of the left TO are represented by TOl.av and TOl.sd, respectively. An equivalent notation is used for the other transitions.
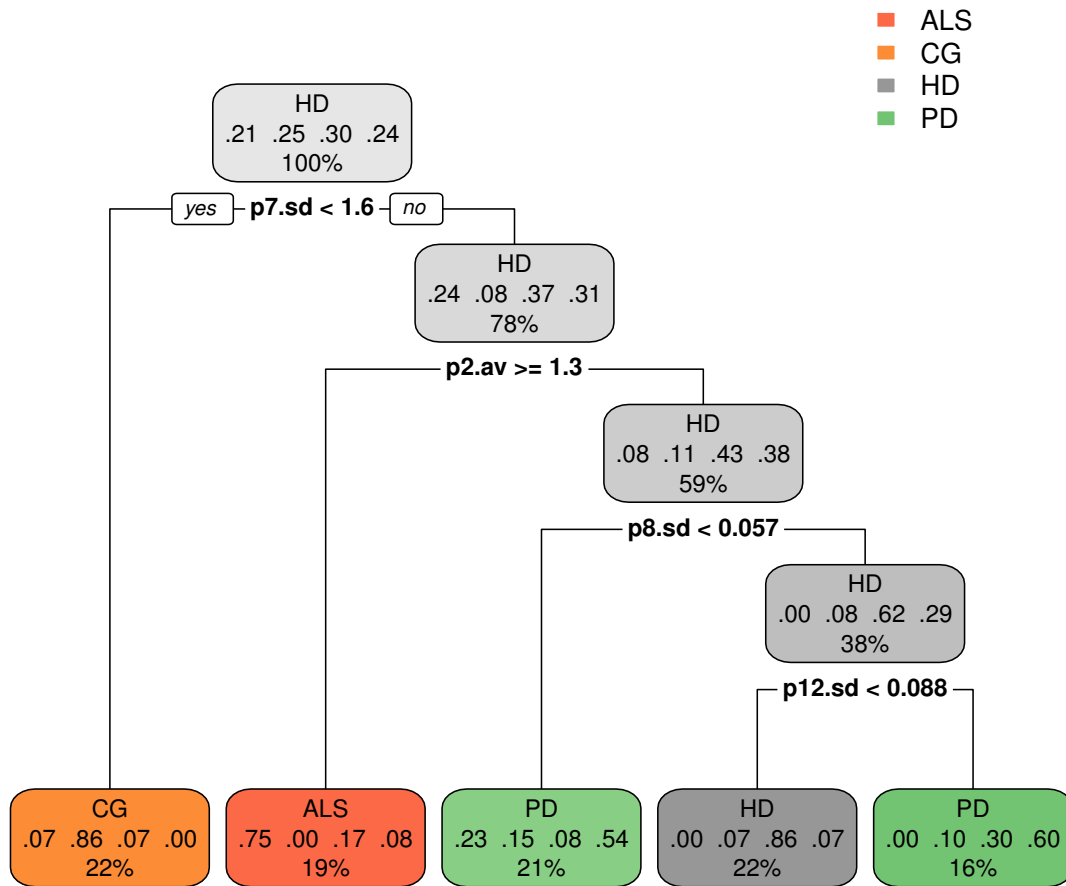
Figure 4.6: CT trained with the twelve means and the twelve standard deviations of the features in Table 3.7. Each node comprises three lines: the predicted class, the probability for each of the four classes, and the percentage of observations in the node. The position of the highest number in the second row indicates the predicted class (1 for ALS, 2 for Control, 3 for Huntington, and 4 for Parkinson). The average and the standard deviation of the predictor p_2 are represented by p2.av and p2.sd, respectively. An equivalent notation is used for the other predictors.
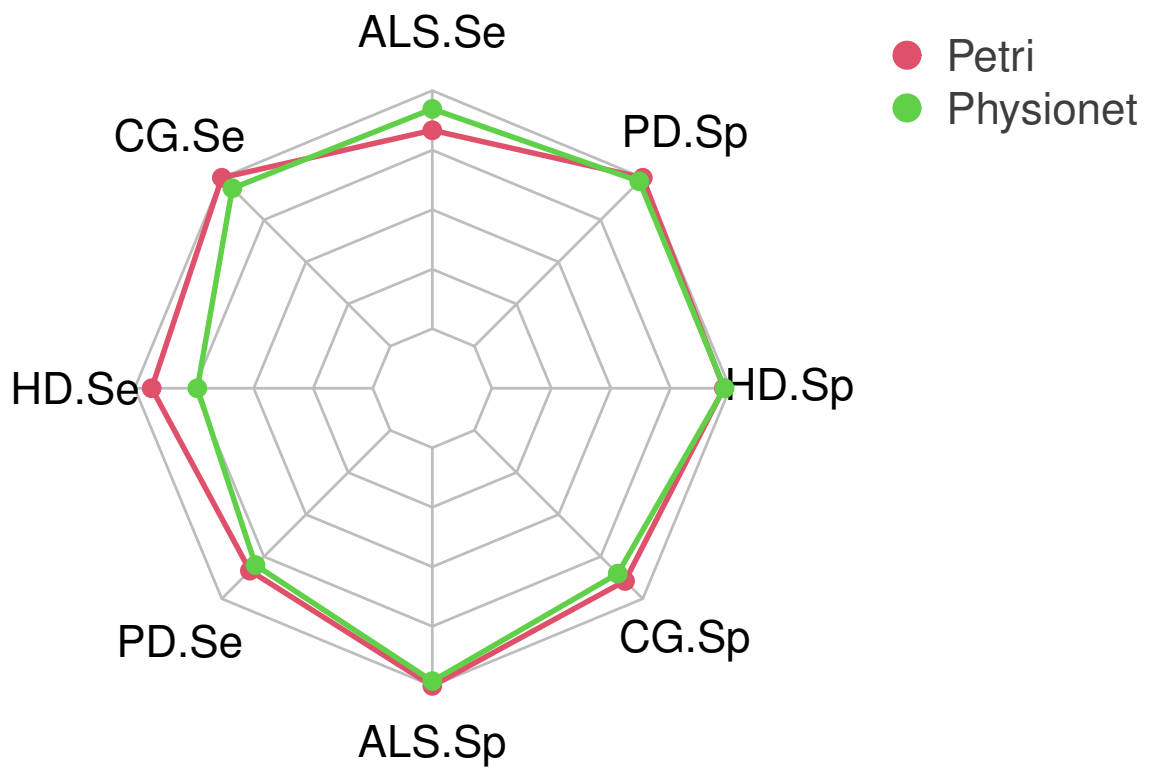
Figure 4.7: CT comparison in terms of the sensitivity and specificity of each class. ALS.Se and ALS.Sp are the sensitivity and the specificity of the class ALS, respectively. An equivalent notation is used for the other three classes.
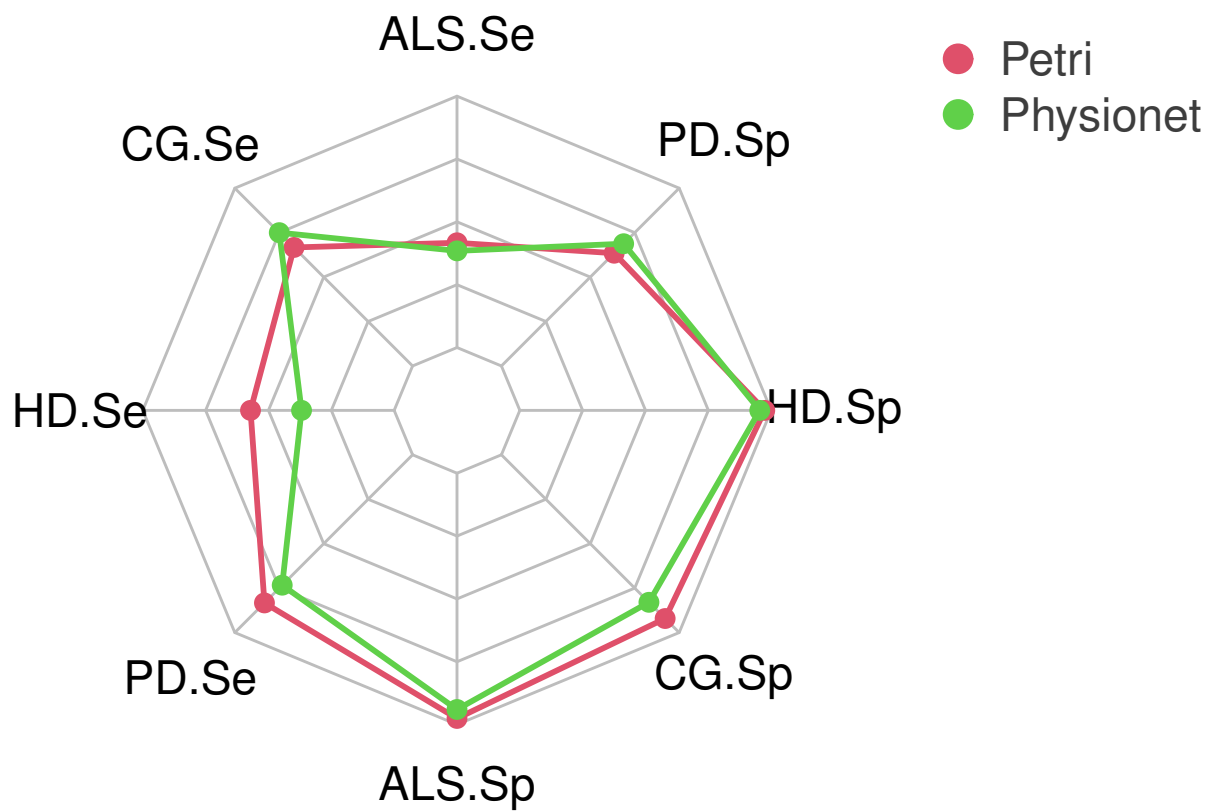
Chapter 5

Discussion

In this dissertation, we proposed to model human gait using timed Petri nets comprising nine phases and nine events. Each event defines the transition from one phase to the next. The moment at which these events occurred were represented as random variables between 0% and 100% of the stride duration. This normalization allowed the transition times of strides of different duration to be considered as samples of the same population. For the PD, HD and ALS groups, we observed that the transition with the lowest percentage of values belonging to the confidence intervals is only TO_l. Therefore, this similarity in the transitions makes not possible a differential diagnosis among neurodegenerative diseases. In addition, the transition times were used as features to train a multilevel CT, which, due to its easy of human interpretation, was devised as a tool to support the differential diagnosis of neurodegenerative diseases.

Previous studies on using VGRFs to diagnose neurodegenerative diseases can be categorized into one of two classes. To the first group belong the studies that use binary classifiers to label gait patterns as healthy or abnormal, such as [36, 42, 37, 38, 35, 39, 40, 41]. The second group comprises the works that present multiple binary comparisons, typically CG versus PD, CG versus HD, and CG versus ALS, for example, [43, 44, 45, 46, 47, 48, 49]. Despite the high accuracy reported in these studies, their main inconvenient is the uncertainty of the classifier output when the evaluated subject does not belong to either of the two categories for which it was trained. Table 5.1 summarizes the studies that present multiple binary classifiers. The approach proposed in [44] was not included because it does not report accuracy, sensitivity, nor specificity. Although we use a four-level classifier (ALS, CG, HD, and PD), instead of four binary classifiers, the RF performance for the Petri net features (Table 4.5) is comparable to that of the Table 5.1. However, this is not the case for the CT, whose performance is significantly lower.

[5] proposed a solution for the multilevel classification problem, which consists of a two-steps process. The participants are first classified as CG or NDD, then NDD gait patterns are sent to three binary classifiers: ALS versus PD, ALS versus HD, and HD versus PD. The final label of someone in the NDD group is defined by the voting algorithm described in [57], which is applied to the output of all three subclassifiers. Table 5.2 shows that

Study	Approach	CG-ALS			CG-HD			CG-PD		
		Acc.	Se.	Sp.	Acc.	Se.	Sp.	Acc.	Se.	Sp.
[43]	SVM	96.8	95.1	98.8	90.3	95.9	85.4	89.3	89.8	89.8
[45]	LS-SVM	100	100	100	100	100	100	100	100	100
[46]	NNLS	98.0	100	95.0	88.0	100	75.0	91.0	100	85.0
[47]	TDA: PL+RF	75.9	81.2	69.2	91.7	87.5	95.0	90.3	87.5	93.3
[48]	NNLS	94.0	87.0	100	95.0	90.0	100	93.0	87.0	100
[49]	ConvLSTM	97.7	97.5	97.7	92.9	92.8	93.0	94.0	94.0	94.1
[51]	ANFIS	93.1	92.3	93.8	94.4	93.8	95.0	90.3	93.8	86.7
[52]	RBF-NN	89.7	92.3	87.5	83.3	85.0	81.2	89.7	92.3	87.5
[53]	RQA + SVM	96.1	100	94.0	100	100	100	100	100	100
[54]	NNLS	100	100	100	99.9	99.8	100	99.8	99.5	100

Table 5.1: Summary of studies with multiple binary classifiers. If more than one machine learning approach was used, only the classifier that led to the highest accuracy was reported. SVM stands for Support Vector Machine, LS stands for Least Squares, NNLS stands for Non-Negative Least Squares, ConvLSTM stands for Convolutional Short-Term Long-Term Memory, ANFIS stands for Adaptive Neurofuzzy Inference System, RBF stands for Radial Basis Function, NN stands for neural network, and RQA for recurrence quantification analysis.

for the Control and Huntington groups, the RF classifier trained with Petri net features (Table 4.5) is 10% more accurate than the multiclass approach described by [5]. For the Parkinson group, this difference is only 1%, and for the ALS group, the classifier of [5] is 2% more accurate than the one proposed here.

Group	Petri net - RF	[5]
ALS	91.7	93.7
Control	94.7	81.5
Huntington	95.1	84.7
Parkinson	91.7	90.7

Table 5.2: Percentage of accuracy obtained using the Petri Net features (Table 4.5) and with multiclass approach proposed by [5].

Despite of the impressive results obtained by machine learning approaches in the last decade, the main obstacle to their widespread adoption in a clinical context is the lack of interpretability of the resulting models. RFs, support vector machines, and neural networks generate black boxes that map a given set of features to a class (i.e., ALS, CG, PD, or HD, in the present paper). However, as well as being extremely complex, the rules behind these maps do not contribute to establishing correlations between diseases and deviations of features from nominal values. Conversely, CTs provide insight into this

correlation. For example, Figure 4.5 reveals that the TO and HS are important for differentiating between ALS, CG and PD. Likewise, the lowest-right node of this tree indicates that the occurrence of the VGRFs local minimum is affected differently in HD and PD.

The present study has four main limitations. First, the average age of the CG was much younger than in the other three groups (Table 3.1). Since gait changes with age this is a potential problem for the external validity of the proposed CT. In addition, a group with an average age of 39.313 ± 18.514 could not be compared with groups with higher ages (55.615 ± 12.829 for ALS, 46.650 ± 12.596 for HD, and 66.800 ± 10.851 for PD). As mentioned earlier and according to [42], another limitation derived from the usage of PhysioNet database is that during VGRF measurements, participants turned back when they reached the end of the $77\ m$ long corridor in which the gait test was conducted. This change in walking direction generates strides with different events to those of a straight-line gait. Finally, a disadvantage of normalizing the stride duration to $[0\%, 100\%]$ is the loss of the variability of this gait parameter, which has been recognized as significantly influenced by aging [78].

Further studies will need to consider the laterality of participants, i.e., the dominance of one side of the brain in controlling pairs of organs e.g., the eyes, the legs or hands. Laterality could be useful in explaining why each branch of the tree presented in Figure 4.5 depends on one side of the body and not on the other. From a theoretical perspective, future work should explain why a given disease affects certain transitions more than others.

Bibliography

[1] C. Kirtley, "Clinical gait analysis: theory and practice." Elsevier Health Sciences, 2006.

[2] M. W. Whittle, "Gait analysis," in *The Soft Tissues*. Elsevier, 1993, pp. 187–199.

[3] J. Perry, J. R. Davids *et al.*, "Gait analysis: normal and pathological function," *Journal of Pediatric Orthopaedics*, vol. 12, no. 6, p. 815, 1992.

[4] L. Rokach and O. Maimon, "Decision trees," in *Data mining and knowledge discovery handbook*. Springer, 2005, pp. 165–192.

[5] Y. Xia, Q. Gao, and Q. Ye, "Classification of gait rhythm signals between patients with neuro-degenerative diseases and normal subjects: Experiments with statistical features and different classification models," *Biomedical Signal Processing and Control*, vol. 18, pp. 254 – 262, 2015. [Online]. Available: https://doi.org/10.1016/j.bspc.2015.02.002

[6] R. Pahwa and K. E. Lyons, *Handbook of Parkinson's disease*. Crc Press, 2013.

[7] L. M. De Lau and M. M. Breteler, "Epidemiology of Parkinson's disease," *The Lancet Neurology*, vol. 5, no. 6, pp. 525–535, 2006. [Online]. Available: https://doi.org/10.1007/s00702-017-1686-y

[8] Y. Moon, J. Sung, R. An, M. E. Hernandez, and J. J. Sosnoff, "Gait variability in people with neurological disorders: a systematic review and meta-analysis," *Human movement science*, vol. 47, pp. 197–208, 2016. [Online]. Available: https://doi.org/10.1016/j.humov.2016.03.010

[9] S. J. Evans, I. Douglas, M. D. Rawlins, N. S. Wexler, S. J. Tabrizi, and L. Smeeth, "Prevalence of adult Huntington's disease in the UK based on diagnoses recorded in general practice records," *J Neurol Neurosurg Psychiatry*, vol. 84, no. 10, pp. 1156–1160, 2013. [Online]. Available: http://dx.doi.org/10.1136/jnnp-2012-304636

[10] R. Reilmann, "Parkinsonism in Huntington's disease." *International review of neurobiology*, vol. 149, pp. 299–306, 2019. [Online]. Available: https://doi.org/10.1016/bs.irn.2019.10.006

[11] R. M. Bonelli and M. F. Beal, "Chapter 30 - Huntington's disease," in *Neurobiology of Psychiatric Disorders*, ser. Handbook of Clinical Neurology, M. J. Aminoff, F. Boller, and D. F. Swaab, Eds.   Elsevier, 2012, vol. 106, pp. 507 – 526. [Online]. Available: https://doi.org/10.1016/B978-0-444-52002-9.00030-9

[12] R. H. Brown, "Amyotrophic lateral sclerosis: insights from genetics," *Archives of Neurology*, vol. 54, no. 10, pp. 1246–1250, 1997. [Online]. Available: http://dx.doi.org/10.1001/archneur.1997.00550220050013

[13] M. A. Van Es, H. J. Schelhaas, P. W. Van Vught, N. Ticozzi, P. M. Andersen, E. J. Groen, C. Schulte, H. M. Blauw, M. Koppers, F. P. Diekstra *et al.*, "Angiogenin variants in Parkinson disease and amyotrophic lateral sclerosis," *Annals of neurology*, vol. 70, no. 6, pp. 964–973, 2011. [Online]. Available: https://doi.org/10.1002/ana.22611

[14] A. J. Hughes, S. E. Daniel, Y. Ben-Shlomo, and A. J. Lees, "The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service," *Brain*, vol. 125, no. 4, pp. 861–870, 2002. [Online]. Available: https://doi.org/10.1093/brain/awf080

[15] S. Keus, A. Nieuwboer, B. Bloem, G. Borm, and M. Munneke, "Clinimetric analyses of the modified Parkinson activity scale," *Parkinsonism & Related Disorders*, vol. 15, no. 4, pp. 263 – 269, 2009. [Online]. Available: https: //doi.org/10.1016/j.parkreldis.2008.06.003

[16] Huntington Study Group, "Unified Huntington's disease rating scale: Reliability and consistency," *Movement Disorders*, vol. 11, no. 2, pp. 136–142, 1996. [Online]. Available: https://doi.org/10.1002/mds.870110204

[17] J. M. Cedarbaum and N. Stambler, "Performance of the amyotrophic lateral sclerosis functional rating scale (ALSFRS) in multicenter clinical trials," *Journal of the Neurological Sciences*, vol. 152, pp. s1 – s9, 1997. [Online]. Available: https://doi.org/10.1016/S0022-510X(97)00237-2

[18] B. Rana, A. Juneja, M. Saxena, S. Gudwani, S. S. Kumaran, M. Behari, and R. Agrawal, "Relevant 3D local binary pattern based features from fused feature descriptor for differential diagnosis of Parkinson's disease using structural MRI," *Biomedical Signal Processing and Control*, vol. 34, pp. 134 – 143, 2017. [Online]. Available: http://doi.org/10.1016/j.bspc.2017.01.007

[19] C. Buckley, L. Alcock, R. McArdle, R. Ur Rehman, S. Del Din, C. Mazzà, A. Yarnall, and L. Rochester, "The role of movement analysis in diagnosing and monitoring neurodegenerative conditions: Insights from gait and postural control," *Brain Sciences*, vol. 9, no. 2, 2019. [Online]. Available: https: //doi.org/10.3390/brainsci9020034

[20] P. J. Blijham, H. J. Schelhaas, H. J. ter Laak, B. G. van Engelen, and M. J. Zwarts, "Early diagnosis of ALS: the search for signs of denervation in clinically normal

muscles," *Journal of the neurological sciences*, vol. 263, no. 1-2, pp. 154–157, 2007. [Online]. Available: https://doi.org/10.1016/j.jns.2007.07.008

[21] D. Joshi, A. Khajuria, and P. Joshi, "An automatic non-invasive method for Parkinson's disease classification," *Computer methods and programs in biomedicine*, vol. 145, pp. 135–145, 2017. [Online]. Available: https://doi.org/10.1016/j.cmpb.2017.04.007

[22] S. Bilgin and Z. E. Akin, "Gait pattern discrimination of ALS patients using classification methods," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 26, no. 3, pp. 1367–1377, 2018. [Online]. Available: https://doi.org/10.3906/elk-1708-221

[23] S. T. Schwarz, T. Rittman, V. Gontu, P. S. Morgan, N. Bajaj, and D. P. Auer, "T1-weighted MRI shows stage-dependent substantia nigra signal loss in Parkinson's disease," *Movement Disorders*, vol. 26, no. 9, pp. 1633 – 1638, 2011. [Online]. Available: https://doi.org/10.1002/mds.23722

[24] F. B. Rodrigues, L. M. Byrne, and E. J. Wild, *Biofluid Biomarkers in Huntington's Disease.* New York, NY: Springer New York, 2018, pp. 329–396. [Online]. Available: https://doi.org/10.1007/978-1-4939-7825-0_17

[25] F. N. Emamzadeh and A. Surguchov, "Parkinson's disease: Biomarkers, treatment, and risk factors," *Frontiers in neuroscience*, vol. 12, no. 30214392, pp. 612–612, Aug. 2018. [Online]. Available: https://doi.org/10.3389/fnins.2018.00612

[26] Y. Xia, Q. Gao, Y. Lu, and Q. Ye, "A novel approach for analysis of altered gait variability in amyotrophic lateral sclerosis," *Medical & biological engineering & computing*, vol. 54, no. 9, pp. 1399–1408, 2016. [Online]. Available: https://doi.org/10.1007/s11517-015-1413-5

[27] A. Khajuria, P. Joshi, and D. Joshi, "Comprehensive statistical analysis of the gait parameters in neurodegenerative diseases," *Neurophysiology*, vol. 50, no. 1, pp. 38–51, 2018. [Online]. Available: https://doi.org/10.1007/s11062-018-9715-5

[28] N. Chastan, W.-N. Bair, S. M. Resnick, S. A. Studenski, and L. M. Decker, "Prediagnostic markers of idiopathic Parkinson's disease: Gait, visuospatial ability and executive function," *Gait & posture*, vol. 68, pp. 500–505, 2019. [Online]. Available: https://doi.org/10.1016/j.gaitpost.2018.12.039

[29] J. M. Hausdorff, S. L. Mitchell, R. Firtion, C. K. Peng, M. E. Cudkowicz, J. Y. Wei, and A. L. Goldberger, "Altered fractal dynamics of gait: reduced stride-interval correlations with aging and Huntington's disease," *Journal of Applied Physiology*, vol. 82, no. 1, pp. 262 – 269, 1997. [Online]. Available: https://doi.org/10.1152/jappl.1997.82.1.262

[30] J. M. Hausdorff, A. Lertratanakul, M. E. Cudkowicz, A. L. Peterson, D. Kaliton, and A. L. Goldberger, "Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis," *Journal of Applied Physiology*, vol. 88, no. 6, pp. 2045–2053, 2000. [Online]. Available: https://doi.org/10.1152/jappl.2000.88.6.2045

[31] T.-Z. Chen, G.-J. Xu, G.-A. Zhou, J.-R. Wang, P. Chan, and Y.-F. Du, "Postural sway in idiopathic rapid eye movement sleep behavior disorder: A potential marker of prodromal Parkinson's disease," *Brain research*, vol. 1559, pp. 26–32, 2014. [Online]. Available: https://doi.org/10.1016/j.brainres.2014.02.040

[32] A. Mirelman, H. Bernad-Elazari, A. Thaler, E. Giladi-Yacobi, T. Gurevich, M. Gana-Weisz, R. Saunders-Pullman, D. Raymond, N. Doan, S. B. Bressman *et al.*, "Arm swing as a potential new prodromal marker of Parkinson's disease," *Movement Disorders*, vol. 31, no. 10, pp. 1527–1534, 2016. [Online]. Available: https://doi.org/10.1002/mds.26720

[33] A.-B. Liu and C.-W. Lin, "Multiscale approximate entropy for gait analysis in patients with neurodegenerative diseases," *Entropy*, vol. 21, no. 10, 2019. [Online]. Available: https://doi.org/10.3390/e21100934

[34] J. M. Hausdorff, S. L. Mitchell, R. Firtion, C.-K. Peng, M. E. Cudkowicz, J. Y. Wei, and A. L. Goldberger, "Altered fractal dynamics of gait: reduced stride-interval correlations with aging and huntington's disease," *Journal of applied physiology*, vol. 82, no. 1, pp. 262–269, 1997.

[35] M. Nieto-Hidalgo, F. J. Fernández-Pastor, R. J. Valdivieso-Sarabia, J. Mora-Pascual, and J. M. García-Chamizo, "A vision based proposal for classification of normal and abnormal gait using RGB camera," *Journal of Biomedical Informatics*, vol. 63, pp. 82 – 89, 2016. [Online]. Available: http://doi.org/10.1016/j.jbi.2016.08.003

[36] Y. Wu and S. Krishnan, "Statistical analysis of gait rhythm in patients with Parkinson's disease," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 2, pp. 150–158, 2010. [Online]. Available: https://doi.org/10.1109/TNSRE.2009.2033062

[37] A. Procházka, O. Vyšata, M. Vališ, O. Ťupa, M. Schätz, and V. Mařřík, "Bayesian classification and analysis of gait disorders using image and depth sensors of Microsoft Kinect," *Digital Signal Processing*, vol. 47, pp. 169 – 177, 2015. [Online]. Available: http://doi.org/10.1016/j.dsp.2015.05.011

[38] F. Wahid, R. K. Begg, C. J. Hass, S. Halgamuge, and D. C. Ackland, "Classification of Parkinson's disease gait using spatial-temporal gait features," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 6, pp. 1794–1802, 2015. [Online]. Available: https://doi.org/10.1109/JBHI.2015.2450232

[39] T. D. Pham and H. Yan, "Tensor decomposition of gait dynamics in Parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 8, pp. 1820–1827, 2018. [Online]. Available: https://doi.org/10.1109/TBME.2017.2779884

[40] B. Vidya and S. P, "Gait based parkinson's disease diagnosis and severity rating using multi-class support vector machine," *Applied Soft Computing*, vol. 113, p. 107939, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494621008619

[41] S. Farashi, "Analysis of the stance phase of the gait cycle in parkinson's disease and its potency for parkinson's disease discrimination," *Journal of Biomechanics*, vol. 129, p. 110818, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0021929021005789

[42] Y. Wu and L. Shi, "Analysis of altered gait cycle duration in amyotrophic lateral sclerosis based on nonparametric probability density function estimation," *Medical Engineering & Physics*, vol. 33, no. 3, pp. 347 – 355, 2011. [Online]. Available: https://doi.org/10.1016/j.medengphy.2010.10.023

[43] M. R. Daliri, "Automatic diagnosis of neuro-degenerative diseases using gait dynamics," *Measurement*, vol. 45, no. 7, pp. 1729 – 1734, 2012. [Online]. Available: https://doi.org/10.1016/j.measurement.2012.04.013

[44] P. Ren, S. Tang, F. Fang, L. Luo, L. Xu, M. L. Bringas-Vega, D. Yao, K. M. Kendrick, and P. A. Valdes-Sosa, "Gait rhythm fluctuation analysis for neurodegenerative diseases by empirical mode decomposition," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 52–60, 2017. [Online]. Available: https://doi.org/10.1109/TBME.2016.2536438

[45] T. D. Pham, "Texture classification and visualization of time series of gait dynamics in patients with neuro-degenerative diseases," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 1, pp. 188–196, 2018. [Online]. Available: https://doi.org/10.1109/TNSRE.2017.2732448

[46] P. Ghaderyan and S. M. G. Beyrami], "Neurodegenerative diseases detection using distance metrics and sparse coding: A new perspective on gait symmetric features," *Computers in Biology and Medicine*, vol. 120, p. 103736, 2020. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2020.103736

[47] Y. Yan, K. Ivanov, O. Mumini Omisore, T. Igbe, Q. Liu, Z. Nie, and L. Wang, "Gait rhythm dynamics for neuro-degenerative disease classification via persistence landscape- based topological representation," *Sensors*, vol. 20, no. 7, p. 2006, Apr 2020. [Online]. Available: https://doi.org/10.3390/s20072006

[48] M. Saljuqi and P. Ghaderyan, "A novel method based on matching pursuit decomposition of gait signals for parkinson's disease, amyotrophic lateral sclerosis and huntington's disease detection," *Neuroscience Letters*, vol. 761, p. 136107, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0304394021004857

[49] Çağatay Berke Erdaş, E. Sümer, and S. Kibaroğlu, "Neurodegenerative disease detection and severity prediction using deep learning approaches," *Biomedical Signal Processing and Control*, vol. 70, p. 103069, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809421006662

[50] Ö. F. Ertuğrul, Y. Kaya, R. Tekin, and M. N. Almalı, "Detection of parkinson's disease by shifted one dimensional local binary patterns from gait," *Expert Systems with Applications*, vol. 56, pp. 156–163, 2016. [Online]. Available: https://doi.org/10.1016/j.eswa.2016.03.018

[51] Q. Ye, Y. Xia, and Z. Yao, "Classification of gait patterns in patients with neurodegenerative disease using adaptive neuro-fuzzy inference system," *Computational and mathematical methods in medicine*, vol. 2018, p. 9831252, 2018. [Online]. Available: https://doi.org/10.1155/2018/9831252

[52] W. Zeng and C. Wang, "Classification of neurodegenerative diseases using gait dynamics via deterministic learning," *Information sciences*, vol. 317, pp. 246–258, 2015. [Online]. Available: https://doi.org/10.1016/j.ins.2015.04.047

[53] P. Prabhu, A. Karunakar, H. Anitha, and N. Pradhan, "Classification of gait signals into different neurodegenerative diseases using statistical analysis and recurrence quantification analysis," *Pattern Recognition Letters*, vol. 139, pp. 10–16, 2018. [Online]. Available: https://doi.org/10.1016/j.patrec.2018.05.006

[54] S. M. G. Beyrami and P. Ghaderyan, "A robust, cost-effective and non-invasive computer-aided method for diagnosis three types of neurodegenerative diseases with gait signal analysis," *Measurement*, vol. 156, p. 107579, 2020. [Online]. Available: https://doi.org/10.1016/j.measurement.2020.107579

[55] A. Mengarelli, A. Tigrini, S. Fioretti, and F. Verdini, "Recurrence quantification analysis of gait rhythm in patients affected by parkinson's disease," in *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE, 2021, pp. 1–4. [Online]. Available: https://doi.org/10.1109/BHI50953.2021.9508550

[56] S. Lahmiri, "Gait nonlinear patterns related to parkinson's disease and age," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 7, pp. 2545–2551, 2018. [Online]. Available: https://doi.org/10.1109/TIM.2018.2866316

[57] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.

[58] W. Zuberek, "Timed Petri nets definitions, properties, and applications," *Microelectronics Reliability*, vol. 31, no. 4, pp. 627 – 644, 1991. [Online]. Available: https://doi.org/10.1016/0026-2714(91)90007-T

[59] J. Bengtsson and W. Yi, *Timed Automata: Semantics, Algorithms and Tools*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 87 – 124. [Online]. Available: 10.1007/978-3-540-27755-2_3

[60] D. Muñoz-Añasco, A. Correcher, E. Garcia, and F. Morant, "Identification of stochastic timed discrete event systems with st-IPN," *Mathematical Problems in Engineering*, vol. 2014, pp. 1–21, 07 2014. [Online]. Available: https://doi.org/10.1155/2014/835312

[61] E. A. Ihlen, O. Sletvold, T. Goihl, P. B. Wik, B. Vereijken, and J. Helbostad, "Older adults have unstable gait kinematics during weight transfer," *Journal of Biomechanics*, vol. 45, no. 9, pp. 1559 – 1565, 2012. [Online]. Available: https://doi.org/10.1016/j.jbiomech.2012.04.021

[62] Y. Liu, K. Lu, S. Yan, M. Sun, D. K. Lester, and K. Zhang, "Gait phase varies over velocities," *Gait & Posture*, vol. 39, no. 2, pp. 756 – 760, 2014. [Online]. Available: https://doi.org/10.1016/j.gaitpost.2013.10.009

[63] F. Hebenstreit, A. Leibold, S. Krinner, G. Welsch, M. Lochmann, and B. Eskofier, "Effect of walking speed on gait sub phase durations," *Human Movement Science*, vol. 43, pp. 118–124, 2015. [Online]. Available: https://doi.org/10.1016/j.humov.2015.07.009

[64] Gillette Children's Hospital, "Normal walking: An overview based on gait analysis." St. Paul, Minnesota: Gillette Children's Hospital, 1993.

[65] I. Birch, W. Vernon, J. Walker, and M. Young, "Terminology and forensic gait analysis," *Science & Justice*, vol. 55, no. 4, pp. 279–284, jul 2015.

[66] S. C. Vázquez, "Análisis de la marcha humana con plataformas dinamométricas influencia del transporte de carga." Madrid: Universidad Complutense de Madrid, 2002, p. 343.

[67] M. P. Murray, "Gait as a total pattern of movement: Including a bibliography on gait," *American Journal of Physical Medicine & Rehabilitation*, vol. 46, no. 1, pp. 290–333, 1967.

[68] E. Ayyappa, "Normal human locomotion, part 1: Basic concepts and terminology," *JPO: Journal of Prosthetics and Orthotics*, vol. 9, no. 1, pp. 10–17, 1997.

[69] C. G. Cassandras and S. Lafortune, *Introduction to discrete event systems.* Springer Science & Business Media, 2009.

[70] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Señales y sistemas.* Pearson Educación, 1998.

[71] J. Peter, W. M. Wonham *et al.*, "The control of discrete event systems," *Proceeding of the IEEE*, vol. 77, no. 1, pp. 81–98, 1989.

[72] S. Engell, G. Frehse, and E. Schnieder, *Modelling, analysis and design of hybrid systems.* Springer, 2003, vol. 279.

[73] T. K. Ho, "Random decision forests," in *Proceedings of 3rd international conference on document analysis and recognition*, vol. 1.   IEEE, 1995, pp. 278–282.

[74] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, no. 2, pp. 197–227, 2016.

[75] J. M. Hausdorff, A. Lertratanakul, M. E. Cudkowicz, A. L. Peterson, D. Kaliton, and A. L. Goldberger, "Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis," *Journal of applied physiology*, 2000.

[76] D. M. M. Añasco, "Identificación y diagnóstico de fallos en sistemas de eventos discretos estocásticos." Ph.D. dissertation, 2015.

[77] M. Kuhn, "Building predictive models in R using the caret package," *Journal of Statistical Software*, vol. 28, no. 5, pp. 1–26, 2008. [Online]. Available: https://doi.org/10.18637/jss.v028.i05

[78] S. Lord, T. Howe, J. Greenland, L. Simpson, and L. Rochester, "Gait variability in older adults: A structured review of testing protocol and clinimetric properties," *Gait & Posture*, vol. 34, no. 4, pp. 443–450, 2011. [Online]. Available: https://doi.org/10.1016/j.gaitpost.2011.07.010