Faculdade de Engenharia da Universidade do Porto

Parkinson's Disease Diagnosis: A Machine Learning and Data Mining based Approach

Juliana Rajão Guerra

Dissertation for the degree of Master in Biomedical Engineering

Supervisor: Prof. Doutor João Manuel R. S. Tavares,
Faculdade de Engenharia da Universidade do Porto

June 2019

Abstract

Parkinson's disease (PD) is the second most common neurodegenerative disorder worldwide after Alzheimer's disease and its prevalence is increasing as the world's population ages. This condition occurs due to a chronic disturbance of the central nervous system. The disruption of dopamine-producing neurons leads to the loss of muscle function, to limb rigidity and to several other problems regarding posture, gait, and balance. Yet, it is still a disease with no cure leading to a continuous study by many researchers. However, available treatments reduce its severity and delay its progression. As for the diagnosis, it can be evidently conclusive on advanced stages of the disease based on the main motor or non-motor symptoms that can be only noticeable later on, when cognitive function has already been lost. Thus, this situation encourages investment in efficient detection for these patients in order to provide an early and accurate diagnosis.

Overall, the motivation for this proposed approach on the problem, besides the wellbeing and diagnosis of PD patients, was to explore classification models, its differences and stronger points for which could provide a more promising support mechanism for doctors in the near future.

Once comprehending Parkinson’s Disease as for symptoms and current diagnosis process, this dissertation proposes a machine learning-based methodology for PD diagnosis from the Parkinson’s Progression Markers Initiative database. In order to classify PD and healthy subjects, a supervised machine learning algorithm was developed in *Python* and is presented in this dissertation. The proposed methodology allowed comparison and analysis of classification results obtained by implementing several different classifiers to develop a high-performance system. The selected classifiers consist on the most reliable algorithm considering this approach: Random Forest, Logistic Regression, Naïve Bayes, Support Vector Machine, Nearest Neighbors and Neural Network.

This methodology was divided into two approaches, one using all data available and other implementing a feature dimensional reduction for feature selection, as well as, a hyperparameter search for model optimization in order to provide better results. Considering both, the classification accuracy was between about 77-89%. Once having the classification results for each method, the most were evaluated, concluding that the best one was Random Forest with reduced features dataset. This implementation achieving an 88.7% accuracy with 87% precision, recall and F1-score. Considering the proposed methodology results satisfactory, suggest that this process could be applied for real PD detection as for now since further developments are required to better validate and improve the results.

Resumo

A doença de Parkinson é o segundo distúrbio neurológico mais comum mundialmente depois da doença de Alzheimer e a sua prevalência tem vindo a aumentar à medida que a população mundial envelhece. Esta condição ocorre devido a uma perturbação crônica do sistema nervoso central, em que os neurónios produtores de dopamina presentes numa área específica do cérebro são afetados. Esta condição provoca a perda de funções musculares como a bradicinesia, rigidez dos membros, assim como problemas de postura, marcha e desequilíbrio. Atualmente, ainda não existe cura, sendo uma doença em constante estudo. No entanto, os tratamentos disponíveis reduzem sua gravidade e retardam sua progressão. Quanto ao diagnóstico, este pode ser conclusivo já em estados avançados da doença recorrendo a testes baseados nos principais sintomas motores ou não motores. Estes podem apenas manifestar-se tardiamente, ou seja, quando a função cognitiva já tiver sido perdida. Sendo assim, existe um incentivo para o investimento em procedimentos de diagnostico eficiente nos estados iniciais da doença de modo a promover a deteção e início de tratamentos mais precoce. Assim, a motivação para a abordagem proposta desta dissertação sobre o problema foi explorar modelos de classificação, as suas diferenças e pontos mais fortes de modo a reforçar quais poderiam fornecer um mecanismo de apoio mais promissor para os médicos futuramente.

Compreendendo os conceitos básicos relativamente à Doença de Parkinson, esta dissertação propõe uma metodologia baseada em *machine learning* para o diagnóstico da patologia utilizando uma base de dados internacional *Parkinson's Progression Markers Initiative*. Foi assim feito o desenvolvimento e apresentação de um algoritmo em *Python* de aprendizagem supervisionada para a classificação entre indivíduos com *PD* e saudáveis. A metodologia proposta permitiu uma análise e comparação de classificação obtidos pela implementação de vários classificadores com o objetivo de desenvolver um sistema de alto desempenho. Os classificadores selecionados foram: *Random Forest, Logistic Regression, Naïve Bayes, Support Vector Machine, Nearest Neighbors* e *Neural Network*. Esta metodologia foi dividida em duas partes, uma utilizando todos os dados disponíveis e outra implementando uma redução dimensional de parâmetros (*features)* para a sua seleção. Ainda foi desenvolvido um estudo de híper-parâmetros dos classificadores para a otimização do modelo, de forma a obter melhores resultados. Considerando ambas partes, a exatidão da classificação abrange valores entre 77-89%. Comparando resultados, os mais promissores foram avaliados, concluindo que o melhor foi o *Random Forest* na segunda abordagem. Esta implementação alcançou uma precisão de 88,7% com precisão, sensibilidade e *F1-score* de 87%. Considerando os resultados da metodologia proposta satisfatórios, é possível, então, sugerir que este processo com aplicação para a deteção da patologia. No entanto, seriam necessários mais desenvolvimentos para uma melhor validação e obtenção de resultados.

Acknowledgments

Primeiramente, ao impulsionador desta dissertação, ao Professor Doutor João Manuel R. S. Tavares, pela sua orientação e apoio no desenvolvimento desta fase do Mestrado.

À minha família, à minha grande família, que fosse num encontro familiar, Natal, Páscoa, aos Domingos, sempre me mostraram o seu apoio incondicional e tentaram transmitir a sua sabedoria e os melhores conselhos. Mas, um agradecimento mais forte à minha mãe, pai, irmã e Hugo, que mesmo em fases menos boas conseguiram sempre dar-me o melhor espaço possível para estudar e lutar pelos meus objetivos. São os pilares da minha vida, um apoio incondicional e juntos conseguimos ultrapassar qualquer obstáculo. Esta dissertação também vos pertence.

Um grande agradecimento aos meus amigos, de norte a sul do país, que não deixaram a distância condicionar o seu apoio e aos meus companheiros para a vida, desde os tempos da Eça até agora, que estiveram sempre lá e me viram crescer ao longo do curso! Oura, Inês e Mariana, obrigada pelo apoio motivacional e emocional infinito.

Obrigada ao "Congresso MEBiano", um grupo de 10 jovens que eu adoro! Pela bela turma que fomos, pela companhia nas aulas e nas tardes/noitadas de estudo, pelo trabalho árduo, puxões de orelhas, conselhos e entreajuda, mas também pelas viagens, jantares, consultas de psicologia, crises existenciais coletivas, gargalhadas e muito mais: Bia, Bruno, Cat, Ciclas, Luís, Marisa, Miguel, Raquel e Tomé, as saudades da Casa do Povo serão sentidas! Obrigada por estes 2 anos! Ainda, ao Luís por ter sido um anjo da guarda para os maiores desafios. Por fim, ao Buno, por tudo, mas tudo mesmo, mesmo. Desde risos a choros, obrigada por estares sempre lá! Somos a melhor equipa.

Juliana Guerra

“A journey of a thousand miles begins with a single step.”

Lao Tzu, Chinese Proverb

Contents

List of Figures

List of Tables

Abbreviations, Acronyms and Symbols

List of abbreviations and acronyms

ACC	Accuracy
AI	Artificial Intelligence
AIME	Artificial Intelligence in Medicine Europe
ANN	Artificial Neural Networks
ap	Anterior-posterior
AUC	Area Under the Curve
CISI	Clinical Impression of Severity Index
CNS	Central Nervous System
COM	Centre of Mass
COP	Centre of Pressure
CSF	*Cerebrospinal* fluid
DBS	Deep Brain Stimulation Therapy
DT	Decision Tree
EEG	Electroencephalogram
FDA	Food and Drug Administration
fMRI	functional Magnetic Resonance Imaging
FN	False Negatives
FP	False Positives
FS	Feature Selection
GPS	General Problem Solver
IBK	Linear k-Nearest Neighbors
IG	Information Gain
kNN	k-Nearest Neighbors
LDA	Linear Discriminant Analysis
LR	Logistic Regression
MDS	Movement disorder specialist
MDS-UPDRS	Movement Disorder Society-sponsored revision of the UPDRS
MFEA	Multiple Feature Evaluation Approach
ml	medial-lateral
MLP	Multi-Layer Perceptron
MLR	Multivariate Logistic Regression

MRI	Magnetic Resonance Imaging
MSc	Master of Science
NB	Naïve Bayes
NN	Neural Network
OPF	Optimum-Path Forest
PD	Parkinson's Disease
PET	Positron Emission Tomography
PLS	Partial Least Squares
PPMI	Parkinson's Progression Markers Initiative
RBF	Radial Basis Function
REPTree	Regression Trees
RF	Random Forest
ROC	Receiver Operating Characteristic
SBR	Striatal Binding Ratio
SCM	Structural Cooccurrence Matrix
SPECT	Single Photon emission computed tomography
SU	Symmetrical uncertainty
SVM	Support Vector Machines
TN	True Negatives
TP	True Positives
UPDRS	Unified Parkinson's Disease Rating Scale
VGRF	Vertical Ground Reaction Force
WEKA	Waikato Environment for Knowledge Analysis
WHO	World Health Organization

Chapter 1

Introduction

The initial chapter presents a brief framework of the subject under study, as well as the outlined objectives and methodology that is planned to be applied. The chapter subsequently describes the structure of the present document and provides a chapter by chapter overview of the dissertation.

1.1 – Problem identification and Motivation

Parkinson's Disease (PD) is the second most common neurodegenerative disease after Alzheimer's disease and its prevalence is rising as the world's population ages. It is estimated that there are 7 to 10 million people living with Parkinson worldwide and, in Portugal, it affects 180 per 100 000 (one hundred thousand) inhabitants [1]–[3]. This condition occurs due to a chronic disturbance of the central nervous system: dopamine-producing ("dopaminergic") neurons on a specific area of the brain, called *substantia nigra*, are affected, leading to loss of muscle function which then results in bradykinesia (slowed movement), limb rigidity, impaired posture and gait and balance problems [4].

Despite being a progressive disease, the individual ends up achieving a state of dependency, which can lead to depression, confusion and/or loss of personality [4]. According to a 2014 study, PD is increasing rapidly and between 2005 and 2030 the number of cases will globally increase at least the double, as presented in Figure 1.1. This means a greater concern about this disease [5].

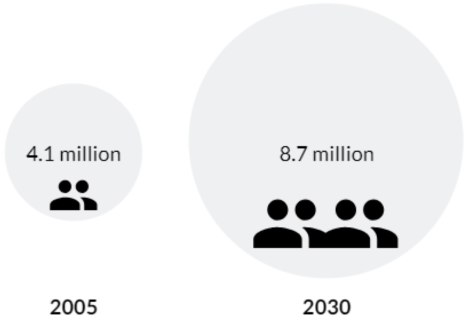

Figure 1.1 - Projected number of people with PD in the world from 2005 to 2030 [5].

Although PD is more common among elders, it is important to keep studying and learning about the disease such as ways to optimize prior diagnosis to effective treatments. According to a World Health Organization (WHO) statement, thus PD can be extremely variable in patients since it runs a chronic slowly progressive course. During the initial years, motor disability may not be noteworthy but if untreated, as the years sum up, it can lead to difficulties in daily life [6]. Degenerative disorders like dementia, strokes and PD, are becoming more common due to population ageing tendency in Europe [7].

These concerns support a more effective and early diagnosis of disease in order to be possible to provide a sustainable lifestyle to human population. Nowadays, the leading source of disability globally are neurological disorders, and ageing is increasing the burden of neurodegenerative disorders, including Parkinson's disease. Thus, early management and diagnosis of PD become crucial as time follows clinical symptoms become noticeable and more than 60% of the dopaminergic neurons have already been lost [8], [9]. According to Scopus [10], over the period of 1990 to 2018, there has been an increasing interest by researchers in general on PD diagnosis over the years. These results were found when searching concerning the keywords: “Parkinson’s disease” + “Diagnosis”. Represented in Figure 1.2, it is possible to verify the importance of researching and understanding PD as well as the growing interest in this area of research.

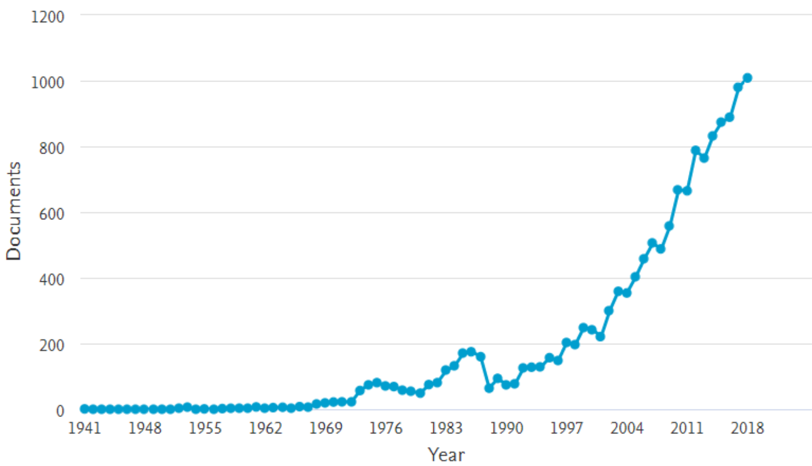

Figure 1.2 - Analysis of search results for Parkinson's Disease Diagnosis from Scopus [10].

Living in the digital era means easy access to data as well as to collecting and storing it, making information systems a common tool for both personal and professional use on a daily basis. Given the fact that equipped hospitals can monitor and collect data from countless devices, it is possible to associate the medical field with the use of expert systems as Machine Learning algorithms which were designed to analyse medical datasets. These expert systems consist of a process structured in several steps, from feature selection until the results.

Concerning minimum possible redundancy content, features must be properly selected to encode as much information as possible given the task of interest, taking into consideration the proximity measure as it will quantify the similarity or distance between the feature vectors. Since not always nor all features have the same impact on the analysis, clustering criterions are also an important part of the process, depending on the pretended sensibility. Once the results of the clustering algorithm are ready, validating becomes an important step as well as the interpretation of these results for taking conclusions. Furthermore, a new patient's diagnosis can be quicker and improved in accuracy and reliability, following this simplified logic along with traditional medical analysis [11], [12].

Currently, visual examination of patient motor performance continues to be the procedure for clinical diagnosis and evaluation of PD. Therefore, expert systems application by the analysis of previous cases can lead to a more objective quantification of the patient behaviour concerning the disease characteristics and symptoms, overtaking issues associated with subjectivity and inter-professional variability.

1.2 – Main goals and Methodology

An increasing number of research fields is following the rapid evolution of technology and the implementation of intelligent knowledge into their processes and systems. Medicine is no exception, as the world is investing in a fast evolution process and companies are growing, pursuing and reconciling several areas such as Engineering with Medicine. Since PD is a delicate disease, it is crucial to focus on research and studies not only to provide a better life for those already suffering but also to further improve the diagnosis and treatment of new cases. Adopting an intelligent system's knowledge can not only lead to faster but also more accurate diagnosis and treatment procedures since it can evaluate previous cases and learn from those. Regarding the specialist's knowledge, Machine Learning can work as a support tool for better and quicker response to exams data [3], [12], [13].

Considering the several exams performed on these patients, there is a large variety of medical areas and extraction of features to be approached, this dissertation aims to apply Machine Learning and Data Mining knowledge to its diagnosis. This provides an opportunity for implementing and testing new methods to reach a better diagnosis practice for PD. Besides, the first steps consisted of acquiring theoretical foundations for the dissertation as concepts of the pathology, focusing on research and learning about intelligent systems. This process approached the understanding of Data Mining and Machine Learning algorithms applications, such as diagnosis, as it consists of an area of interest concerning the biomedical field.

In this context, the main objectives for the dissertation are the development, analysis, and comparison between methods for a Parkinson's Disease Database analysis applying a Machine Learning algorithms approach by developing potentialities for assisting health professionals in the identification and classification of Parkinson.

1.3 – Dissertation achievements

The development of this dissertation provided a study focused on the data acquired in the most recent trial of the PPMI database, a major observational clinical study of PD markers. From the theoretical part, it also contributed to the familiarization with Machine Learning concepts and in-depth research on PD principles, as an important foundation for the algorithm implementation phase.

The main focus contributes to a wide performance comparison of classification models, with different learning methods, applied to a multi-type-feature dataset for PD diagnosis. The proposed methodology also provided a hyperparameter search for model optimization, considering two approaches on feature selection. This allowed concluding the approach and model that could deliver better classification results and be used as a support mechanism for experts.

1.4 – Dissertation structure

This document is structured in more six independent chapters, followed by bibliographical references. The six chapters that follow focus on the theoretical introduction of some fundamentals about Parkinson's Disease and the basic principles of Machine learning and Data Mining. Subsequently, a description of the state of the art is made regarding research studies of different approaches for PD diagnosis, using as well as diverse methods of Machine Learning.

The last chapter presents some important conclusions obtained through the work developed in this report in the scope of the Dissertation in Biomedical Engineering.

The following is a summary of the structuring content of each of the remaining chapters of this document:

- Chapter 2: Parkinson's Disease theoretical principles

This chapter addresses basic concepts for understanding the pathology of Parkinson's Disease. It is organized into five sections, where the description of the disease is carried out, its symptoms and clinical characteristics, an approach to the criteria used for diagnosis and some treatment methods. Finally, some final remarks about its principles.

- Chapter 3: Principles concepts of Machine Learning

Chapter 3, organized in eight sections, presents the basic theoretical foundations of Machine Learning, namely an overview and the physical principles inherent to this technique. Some important concepts are also described for Data Mining, Feature Selection and classification algorithms. Thus, this chapter describes the main principles of Machine Learning and Data Mining as well as its role on Diagnosis, the core point of this study.

- Chapter 4: Literature review: Machine learning technics applied to Parkinson's Disease Diagnosis

In this chapter, several methodologies currently in the field of the dissertation are presented and various applications examples are indicated. After introducing some fundamentals of Machine learning in the previous chapter, this chapter presents some studies that applied this technique on different scenarios for PD diagnosis. Chapter 4 closes with a summary of the work done over the previous semester.

- Chapter 5: Database characterization and Methodology for Parkinson's Disease Detection

Chapter 5 describes the main development steps of the proposed methodology are presented along with the theoretical fundamentals needed to understand each step. Hence, a detailed description of the public dataset using during the dissertation and the process of data organization is done.

- Chapter 6: Results and Discussion

This chapter contains the presentation, detailed analysis and discussion of the results obtained by the proposed and followed methodology of the previous chapter. It provides a comparison between the chosen models after its optimization, concluding which could provide a better solution for the problem.

- Chapter 7: Conclusions

The last chapter sets out the final conclusions of the dissertation and consequent critical reflection on the followed methodology. It emphasizes the importance of this theme research and study, focusing on the most critical and current aspects of Machine Learning application in Parkinson's Disease diagnosis as well as some final remarks on the classification results by the implemented methodology. Finally, some perspectives on aspects that could be improved or carried out as continuous future work.

Chapter 2

Parkinson's Disease theoretical principles

Parkinson's Disease (PD) is one of the most common neurodegenerative diseases worldwide, mainly characterized by a chronic disturbance of the central nervous system. These lesions cause intermittent neurological signs and symptoms that progressively worsen as the disease advances. This disease is more common in the elderly since ageing sparks modifications in our brains, declining its performance due to deterioration of synaptic contact and in neurotransmitters and neurohormones.

This chapter addresses the central theme of this study, Parkinson's disease. Initially, a historical context is made about the pathology, then proceed to its description, causes, predominance, and forms of manifestation. Next, the diagnostic criteria, as well as the treatments used by health professionals, are described. At the end of the chapter, the conclusions of the exposed information are presented.

2.1 – Disease Overview

Parkinson's Disease was first described by James Parkinson, approximately 200 years ago, in 1817. According to his book "An Essay on the Shaking Palsy", Parkinson describes the disease as an "involuntary tremulous motion, with lessened muscular power, in parts not in action and even when supported; with a propensity to bend the trunk forwards, and to pass from a walking to a running pace: the senses and intellects being uninjured" [14].

After Alzheimer's disease, it is the second most common neurodegenerative disorder and affects approximately 1% of the worldwide population over the age of 65. Primarily it was described as a condition characterized by diminished muscle strength leading to involuntary tremulous movement. [6], [15], [16].

PD incidence in Europe is estimated from 11 up to 19 new cases per 100 000 inhabitants per year and a prevalence of 108 to 257 cases per 100 000 inhabitants per year [17],[7]. In Portugal, according to a 2016 study carried by the Ethics Committee of the Faculty of Medicine of the University of Lisbon, for the mainland Portuguese population over 50 years of age, the number of PD cases was estimated around 240 per 100 000 inhabitants. But considering the total Portuguese population the estimative drops to 180 per 100 000 inhabitants [2], [3]. Nevertheless, when considering ages greater than 60 years, worldwide, these values tend to increase [18].

Overall, it is a chronic and progressive disease as well as a neurodegenerative disorder of the central nervous system (CNS). It is mainly characterized by symptoms involving the motor domain, presenting a peak between the ages of 60 and 65 years which can be considered the most critical period to its appearance. In men, it also shows an early appearance comparing in women with a ratio: 1.5 to 2.0 [18].

Taking essentially the motor domain into account, PD is mainly linked to tremors, muscle rigidity, a slowing of physical movement, and can also cause cognitive and mood disturbances. These symptoms are tightly associated to the concentration of dopamine in the brain and the loss of nerve cells (Dopaminergic neurons) in part of the midbrain known as the *substantia nigra* [6], [15], [16]. Dopamine is a neurotransmitter produced by the neurons on this part of the brain, that has an important role in movement control, Figure 2.1 [19].

Consequently, with this progressive deterioration of the dopaminergic neurons, there will be a significant decrease of dopamine available, reflecting that a loss around 80% results in PD symptoms becoming evident. Moreover, motor deficits appear. Although there are many studies concerning this pathology, there isn’t yet an absolutely clear cause of PD, some point to Dopaminergic neurons loss association with mutations or to some known toxins or chemicals that may also origin the disease [6], [15], [16], [19].

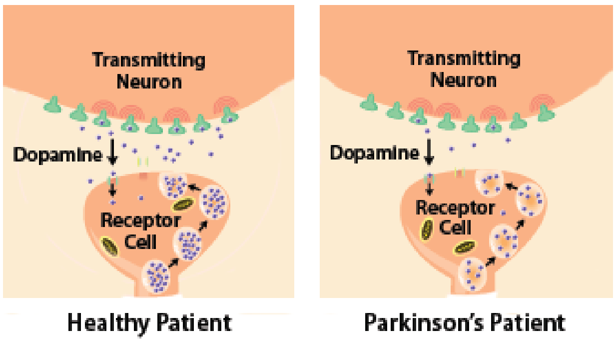

Figure 2.1 - Difference between dopamine production in a healthy person and Parkinson’s patient [20].

Despite being a progressive disease, the individual ends up achieving a state of dependency, which can result in depression, confusion and/or loss of personality. As stated before, according to some authors in epidemiology studies, risk factors that may increase the probability of PD appearance like high blood pressure, exposure to pesticides, tobacco consume, women on postmenopausal hormone replacement therapy treatments and so on [21]. The factors that contribute to the onset of the disease are mostly genetic, however, the exposure to pesticides or industrial toxins and ageing itself (above 50 years old) also play a fundamental role in its development [1], [22].

Even though there is currently no cure to this disease, symptoms can be controlled through drugs that stimulate the release of dopamine, with the condition that the patient still has neurons that produce this chemical [22]. If these cells are non-existent, the medication passes through the administration of levodopa, which is converted into dopamine at the brain level. Other therapies are used, such as Deep Brain Stimulation Therapy (DBS), but it has the disadvantage of restricting treatment to people over the age of 75 (most affected by the disease) and, on the other hand, is quite invasive, with surgical risks [23].

Meanwhile, even presenting a solid understanding of the physiologic mechanisms that lead to motor impairments in PD and the global effort to know the causes of Parkinson's Disease, there is still a gap between the present knowledge and understanding of the progressive dopaminergic neurons' degeneration and its implications. Therefore, an important part of this research effort relies on understanding the disease pathogenesis as well as cell death mechanics. Hence, it consequently offers a lack of effective therapies to both motor and non-motor symptoms prevention or improvement [6], [24].

2.2 – Parkinson's Disease Symptoms

PD symptoms and its manifestations are different in each patient and vary not only with disease evolution or therapy but also emotional and environmental factors. Nonetheless, there are four main features characterizing PD: Tremor while resting, muscular Rigidity, Akinesia (or bradykinesia) and Postural instability. These specific group of symptoms are commonly referred to as T.R.A.P., an acronym easy to remember. Although it can be considered five, depending on the author, adding walking or gait problems to the picture [7], [17], [25].

Meanwhile, motor symptoms only become noticeable much later than the pathological processes leading to the onset of PD since the diagnosis happens when many neurons cell already degenerate. Whiting the diagnosis process, approached on section 2.3, it is possible to find several rating scales for the evaluation of motor impairment and disability, even though most scales have not been fully evaluated for validity and reliability [25], [26].

Table 2.1 presents the main symptoms associated with PD. These motor and non-motor symptoms will then be briefly explained in the sections 2.2.1 and 2.2.2.

Table 2.1 - Most common symptoms of PD, based on [25].

Motor Symptoms	Non-motor Symptoms
Tremor	Disturbances in the Sense of Smell
Bradykinesia	Sleep Problems
Postural Instability	Depression and Anxiety
Walking or Gait Difficulties	Psychosis
Vocal Symptoms	Cognitive impairment
Decreased arm swing	
Difficulty in daily life activities	

2.2.1 – Motor Symptoms

This sub-section provides a description of five motor symptoms, which are considered the most characteristic symptoms. However, these can vary greatly from intensity and progression to time appearance in each patient.

- Tremor

Affecting almost 70% of PD patients, tremor sensation while resting became the most common symptom and the most easily recognized. However, tremors are not proof of Parkinson's nor every patient experience it [25], [27]. Tremors are unilateral and prominent in the distal part of an extremity occurring at a frequency 4 to 6 Hz. It begins affecting one side but eventually affects both sides of the body and can also occur in the jaw, mouth or tongue and leg. However, unlike essential tremor, rarely involves the neck/head or voice [25]–[27].

- Rigidity

Rigidity manifestation refers to an increased limb resistance, as limb tightness or stiffness, limiting limb passive movements range like flexion, extension or even rotation. During the disease initial stage, when it first starts to be perceptible, it is largely mistaking or misdiagnosed as orthopaedic problems or arthritis, such as a rotator cuff injury. Sometimes it is associated with pain, for example, shoulder pain [25], [27].

- Bradykinesia

Present in several movement disorders but also in PD, Bradykinesia refers to the slowing of movements. Adding up to the slowness, it also covers difficulties with planning, initiating and executing movements, as well as sequential/simultaneous tasks performance. These difficulties initiate as daily living activities and movement execution become slower and reaction time starts increasing [25], [27].

Other demonstrations of Bradykinesia include trouble turning over in bed, loss of spontaneous movements and gesturing, impaired swallowing (which provokes drooling) and eye blink rate decreasing [25], [27]. As occurring in other symptoms common, these manifestations depend on the patient emotional state [25].

- Postural Instability

Contrary to rigidity, which manifests itself in precocious stages, postural instability usually occurs after the onset of other clinical symptoms in later stages. It is the consequence of postural reflexes loss and consists of the inability to maintain a straight and steady posture, preventing falling episodes. This postural instability leads to an increasing risk of hip fracture and other physical injuries in the patient [27].

Balance problems and falling (especially backwards) is a tendency associated with this symptom [25], [27]. In order to test it, a patient is pulled backwards or forward by the shoulders in a quick form measuring the retropulsion or propulsion degree, respectively. An abnormal response is pointed if the patient needs more than two steps before stabilizing. Other PD symptoms contributing to instability are also orthostatic hypotension, age-related sensory changes as well as the own fear of falling [25].

- Walking or Gait Difficulties

As this pathology progresses, due to postural instability and bradykinesia PD patients begin to experience walking or gait difficulties. In early stages, it is common to detect a decreasing on one or both arms swinging movement as well as slower and smaller steps in later stages. Additionally, advanced cases may experience freezing episodes and short step with the almost nonexistent lifting of the feet [27].

Finally, among others and in addition to the five core motor symptoms aforementioned, many PD patients experience voice changes, for example becoming softer and fading at the end of sentences. These changes are allegedly a bradykinesia consequence and their speech can become monotone. However speedy and hesitating speech may appear in later stages [27].

2.2.2 – Non-motor Symptoms

PD is mainly considered a movement disorder, which can lead to other details being overlooked. However, it is important to pay attention to some non-motor symptoms as well [27]. Nevertheless, some of them can ever precede motor manifestations by years as changes in smell capacity. In addition to the presented symptoms, some patient can also experience depression [26].

- Sleep disorders

Previously researchers commonly considered that perturbations during sleep were just connected to people suffering from Parkinson. Currently, due to rapid eye movement sleep behaviour disorder observations, this disturbance is considered a feature and an integral part of PD. This is confirmed in approximately one-third of patients. Furthermore, episodes of falling asleep incapability, insomnia and more rarely to stay asleep and vivid dreams are also perceptible. Yet the least common are typically considered side effects of medications [25], [27].

- Cognitive and neurobehavioral abnormalities

Although not all patients experience these abnormalities, cognitive symptoms are considered very common among PD. These result in slower thinking capacity or task planning difficulties. Some domains can be severally affected like executive function and memory, eventually can lead to dementia which is a strong reason for nursing support [27], [28].

Thought studies state around 20 to 50% of PD patient to have a mild cognitive impairment, in other cases, it can be severe and have an impact in daily living functions, albeit not everyone even experience this [27].

- Autonomic dysfunction

This dysfunction is caused by central and peripheral autonomic nervous systems failure, compromising many functions: sweating abnormalities (excessive sweating, normally in the upper body, especially when untreated), sexual concerns (underrecognized feature that reveals desire reduction or libido), bladder dysfunction (urinary frequency and urgency that may be worse at night), light-headedness (faint feeling, contributes to gait problems) and orthostatic hypotension (blood pressure decreasing when quickly standing up) [25], [27], [28].

- Sensory abnormalities

Although often overlooked, some early PD symptoms encompass odour sensitivity decreasing or even smell capacity loss and may be experienced months or years before noticeable major motor symptoms [27].

- Psychosis

Over the course of their lives, the majority of patients with Parkinson's will develops several symptoms already mentioned, however, it is also possible to experience hallucinations and delusions episodes [27].

2.3 – Diagnosis

Regarding PD diagnosis, it is important to note that presently there is not a definitive diagnostic test and it is based on clinical criteria like the evaluation of the motor and non-motor symptoms [25]. However, there is a staging phase or, more precisely, Premotor/Prodromal phase that can last between 5 to 20 years for the onset of neurodegenerative and motor symptoms manifestation [8].

These clinical diagnoses are performed by a physician (neurologist trained to diagnose and treat neurologic disorders) evaluating the presence/absence of the possible symptoms by running a detailed neurologic examination. As a form to prevent misdiagnosis, beyond the neurologist evaluation, it is recommended to consult a movement disorder specialist (MDS) as well. However, PD diagnosis will only be considered if the person reveals at least two of the main motor symptoms during examinations [29].

Typically, the neurologists will evaluate the situation by recurring to motor symptom examinations scales like the Unified Parkinson's Disease Rating Scale (UPDRS) or the updated version the Movement Disorder Society-sponsored revision of the UPDRS (MDS-UPDRS), the Hoehn and Yahr scale method or even several technical imaging analysis [29].

- MDS-UPDRS

When in examination, the patient is asked to perform specific tasks as the doctor will assign scores for each of them as required and defined in the full MDS-UPDRS, the most renowned scale for evaluating disability and impairment. This scale has 4 parts. Each of them addressing different aspects and normal symptoms of the disease, as indicated in Table 2.2, rating 65 items with 48 questions for a 0 to 4 evaluation and 7 with yes/no responses [25], [30]–[33].

Table 2.2 - MDS-UPDRS constitution based on [32], [33].

Part	Description
Part I	non- motor experiences of daily living
Part II	motor experiences of daily living
Part III	motor examination
Part IV	motor complications

Part III presents a bigger influence on UPDRS score compared to the others and Part IV is not used on non-treated patients. The motor-UPDRS (Part III) scores between 0 and 108, where 0 represents a healthy patient and 108 for severe disabilities [34].

PD progression tracking studies that used UPDRS conclude that its course is not linear and with a variable rate of deterioration but quicker in early stages of the disease or in people with postural instability [25].

- Hoehn and Yahr Scale

Beyond MDS-UPDRS, there is the Hoehn and Yahr scale, which is commonly used to compare groups of patients and to keep up with a disease progression evaluation. It offers an assessment of stages 0 to 5 (Table 2.3), assigning an overall score to the patient based on the pathological progress [25], [30], [31]. The 0 stage will represent no disease manifestations and 5 to wheelchair bound or bedridden unless assisted [25]. Still, both methods require being performed by specialists with high experience and skill in order to state the patient psychological and some physical and conditions [34].

Table 2.3 - Hoehn and Yahr Scale description based on [35].

Scale	Description
0	No signs of disease
1	Unilateral disease
2	Bilateral disease (without balance impairment)
3	Mild to moderate bilateral disease (some postural instability)
4	Severe disability (still able to walk and stand unassisted)
5	Wheelchair-bound or bedridden

In the end, all these clinical scales are subjective and essentially rely on the experience of the specialist performing the patient evaluation. Consequently, this provides a high inter-rater variability among different neurologists or different medical centres an even change over time. Although diagnosis plays an important role in the disease progression, studies state that around 25% is incorrect, even when core motor symptoms as tremors are manifested [30], [31].

- Brain imaging and other tools

Brain imaging is not the usual procedure performed by neurologists or movement disorder specialists when they are considering a PD diagnosis. However, in order to perform a detailed neurologic examination, brain imaging is sometimes used to help support the diagnose. These are normally only applied to selected patients, due to their limitations, but an essential help when uncertain diagnosis occurs also providing a follow up to brain changes [29].

PD Imaging studies include Magnetic resonance imaging (MRI) for brain structure examination and DaTscan (approved by the Food and Drug Administration (FDA)) that allows dopamine function detection in the brain. Although not frequently used, there is also functional MRI (fMRI) - a specialized brain imaging examination for brain function - and Positron Emission Tomography (PET) for certain brain functions measurements [29].

Concerning what was stated before, PD diagnosis is generally supported by the time brain neurodegenerative process is already triggered. Nonetheless, since the disease begins to involve before symptoms manifestation turns noticeable, it would be important to optimize this pathology diagnose in early stages, allowing a better quality of life to patients [30].

As to guide for an accurate diagnosis, it is possible to associate all this clinical knowledge in order to perform an exhaustive study on the pathology, from olfactory and sleep exams to MRI [30], [31].

Many organizations are dedicated to PD research such as government institutions or supported laboratories, universities and even private research facilities and it was contributed as part of research and technology/pharmaceutical areas development [6]. These areas are focused from genetics to pathogenesis to early diagnosis markers, which is a major area of interest for biomedical investigation. A review of previous studies applying machine learning concepts for PD diagnosis can be found in Chapter 4.

2.4 – Treatments and Medication

Unfortunately, nowadays there is still no definite cure for Parkinson's Disease, however, there are several treatments that are focused on symptoms attenuation. Eventually, almost every patient will have the need to take medication for motor symptoms moderation and there is a large variety of medication. Although therapies based on Carbidopa/Levodopa use remain the most effective, each treatment is highly individualized and adjustable concerning side effect manifestation and symptoms development [25], [30], [36].

Levodopa therapy is the most common and probably the oldest pharmacological therapy applied to PD patients, and later introducing dopamine agonists. Hence consisting of a treatment providing a significant improvement on patient's quality of life while decreasing the mortality rate. Levodopa conversion to dopamine will act by compensating the absence of dopamine concentration on the brain [6]. However, this therapy presents several side effects, such as uncontrolled/involuntary movements (dyskinesias) and overall motor complications [6], [30].

These effects are commonly registered in young people, which contributes to a postponement on Levodopa based medication intake. Considering elder patients and their symptom range, the scenario changes, as motor complication risk is lower and safety profile of levodopa is better within those ages, which presents better results when medicated comparing to younger patients [6], [25], [30]. Medication effects can decrease as time pass which can lead to an ON-OFF phenomenon: ON meaning the PD patients in under medication and OFF when it is gone and in a waiting period for another dose (6 to 8h) leading to a bigger possibility of motor instability [30].

Although medication is the most requested treatment, another possibility is a combination between various factors and modalities. Some of them involving physical, occupational and speech therapy for a fuller treatment – DBS. This form of treatment consists of a surgical option implanting an electrode into a targeted area of the brain or complementary medicine/therapies for some symptoms and lifestyle changes for a better quality of life. Even so, each combination and full treatment is adapted and personalized for each patient [36].

On initial years, even with medication being a benefit, PD patients can live normal lives and perform daily activities. When the severity of the symptoms increases and medication alone is not effective enough and their life quality drops, this kind of side treatments as mentioned before (physical therapy, psychological support and so on) take a fundamental role in the process [6], [36]. At last, and an important step for future research is participating in Clinical Trials to help studying and revolutionize PD treatments as more information can be gathered [36].

2.5 – Final remarks

The concepts abroad in this chapter aimed mainly to facilitate the understanding of the disease studied in this project. Parkinson's Disease is a condition that affects many people worldwide and it is still a concern for the medical field due to the numerous adverse and disabling symptoms/effects and the fact that there is no cure. Although it is more common among elders, the main difficulty remains in an early diagnosis since the symptoms are manifold and appear with varying time intervals. Thus, clinical evaluation and the use of complementary tests can contribute as support for physicians, allowing to reach more accurate and timely conclusions about the disease.

Chapter 3

Principal Concepts of Machine Learning

Traditional medicine keeps expanding as engineering continues to step up into the medical field bringing them together to a brighter future. In modern healthcare system, there has been a quick expansion on computational complexity. This phenomenon occurs for more efficient treatment decisions in order to provide more treatment option and information exploring [37].

The foundation of an expert system lies on a solid data set, making it is crucial to develop data-driven models as a path to individualize medicine. By analysing previous cases and scenarios and constructing a system capable of analysing and learning from the available data, it will allow a focused treatment to fit certain conditions characteristics. Technology is becoming a closer reality in the medical field and nowadays, it is possible to refer to technology in several steps of medical treatment, pre-treatment of medical consultation, from the moment of disease diagnosis [37], [38].

This chapter is mostly focused on some concepts and deepening knowledge concerning expert systems and its historical overview, Machine Learning and data mining as an overall view of its mechanisms and its inclusion on the medical field. Moreover, a description of learning methods, features selection methods and classification algorithms are approached.

3.1 – Expert Systems

Considering a computer system able to behave and simulate a human expert decision-making ability is considering an expert system, capable of advising and explaining the logic and/or the relevant steps behind it. These systems use reasoning knowledge to analyse and solve complex problems. While developing these systems, researchers have realized that there are three basic problems with AI application such as knowledge representation, utilization and acquisition [39].

AI is currently considered an Engineering branch capable of finding solutions to complex challenges by implementing innovating concepts and solutions. Nonetheless, in order to influence their practical use into real-world application, it is important to develop capable environments for it. This would allow knowledge-based tools to evolve and involve several applications [40], [41]. Researcher working with AI in Biomedical informatics is slowing showing their work, gradually in key decision makers, trying to provide the necessary change for a full impact of their technology can be appreciated and applied [40]. The continuous technologic and electronic progress plus software programming g might result in a computer's intelligence comparison to human's, however, this will not compromise the important human capacities and contribution for AI development [41].

Opinions shift from knowledge-intensive to data-intensive systems application. Despite expert systems capacities and impressive performance, they were not yet fully deployed in clinical settings due to their hardware needs [42]. Modern healthcare system not always gets the advantage of all these systems capacities even though its rapidly expanding costs/complexity and all exploding information streams. This is a decline because these systems do not reach the front lines. Consequently, the ability to choose ideal treatment decisions is decreased [37]. Furthermore, nowadays health systems need to learn to take advantage of expert systems and focus on continuous implementation of process improvement. Although these systems can be used in many categories within an area, their programming purpose can involve many health application such as diagnosis [38], [41].

As machine learning concepts like neural network and statistical learning keep expanding, it turns out to be a research hotspot in AI. Consequently, machine learning application to expert systems could be the answer to possibly promote its performance. Nonetheless, the urge to push expert system in order to solve more complex problems is currently on demand [39]. Machine learning concept is approached in section 3.2.

3.2 – Machine Learning

People are overwhelmed with new information daily, making learning a constant process in human lives. The main goal of learning is retrieving the existing data and try to make sense of it, discover patterns and learn from it. However, learning is a complex concept that englobes a large range of processes, making it difficult to define. Nonetheless, Humans are innately curious beings who will continuously progress to try and understand or decipher new problems of all tiers of complexity, throughout their lives [40], [43]-[45].

Therefore, when trying to define Machine Learning, the same process occurs, recurring to AI, systems change and adapt to perform tasks. Based on these processes, expert systems evolve with task performance and can be used in fields like diagnosis, planning, predictions and many others [43], [44].

According to Mario Stefanelli et al. (2001), AI in medicine has been brought by the new millennium and medical evidence, values and resources of data should be the base of Clinical decisions. Nevertheless, by applying these base guidelines could be possible to improve patient care outcomes, reduce variability in practical cases and control associated costs [46].

Giving the amount of available data combined with technology evolution, a natural progression is happening in this field. This capacity of data collecting, and storing is mainly enabled by digital revolution, providing available data to analyse. According to many scientists, Machine Learning is identical to AI, since the intelligence characteristic of the systems allows a learning possibility [12], [47]. Moreover, AI application in many areas becomes a more common and useful procedure. Its main propose is computer system development with learning and adapting capability from its experience [40], [47].

The potential of Machine Learning algorithms was early approached by healthcare and medicine applications, essentially to medical dataset analysis. Patients records could be stored and analysed in order to apply the acquired knowledge to future cases. In order to do this, labelled input of patient records with known correct diagnosis is inserted into a computer system to run a learning algorithm. These systems can act as classifiers and used as a support for students or non-specialists to rely on future diagnose cases and as support opinion, where a physician is assisted in diagnosis to improve faster diagnosis with higher accuracy. An example of medical acquired information can be done by wearable health sensors [12], [40].

Furthermore, AI application in medicine can be divided into two main branches: virtual component, referring to Machine Learning, and physical component. The virtual component is mainly represented by the constantly improving mathematical algorithms that learn from experience [41]. Chronic mental diseases treatment can be referred to as a complex example of this kind of process applied to medicine in development [48]. The physical component consists of the application of medical devices and robots. These robots progressively becoming more sophisticated and being an important part of care delivery as *carebots* [41].

When referring to the tasks that perform changes in a system, it is the same as referring to machine learning applied to artificial intelligence systems [44]. Additionally, machine learning can also be referred concerning not only early detection and targeted prevention but also personalized medicine. This enables an aim for clinical value increasing and health costs decreasing [42], [49].

3.2.1 – Categories of Machine Learning Tasks

Machine learning algorithm can be categorized into many types. However, there are three types that are typically mentioned: supervised learning, unsupervised learning and reinforcement learning. Shortly, supervised relies on labelled previous cases for classification and prediction, unsupervised consists on the system capability to find pattern even with unlabelled data and, lastly, reinforcement for system interaction with dynamic environment [41], [47].

The following points present a brief description of these main types of learning, providing a better overall perception of learning algorithms taxonomy.

- Supervised Learning

In this type of learning, the system tries to find the best function (target function) that defines the relation between inputs and outputs by looking input-output training examples, giving an expression of a model describing the data. Given a training set of data, for a set of features, and considering the alternative functions, the main goal is to build a concise model of the problem that is capable of predicting the value of a new variable [47], [50].

Supervised learning presents also two main types of tasks: Classification models for class distinction prediction and Regression models for numerical values. Some common examples of these techniques are Decision trees (DT), k-Nearest Neighbours (kNN), Artificial Neural Networks (ANN) and Support Vector Machines (SVM) [47]. In Table 3.1, an example of a table with labelled data.

Table 3.1 - Example of a dataset with n features for each case and a label class associated [50].

Data in standard format				
Case	**Feature 1**	**...**	**Feature n**	**Class**
1	XXX		XX	A
2	XXX		X	B
...	...		...	...

- Unsupervised Learning

Oppositely to the previous case, unsupervised learning systems take data that does not have a corresponding label and trains with its instances. Therefore, the system tries to find associations and patterns within the data variables [47], [51].

Despite the different approaches to data from supervised to unsupervised, an unsupervised component such as a pre-training phase is involved in the best results obtained by supervised learning. However, there are still many questions concerning learning problems difficulties even though it is possible to train deep models applying these algorithms [51].

Furthermore, there is another approach of these categories that falls between the two types previously described, the semi-supervised learning. This tries to take the benefits from each one, using both labelled and unlabelled data. By this it means, for example, using a training set with and/or without input-output pairs [52].

- Reinforcement Learning

In this kind of learning system there is no prior knowledge about the behaviour of the environment, so, it attempts to learn through direct interaction from it to maximize the notion of cumulative reward. Therefore, it works on a trial and error mechanics, as through trial and failure [47], [53].

Data mining application runs tools for extracting regularities from data. Therefore, this technical basis is provided by machine learning and it is possible to define more and different type of algorithms that have been developing among time.

3.3 – Data Mining: learning from data

Machine learning provides a technical basis of data mining. So, as data acquisition and machine development increase, data search and analysis become an opportunity window for data mining implementation keeps growing [43]. Researchers and institutions need to keep promoting and increasing their use of intelligence processes as data is acquired in real time. Moreover, some researchers believe in implementing open access to data as a collaboration in the scientific and clinical community [41]. Data mining consists of the analysis of data available on a database as a solving problem technic, taking raw data, previously unknown yet potentially useful information, from the database. This process provides computer programming for pattern and regularities seeking trough out databases. Hence, besides good performance, explicit knowledge structures take consist of an important part of machine learning application to data mining, as experience in many fields and applications [41], [43].

The classical approach for data mining application to a database for a certain result can be summarized in a process pipeline, **Erro! A origem da referência não foi encontrada.**1. Thus, data will be processed from a pre-processing phase to feature extraction to feature selection algorithms followed by classification algorithms.

Feature extraction methodologies provide a signal analysis through the extraction of the most prominent features. These features are representative variables of the various classes of objects [54]. After the first two phases, Feature selection will reduce the data and computational complexity, by selecting relevant features as the comparison of diverse Classification algorithms will help to conclude the classifier that provides the best results [54]-[56]. Feature Selection and Classification algorithms will be explored in sections 3.5 and 3.6, respectively.

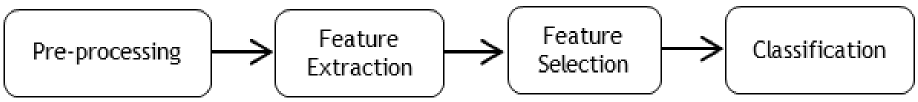

Figure 3.1 - Process Pipeline

- Data Mining and ethics

Concerning data application and its use, especially people's data, it presents a few implications around ethics. Although depending on the application finality, data mining techniques application must take responsibility by being aware of ethical issues. The main concern relies on data confidentiality and integrity protection as people must be informed about the utility before deciding to provide their personal information [43].

Hence, society standards over data acquisition and use, logical and scientific standards must also be followed in order to take information and conclusions from that data.

Consequently, in the whole process, data mining represents only a tool used by people who take conclusions and decide its knowledge application [43].

3.4 – Fielded applications: Diagnosis

Considering the system's capability of learning and its performance in several situations, there is a panoply of fielded applications for machine learning algorithms beyond the medical and healthcare fields. The main criteria are the emphasis and necessity of the ability to perform well when given new examples. Applications such as Decisions involving judgment (loans); Screening images like satellite images analysis for detecting oil slicks; Load forecasting, presenting a faster forecast prediction than done by humans; Marketing and sales which is one of the most active applications allowing companies to find the targets and other applications [43]. However, concerning the goals of the present thesis, the main field is its application in diagnosis.

Data can be complicated to manage as privacy-sensitive and heterogeneous when concerning data mining to medicine. However, this is probably a challenge that researchers outside the medical fields may not be aware of. Currently, as technology evolves more medical procedures employ imaging as a preferred diagnostic tool. Thus, the development of efficient data mining for its databases around medical exams such as imaging techniques or vital signals. This heterogeneity and variety of data require high efficiency, capacity and storage of the systems, for all the available data for each patient. Consequently, this provides a faster analysis of the data as a technician would take more time to process it [57], [58].

- Diagnosis

The application of Machine Learning has proved to be extremely useful even when compared to handcrafted rules applied in expert systems collecting performance. Thus, diagnosis is considered one of the main application areas of expert systems. This learning approach was investigated for this field once, performing tests, the system with expert rules results was slightly superior to the handcrafted ones. Still, expert systems are not only applied for diagnosis for their good performance but since its rules and learning capacity are approved by domain experts [43].

At Principal Investigator and Principal Research Scientist in the MIT Media Lab, Pratik Shah et al. (2017), gave a TED Talk on “How AI is making it easier to diagnose disease”. In the beginning, Shah reminds the audience that AI can perform amazing tasks and can be a great help for medical cases, especially diagnosis. As of today, expert physicians have to order expensive imaging techniques so later another expert diagnose the results, being not practical and time-consuming. However, with today's traditional AI it would be necessary to have thousands of previous data and a trained physician to analyse the same data and combine both to train a system. Nonetheless, on MIT Media Lab, a group of researchers is investing on unorthodox AI architectures for solving this problem in medical imaging and clinical trials for quicker, easier and better diagnosis that does not require as much training data [59].

3.5 – Feature Selection Algorithms

One of the main steps to problem-solving with machine learning is Feature Selection (FS). It consists of a pre-processing technique that is responsible for removing all the non-relevant data which can either be irrelevant or repeated information. This technique is described as data reduction step and its role is to filter the full dataset by selecting the most relevant features. It achieves this by compiling data in a balanced fashion so that the subset that it originates has the desired data size without overlooking its relevancy [34], [60].

However, its these is often subjective as it is directly correlated with the problem's final objective. As such, FS algorithms can follow 2 different paths: Manipulative approach where the algorithm selects, and orders features by their relevancy, and an Exclusion approach where the algorithm selectively excludes part of the dataset that finds least relevant, being the result a subset of the original values. The relevancy factor is, itself, a subjective matter as it is also linked with the author's objectives [34], [60].

FS greatly downscales the dataset which allows for faster interaction, learning and overall effectiveness. Data aspects or characteristics are often called "features", "attributes" or even "variables". These features can present themselves as discreet, continuous or nominal and have a significant toll on the algorithm's output. Conversely, irrelevant features do not as their values are randomly generated. It is important to note that, regarding features, redundancy is applicable when a feature's role can be fulfilled by another. Thus, the ultimate goal of FS is to create a viable subset of data that encompasses independent features while maintaining its relevancy that is needed for the learning process [51].

There are two types of approaches in FS: Forward Selection, starting with no variables and keeps adding them if it is one that can decrease the error the most; and Backward Selection which starts with all the variables and keeps removing them, one at a time, until any further removal increases the error significantly [61].

Nonetheless, it is important to mention the relationship between Feature Selection algorithms and classifiers (approached in the next Section 3.6). Feature selection process can be divided into three main forms: Filter, Wrapper and Embedded. In Table 3.2, the taxonomy of this process is resumed from each form, its strong points, disadvantages and examples [61], [62].

- Filter Methods: based on discriminating criteria that are relatively independent of classification, so the former acts as a filter to irrelevant features before the learning phase. These methods minimum redundancy-maximum relevance feature selection framework [61], [62].

- Wrapper Methods: the classifier is applied as a black box. This provides the subset to score features based on predictive power. The most widely studied in machine learning are the methods based on SVM [61], [62].

- Embedded: in this case, the FS algorithm is part of the classification algorithm. This kind of scheme is used by algorithms like Decision Trees or Neural Networks (NN). However, since the classification algorithm is called several times, it becomes a disadvantage computationally [61], [62].

Table 3.2 - Feature Selection Taxonomy based on [61].

	Filter		Wrapper		Embedded
Examples	Euclidian distance [58], T-test [61], Information gain [56]	Correlation-based feature selection (CFS), Markov blanket filter (MBF) [61], Fast correlation-based feature selection [61]	Sequential forward-selection (SFS) [63], Sequential backward elimination (SBE) [63], Beam search [61]	Simulated annealing, Randomized hill climbing, Genetic Algorithms, Estimation of distribution Algorithms,	Decision trees [61], Weighted naïve Bayes [61], Feature selection using the weight vector of SVM [64]
Pros	Fast Scalable Classifier Independency	Classifier Independence, Better computational complexity (comparing to wrapper methods)	Simple, Classifier interaction, Models feature Dependency, Less computationally intensive	Less prone to local optima, Interacts with the classifier, Dependent on models feature	Interacts with the classifier, Better computational complexity than wrapper methods, Models feature dependencies
Cons	Ignores feature dependencies, Ignores interaction with the classifier	Slower than univariate techniques, Less scalable than univariate techniques, Ignores interaction with the classifier	Risk of overfitting, More prone than randomized algorithms to getting stuck in a local optimum (greedy search), Classifier dependent selection	Computationally intensive, Classifier dependent selection, Higher risk of overfitting than deterministic algorithms	Classifier dependent selection

3.6 – Selection of Classification Algorithms

The main focus of classification techniques relies on information extraction from the dataset and convert it into an understandable structure for further use. Based on Table 3.2, it is possible to confirm that the selection of the classification algorithms is dependent on diverse factors [8], [65]. So, classification, being an important part of data mining is used as a form of predictive modelling [66], [67].

Due to the training sample finite size, not so relevant features can influence negatively the classification decisions. Since features can deeply depend on one another, this can influence the accuracy of Machine Learning classification models [64], [68].

In order to have a condensed guide of some learning algorithms, Table 3.3 presents examples of classifiers, as well as a brief description and its strong and weak points.

Table 3.3 - Classification Algorithms Taxonomy based on [50], [65].

Classification Algorithms	Description	Strong points	Weak Points
SVM	Mostly applied for pattern classification. Linear patterns are easily distinguishable and separated, contrary to non-linear [65], [69].	Better for multidimensional and continuous features; High Accuracy in general; High Classification Speed; Popular in text classification problems;	Requires large sample size; Slow learning process; Computationally expensive;
NN	Used to classify the feature space. Types of neural networks architectures, the most widely used for prediction is the multi-layer feed-forward neural network [65], [67]. Requires large sample size.	Better for multidimensional and continuous features; High Accuracy in general; High Classification Speed;	Impractical if irrelevant feature present; Slow learning process; Difficult in choosing the type network architecture;
Decision Trees	Description, categorization and simplification technique for the dataset. Classifies by repeatably parting the dataset in subsets and each leaf node represents a class label [61], [65].	High Classification Speed; Easy interpretation; It is not affected by non-linear relationships between parameters;	Complex; Possible subset duplication on different path;
Naïve Bayes	Goal: examples or samples classification to a class based on attribute value, by approximating class and attributes joint probability [65], [70]	Can use smaller dataset; Fast computationally with good performance; High Classification Speed; Classification performance improved by irrelevant features exclusion;	Information theoretically infeasible; Computationally complicated;

kNN	Learns by analogy and search a pattern within the k closest training samples of the unknown. The training samples stored in n-dimensional pattern space and each sample is a point in that [65], [71].	Fast Learning process (training); Easy implementation; Suitable for multimodal classes;	Slow Classification Speed; Memory limitation; Sensitive to data structure;

Estimating the accuracy for possible classifiers candidates to a certain problem is the simplest approach for deciding which algorithm will provide the highest accuracy. However, a proposed alternative to improved performance consists of combining classifiers. Nonetheless, the main interest is to provide precise, more certain and accurate system results [50].

3.7 – Credibility: Evaluating the system

Evaluating the system is a fundamental step for making real progress in data mining, as performance tests can be used to validate machine learning algorithms and models. Consequently, there are certain measuring factors that allow an evaluation of the classification performance. Thus, the researcher looks for a system with high accuracy and reliability, suitable for clinical applications, but with low computational cost [66], [67].

Measuring the performance of the classifier is normal when considering classification problems since the classifiers will predict the class of the instances. The main difficulty relies on finding a good classifier. For that, it is essential to perform the training phase with the most data possible. Nonetheless, only relevant features should be accounted for since the presence of redundant features will most likely interfere negatively on the accuracy of supervised classification models [43], [68]. However, considering limited data, it is possible to apply cross-validation where random sampling is picked. By choosing a fixed amount of folds, the data is parted and this is done, possibly, in a stratification way so the rest of the data can be used for testing [43].

Considering diagnosis problems, the classification is mostly done by taking two classes into perspective, as for example: parkinsonism or non-parkinsonism. Hence, the goal of such systems is to determine which instance or sample of data belong to each class, aiming for the maximum accuracy possible. Many performance metrics can be applied, however some of the most well know are: Accuracy, Sensitivity, Specificity and the Area Under the Receiver Operating Characteristics. These metrics are based on four outcomes based on the system's predictions compared to the actual results of the classification [43], [60], [64].

3.7.1. – Confusion Matrix

The Confusion Matrix provides the most intuitive way for representing classification rules by comparing two or more classes, based on the predictions made by the systems and the actual class of a certain instance of data. However, it is not a performance metrics by itself but it is based on four important outcomes namely: True Positives, False Positives, True Negatives and False Negatives, as presented in Table 3.4 [58], [72].

- True Positives (TP): Represent the number of samples labelled correctly, the prediction matches the actual class. Or the correct positive predictions [60].
- True Negatives (TN): Data samples with accurate predictions, which matches the actual class, but considered negative [73].
- False Positives (FP): Samples that were not a match between the prediction and its actual class. It is false due to the system miscaption [60].
- False Negatives (FN): Data samples labelled incorrectly as another class, the system predicted as negative but it is actually positive [73], [74].

Table 3.4 – Example of a Confusion Matrix representation.

		Actual Class	
		Positives (1)	Negatives (0)
Prediction	Positives (1)	TP	FP
	Negatives (0)	FN	TN

Briefly, this matrix provides an overall perception of the number of samples guessed correctly or incorrectly in a system, considering that all assumptions have the same weight over the total number of classifications. This comes to minimize the error rate. Other than this Confusion Matrix, there are other models like: Lift Charts and ROC Curves [58].

3.7.2. – Accuracy

Classification problems' accuracy represents the number or percentage of data instances predicted correctly by the model over the total instances and predictions made. This is presented by the sum of TP and TN (main diagonal of the Confusion Matrix) over the total number of classified samples (Table 3.5) [70], [72].

3.7.3. – Precision

A system's precision is based on the comparison between the actual positive and correctly classified instance (TP) with all the considered positive instances (TP + FP). This is calculated by TP over the sum of TP + FP (Table 3.5) [70], [72].

3.7.4. – Recall or Sensitivity

Sensitivity shows the number of predicted positive cases that were actually Positive. This is calculated by the proportion of TP over the sum of TP + FN (Table 3.5) [67], [75]. ´

3.7.5. – Specificity

Specificity, contrary to Sensitivity, presents the proportion between the samples that are actually negative to the overall negative samples. Which means it counts the ability to be negative when the condition is not present. This is calculated by TN over FP + TN (Table 3.5) [73].

3.7.6. – F1-score

F1-Score is a metric that ranges from 0 to 1 as the harmonic mean between recall and precision. Ideally, the model should present a value closer to 1 since it represents how precise and robust a model/classifier is [74].

3.7.7. – ROC curves

The Receiver Operating Characteristics or simply ROC curve is a predictive measurement that indicates the overall accuracy of a model. However, when concerning training set it shows less accuracy [67], [75]. Taking a model, the curve is represented by four main frequencies and cut-off value. These conditions englobe the four scenarios around predicting events or non-events, the actual outcome and if it matches the predictions. As for the graph itself, specificity represents the x-axis and sensitivity take the y-axis values. The curve is a result of the cut-off points [58]. This means that the metrics represented in ROC curves are:

- True Positive Rate (Sensitivity): TP / (TP + FN)
- False Positive Rate (Specificity): FP / (FP + TN) [73].

Taking these performance metrics into account and considering forward consultation, Table 3.5 presents a summary of the performance metrics mentioned and respective abbreviation and calculation method [57].

Table 3.5 - Performance metrics for evaluation.

Method	Abbreviation	Formula
Accuracy	ACC	$\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TN}+\mathrm{TP}+\mathrm{FN}+\mathrm{FP}}$
Precision	-	$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$
Recall/Sensitivity	-	$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$
Specificity	-	$\frac{\mathrm{TN}}{\mathrm{FP}+\mathrm{TN}}$
F1-Score		$\frac{2*(\mathrm{Precicion}*\mathrm{Recall})}{\mathrm{Precicion}+\ \mathrm{Recall}}$
Area of under the ROC curve	AUC	$\frac{1}{2}\left(\frac{TP}{TP+FN}+\frac{TN}{TN+FP}\right)$

3.8 – Final Remarks

Living in the digital era means easy access to data as to collect and store it, inserting information systems as a normal tool in daily life. As equipped hospitals can monitor and collect data from countless devices, it is possible to associate the medical field the use of expert systems as Machine Learning algorithms which were designed to analyse medical datasets. These expert systems consist of a process structured in several steps from feature selection until the results. Concerning minimum possible redundancy content, features must be properly selected to encode as much information as possible concerning the task of interest, considering the proximity measure as it'll quantify how the similarity or distance between feature vectors. As not always and not all features have the same impact on the analysis, clustering criterions are also an important part of the process, depending on the pretended sensibility.

Once the results of the clustering algorithm are ready, validating becomes an important step as the interpretation of these results for taking conclusions. Furthermore, a new patient's diagnosis can be quicker and improved in accuracy and reliability, following this simplified logic plus traditional medical analysis.

Chapter 4

Literature review: Machine learning technics applied to Parkinson's Disease Diagnosis

Expert systems and Machine Learning application in the medical fields rely on the optimization and evolution of medical procedures including diagnosis. Thus, concerning data analysis, medical or health records become essential tools for personalized medicine, early detection and targeted prevention [41]. Moreover, notorious difficulties are revealed concerning reliable PD diagnosis, reporting 25% of misdiagnosis cases [67].

This chapter goes over the overall importance of expert systems and machine learning application in medicine, focusing on Parkinson's disease. Many studies can be found concerning different features and methods that aimed to improve this condition's diagnosis or treatment, leading to a review in this chapter. Moreover, and to sum it up, a synopsis of the previous work developed before the dissertation, approaching data mining concepts.

4.1 – Parkinson Disease diagnosis applying Machine Learning

After some research, it was possible to gather some papers and studies concerning Machine Learning application to PD diagnosis. Therefore, an important step to the development of this dissertation was to analyse and learn from the previous study cases and to understand the methodology approach for the applied methods.

Diagnosis is the main focus when approaching the computational field around PD, being most available literature about its absence or presence and disease severity identification or feature extraction of handwritten exams. It is more common to use signal exams, turning diagnosis through handwriting exams rarer [76]. However, both forms will be approached.

Consequently, to technology development as expert systems, healthcare is changing from diagnosis to treatments. Future technological innovation such as the examples that will be presented in this chapter comes as a support for even high-skilled experts [77].

According to Scopus [10], when searching for papers concerning the keyword: Parkinson's disease + Diagnosis + Machine Learning, there was a 542 papers result. From these results, presented in Figure 4.1, it is possible to see that has been an increasing interest in PD diagnosis and Machine Learning area over the years. The main growing point starts in 2011, with over 60 papers published only in 2018 on Scopus concerning this theme. These results support the current concern around PD diagnosis and the overall interest of the development of this thesis.

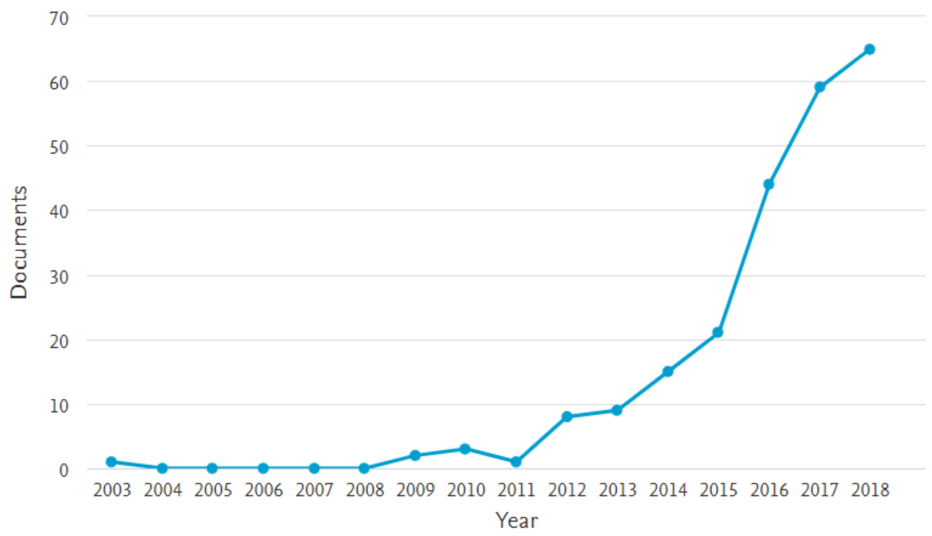

Figure 4.1 - Search results for Parkinson's Disease Diagnosis with Machine Learning from Scopus: number of published papers by year [10].

For analysis and overall perspective of a previous application of Machine Learning and Data mining technics, mentioned on Chapter 3, applied to PD diagnosis, a research process was made in the search engines: PubMed, Science Direct, Mendeley and Scopus, between October of 2018 and January of 2019. From the articles found initially, some were selected for this study based on several selection criteria such as giving diverse approach examples. The search descriptors and engines used are shown in Table 4.1.

Table 4.1 - Search expressions used to identify potential papers to review for previous tested PD diagnosis systems.

Search engines	Search Expressions	
Pubmed [78] Science Direct [79] Mendeley [80] Scopus [10]	Parkinson's Disease Diagnosis or Parkinson's Disease Detection	Machine Learning, Data Mining, Expert System, Computer-aided Diagnose

The selected articles refer to the applied algorithms and techniques, as well as some comparisons between different approaches for the same dataset, in order to provide an overall analysis and comprehension of the studies' pipelines. The main steps in interest for this analysis are presented in Figure 4.2.

Figure 4.2 - Steps in analysis for the state of art.

4.1.1 – Motor Symptoms or Non-motor Symptoms Approaches

- Postural instability and gait problems

Postural instability is one of the core motor-symptoms of PD, being one of the studies focus for many researchers. This reflects on the inability to maintain a straight and steady posture, preventing falling episodes.

Concerning 56 PD patients and 34 healthy volunteers group study, regarding postural sways due to the instability resultant of the disease severity. Even under standardized conditions, a large group of patients reveals in different stages of the disease. The database under analysis was composed by a 37-76-year group, ranging from 1 to 4 on the Hoehn and Yahr scale and that had a positive response to acute oral levodopa test. The protocol involved a computerized force plate in two moments: first, before levodopa dosing and the second moment for 1 and 2 hours after the intake of a fasting dose of levodopa (100 mg) and *benserazide* (25mg) [81].

For this posturographic examination protocol, patients would stand barefoot on the force plate, stay still with relaxed arms with eyes open or closed. The variables analysed were the anteroposterior and Lateral-lateral oscillation, and curve surface (area within the centre shift during the trial time). The results were analysed by an ANOVA statistical approach [81].

In a study, Tahir and Manap et al. (2012) focused on one of the main motor symptoms: gait problems owing to motor impairment provoked by PD [82], [83] The main goal was to compare the classification ability for gait pattern between ANN and SVM. First off, the pre-processing phase consisted of a two-normalization process: intra and intergroup normalization. The main features vectors in this gait study were: basic spatiotemporal, kinematic and kinetic [83].

Taking these features, feature vectors were implemented, and the diagnosis was evaluated in both classifiers. The authors affirm that ANN classifier application for gait problems in slow walking is widely used, presenting previous studies with 98+% accuracy and SVM as well with high accuracy systems for EEG analysis. So, they decided to compare both classifiers for gait pattern recognition. Taking a group of 12 PD patients and 22 healthies capable of walking without any external support were submitted to the study with 37 reflective markers traced by an infrared camera. The participants had to perform a routine of walking in a force plate. The best results, for both classifiers, were considering the extraction of the four features from the basic spatiotemporal on the intragroup normalization. SVM presented 98.2% accuracy compared to a 96.9% with ANN [83].

In a 2018 study, concerned about PD inhibition of controlled movement such as gait impairment and tremors, a group of Indian researches decided to apply machine learning algorithms for faster PD diagnosis. The main focus was on gait analysis: normal or abnormal since gait cycle can be separated in several phases. The dataset was provided by a public database of vertical ground reaction force (VGRF) by Physionet with 279 gait recordings from 93 PD and 73 healthy controls [84].

The data was acquired by 8 sensors underneath each foot displayed following (x,y) coordinates. The sample rate was of 100Hz and for pre-processing a Chebyshev type II high pass filter with a cut-off frequency 0.8 Hz was applied in MATLAB. Several features were studied as: kinetic features (heel and toe forces), temporal features (stance and swing phases) and spectral features (tremors). For classification, the system presented an average of 92.7% ACC, being the maximum of 94.8% obtained by the Medium Gaussian SVM classifier. Further performance metrics as ROC curves show an 85% true positive rate and 1% false positive rate, and AUC of 97% [84].

- Handwriting exams

Contemplating handwriting exams, it is possible to refer two studies that compared each other. Handwriting exams can be on paper, using digitizers or on a smartphone. These exams can be quite diverse, regarding the PD patient's capacity to trace forms: drawing spirals, ellipses, connected syllables, connected words, and more. However, information can be unclear on such exams, complicating the feature extraction and analysis [76].

The first study [85], created a new approach for diagnosis through handwriting exams, performed on paper and relies on underlining of the template is correct, and the second made a SCM-based approach, comparing the two. On the first one, researchers evaluated their approach using: Naïve Bayes (NB), Optimum-Path Forest (OPF), and SVM classifiers. The approach on [76], a Structural Cooccurrence Matrix (SCM) based approach, relying on similarity metrics as attributes from handwriting templates-patient handwrite comparison. This was applied to 3 databases: Meander, Spiral and both (M/S) [76], [85].

Considering the three options and comparing both approaches, using the same classifiers, the second study had better results. Presenting an 85.54% accuracy applying the proposed method for Spiral format exams with SVM classifier, Table 4.2 [76].

Table 4.2 - Performance measure for tested classifiers: Accuracy (%) on [76].

Method	Accuracy (%) Spiral format
NB	82.01 ± 5.53
OPF	75.71 ± 4.39
SVM	85.54 ± 3.62

- Voice and Speech Impairment

According to a 2010 research, approximately 90% of PD patients suffer from some form of vocal impairment, as one of its symptoms is dysphonia. This term aboard disorders concerning the voice as0 pathological or functional problem in it, such as hoarse sounding or effortful. This study approached 31 people being 23 PD patients that recorded 6 tracks of voice each for analysis [67].

After feature selection and data input partition, for computation time of preliminary modelling runs reduction, four independent classification schemas were applied: NN, DMneural, Regression and Decision Tree. The database was randomly parted by 65% for training and the rest for testing. The best result was a classification of 92.9% accuracy when applied the NN classifier, followed by 88.6% for Regression classifier, presented in Table 4.3. By using NN classifier, a few adjustments were made: as a feed-forward, single hidden layer neural network was used for the backpropagation learning algorithm with 10 neurons in the hidden layer. However, Decision tree produced better classification rate at the training stage [67].

Table 4.3 - Performance measure for tested classifiers: Accuracy (%) on [67].

Method	Classification testing (%)
NN	92.9
DMNeural	84.3
Regression	88.6
Decision tree	84.3

On another study, Ozcift et al. (2012) took a 31 people voice record database defined by measuring of 22 features, being 23 PD patients. Dysphonia measures were used, and each column of this dataset represented a particular measure attribute. This study applied a new SVM classification scheme, selecting features to train Rotation Forest (RF) ensemble classifiers. The aim was to present an improvement for PD diagnosis since *dysphonic* indicators consist of an important part of it with speech measurements [64].

The FS algorithms used a SVM strategy to make the number of attributes reduce in order to train the classifiers by finding powerful features from the 22, implemented in WEKA data mining environment. For training, six classifiers were selected and trained with the subset of features and, finally improved by RF ensemble classification strategy. The results evaluation was made recurring to three metrics: Classification Accuracy (ACC), Kappa Error (KE) and Area under the Receiver Operating Characteristic (ROC) Curve (AUC). The Classification performances of the algorithms are shown in Table 4.4, with RF improved classifications. Since the application of the RF ensembled improved the diagnosis in this study, the authors consider it presented pleasant results with an approximately 97% ACC of IBK, which is considerably high performance for PD diagnosis in their perspective [64].

Table 4.4 - Classification performances of the tested algorithms on [64].

Method	ACC (%)	KE	AUC
Multi Layer Perceptron (MLP)	90.8	0.75	0.91
RBF	88.71	0.68	0.89
LADTree	92.82	0.71	0.93
J48	92.3	0.63	0.92
KSTAR	96.41	0.86	0.97
IBK	96.93	0.89	0.97

Another study by Mandal and Sairam et al. (2013), was based on PD diagnosis by dysphonia detection. By telemonitoring applications, the authors proposed new methods for detection PD that were not implemented before in the medical field. This clinical trial took 6 months of collecting data from PD patients who already exhibit motor symptoms like tremors with medication absence during this period. It took some training to work with the device since it was made from their homes [86].

The study included 195 instances from 31 people, 23 with PD from 46 to 85 years. The collected data had 6 speech signals records per patient between 1–36 seconds. As the pre-process phase, noise was removed by Praat software. This study was carried out with a dataset leave-one-out k-fold cross-validation, in this case, 10- fold for training and testing the machine learning algorithms. A total of 12 algorithms were tested from: Linear LR, NN, SVM, SMO, Bayesian network, Ensemble selection, among others. These were performed under confidence level $p<0.001$ and corrected T-test with respect to Linear logistic regression, being evaluated from a total of 8 performance metrics. To deal with such quantity of multivariable, the authors refer to the importance of finding independent features, which was done by MATLAB analysis [86]. On Table 4.5, it is presented the best results by SVM and the average values of all algorithms.

Table 4.5 - Summary of classification performances of the tested algorithms on [86].

Algorithm	ACC (%)	Kappa value	Precision	F-value	AUC
SVM	97.65	0.92	0.95	0.93	0.96
Neural Networks	94.71	0.83	0.86	0.86	0.98
Bayesian Network	97.06	0.90	0.95	0.92	0.99
Average Values	97.060	0.896	0.956	0.913	0.979

Also, Khan, Westin and Dougherty et al. (2013) proceeded to a clinical trial for speech since in PD it is affected by respiration, phonation, articulation compromising its intelligibility. The used database had a total of 80 people with 60 PD patients. The routine provided 240 running speech samples in which 13 features were analysed, such as cepstral separation difference and Mel-frequency cepstral coefficient. The samples were classified according to the UPDRS-S from levels 0 to 4 [87].

Feature extraction consisted of speech acoustic features concerning measurements from phonatory symptoms, articulatory symptoms and prosodic symptoms. Since SVM is highly reliable for biomedical decision systems, this study trained SVM using final 20-fold cross-validation. Evaluating this system performance, it was possible to obtain an 85% accuracy in 3 levels of the dataset and proximally 91% for AUC-ROC in 2. Overall, the authors considered that SVM models present a robust classification capacity [87].

A more recent study concerning voice disorders analysis, researchers aimed once more to improve PD diagnosis by extracting feature sets of a recorded person's voice and testing multiple feature evaluation and classification methods. This study focused on 3 major objectives from the application of a proposed feature evaluation - Multiple Feature Evaluation Approach (MFEA) – to testing 5 different classifiers: DT, NB, NN, Random Forests (RF) and SVM before or after MFEA application, and final accuracy evaluation [88].

The feature selection outcome was feature vectors with 11 features compared to 23 features vectors in the original dataset. Hence, classifiers with 10-fold cross-validation were applied to both, raw and filtered data, providing better results on the second dataset. Table 4.6 provides an overall analysis of the classifiers for filtered data by 4 performance metrics: accuracy, precision, recall and AUC ROC. The highest accuracy was obtained by RF when applied to filtered data. Consequently, the study concludes that MFEA of the multi-agent system helps in finding the best feature set for classifiers' performance improvement [88].

Table 4.6 - Results of classification evaluation of filtered dataset on [88].

Classifier	Accuracy (%)	Precision	Recall	ROC
DT	96.954	0.970	0.970	0.996
NB	89.340	0.898	0.893	0.950
Neural Network	96.950	0.971	0.970	0.986
RF	99.492	0.995	0.995	1.000
SVM	95.431	0.957	0.954	0.982

- Non-motor Approaches

Concerned about the challenges in early PD diagnosis, a study approached Single-photon emission computed tomography (SPECT) images in order to analyse the distribution of the radiopharmaceutical in the brain and the loss of dopamine transporters. Recurring to a database consisted of 94 controls or healthies and 95 PD patient images, the main goal of this study was to identify these two group in an automatic classification system for ^{123}I-ioflupane images. This system was developed through the improved of partial least squares (PLS) approaches with SVM classifier [89].

The first step, pre-processing, consisted of a linear affine transformation and a non-linear warping using basis functions from SPM software. Feature extraction counted with the proposed PLS, a supervised approach. After vectors extraction, both training and test sets went through classification by a SVM classifier. SVM was used with kernel functions. For evaluation, three main performance metrics were calculated such as ACC, sensitivity and specificity. The results outperform previous studies with a 94.7% ACC, 93.7% sensitivity and 95.7% specificity, compared to 84.71% in the baseline approach [89].

On a similar study, an automatic diagnosis in DaTSCAN SPECT imaging system was approached, since ^{123}IFP-CIT images are used to provide dopamine transports density information. The database included 100 healthies and 108 PD patients giving a total of 208 images, that were later pre-processed and normalized with two proposed normalizations: normalization to the maximum and Integral normalization [90].

The classification is seen as a binary scenario to accomplish the final diagnosis and it chose to perform classification with 3 classifiers: SVM, kNN and nearest mean (NM). Further analysis of the classifiers recurs to ROC curves, specificity and sensitivity calculations. The experiment was performed considering 3 cases for the database: raw data, normalization to the maximum data and integral normalization data, forming 3 datasets. Then, for each dataset, the proposed methodology was implemented. From an overall appreciation of the system is possible to admit an 0.9681 (0.9641– 0.9722) value for AUC when data has normalization to the maximum and an estimative of 89% sensitivity and 93.21% specificity. Comparing all classifier, the most reliable values are obtained when trained by SVM with normalization to the maximum data, presenting a greater AUC value comparing to the other methods [90].

In a study from Spain, the authors questioned if it was possible to determine PD stages and its severity by a non-motor symptom score approach. This approach would combine two types of PD follow-ups: Hoehn & Yahr index and clinical impression of severity index for PD (CISI). The database was composed of 410 PD information of non-motor symptoms tests measurements: cognitive impairment, autonomic dysfunction, sleep disturbance or psychiatric complications. The participant's treatment englobed 81.02% taking levodopa and other antiparkinsonian drugs depending on the severity stage [91].

This model used a wrapper feature selection scheme, that predicted the class variables and 5 classifiers for 3 different severity sets: mild, moderate and severe. The ACC range in 72–92%. The classification was performed by NB, kNN, linear discriminant analysis (LDA), C4.5 decision trees (C4.5) and ANN. On Table 4.7, it is presented the best accuracy results in 3 scenarios: considering 3 classes, mild vs moderate and moderate vs severe, for each test HY or CISI [91].

Table 4.7 - Highest accuracy obtained from the 5 classifiers for HY and CISI [91].

Approach	3 classes		mild vs moderate		moderate vs severe	
	HY	CISI	HY	CISI	HY	CISI
ACC (%)	72.51	80.00	78.77	83.24	86.08	92.91
Classifier	kNN	kNN	NB / kNN	LDA	kNN C4.5	kNN

Moreover, the authors defend that this type of approach can be useful for PD detection on different stages of the disease from a limited set of symptoms [91].

4.1.2 – Combine Approach: Parkinson's Progression Markers Initiative

Parkinson's Progression Markers Initiative (PPMI) is an international, observational and multi-centre five-year study to identify PD progression biomarkers, consisting of a private-public partnership by the Michael J. Fox Foundation. This study aims to a better understanding of the disease and tries to provide the necessary tools for successful PD disease-modifying therapeutic trials. Moreover, it aims for researching communities to standardized acquisition protocols, transfer and analysis of clinical and imaging data; investigate existing novel clinical and do comparisons between PD patients and healthy controls for progression, and, finally, optimize preliminary verification studies on promising biological markers. This study combined data from biomarkers, imaging, genetic data, and others [92].

Considering a study about an automatic classification and prediction/prognostic models for early PD in 2013, striatal binding ratio (SBR) values from the PPMI database were analysed. These values are calculated from the ^{123}I-Ioflupane SPECT scans since it was shown the early detection capacity of dopaminergic neuroimaging imaging by SPECT. The database was downloaded in March 2013 and consisted of observations and scans in a time span of 12 months, in 3-month intervals, with a total of 179 healthies and 369 PD Patients. The models under study as diagnostic tools were: SVM and multivariate logistic regression (MLR), both widely used when approaching neuroimaging studies [93].

In this study, only 4 features were used: the striatal regions (left and right caudate, left and right putamen) from the SBR. Later, it was possible to do a comparison between SVM with radial basis function (RBF) kernel and SVM linear Kernel, with 10-fold cross-validation. The best results were obtained with the SVM-RBF with 96.14% accuracy, Table 4.8 [93].

Table 4.8 - Performances measured of the tested classifiers on [93].

Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)
SVM RBF	96.14	96.55	95.03
SVM Linear	92.28	95.33	83.98

Furthermore, the authors conclude that SVM classifiers present a high capacity for early diagnosis application whereas logistic model would be more indicated for estimating the risk of PD. Hence, this kind of models deserves a place in clinical environment for PD diagnosis [93].

A few years later, with 183 healthy normal and 401 early PD subjects from a Parkinson's Progression Markers Initiative (PPMI) database, a group of researchers used non-motor features of RBD and olfactory loss, plus other significant biomarkers such as *Cerebrospinal* fluid (CSF) measurements and dopaminergic imaging markers using Single Photon Emission Computed Tomography (SPECT). Taking these features, a Statistical Analysis of features was made using the Wilcoxon rank sum test, performed by MATLAB statistics toolbox. The classification was made by several classifiers: NB, SVM, Boosted Trees and RF. Concerning the choice of different classifiers, the authors remind that there's no one algorithm that's always better than other [8].

In order to find the classifier that was capable to provide the best results, their performances were shown by their ROC plots. Even though all classifiers had a considerably good performance, SVM provided a better performance with: 96.40% accuracy, 98.88% area under the ROC (AUC), 97.03% sensitivity and 95.01% specificity, Table 4.9. Although, in previous cases, the use of CSF measurements provided a low diagnostic utility, that inspired this study to combine non-motor, CSF and dopaminergic imaging measurements that could potentially discriminate early PD from healthy normal and may help in its diagnosis [8].

Table 4.9 - Performance accuracy for training and testing classification on [8].

Method	Classification training (%)	Classification testing (%)
NB	94.67 ± 0.59	93.12 ± 1.49
Logistic Regression	96.50 ± 0.60	95.63 ± 1.21
Boosted Trees	100 ± 0	95.08 ± 1.26
SVM	97.14 ± 0.45	96.40 ± 1.08

4.2 – Final remarks

Since technological evolution, computers became not only a tool for collecting and storing data on patient illness, treatments and outcomes but also a strong component of the analysis process in healthcare. The examples described along this chapter show a variety of PD factor that can be considered and analysed in order to train systems capable of improving PD diagnosis as the world knows it now [77], [94]. Machine learning and Data Mining have been used in several applications. In the medical area, it has been carried out in the last years a high development of the same ones for diagnosis diseases.

Throughout this state of the art, it is verified that several reviews of the literature have been carried out over the years. Numerous works are now available based on different algorithms applied to PD diagnosis, from postural instability, to voice tremors to biomarkers analysis, that allow analysing this problem and assisting health professionals. One main similarity within the previous examples is the current resort to the SVM classifier, being often mentioned as an excellent approach for several biomedical and diagnose scenarios. This relies on its ability to regularize global optimality in the training algorithm and for having excellent data-dependent generalization bounds to model non-linear relationships.

The outgoing search for new approaches is reflected in the number of articles described and existing for the subject studied. Relatively to the studies analysed for PD diagnosis, it is possible to conclude that different approaches required different ways to abroad the problem and, consequently, the application of different algorithms within Machine Learning. Even on the same study, researchers invest in testing diverse feature selection algorithms and classifiers searching for the best performance and results as high accuracy systems rise.

Chapter 5

Database characterization and Methodology for Parkinson's Disease Detection

This chapter is dedicated to the development steps of the methodology for PD detection proposed in this dissertation, as well as important aspects concerning the dataset that was used. Since this dissertation was based and tested in a publicly available study dataset, a good understanding of the acquisition conditions and protocol was fundamental.

Throughout the sections, it is possible to find a detailed overview of the dataset and its constitution as well as some pre-processing analysis of the data. The implementation pipeline is described across the chapter. The main steps carried out through the proposed methodology process are presented until the classification phase and performance evaluation.

5.1 - Dataset Characterization

In order to study PD detection for this dissertation, data from a public dataset - Parkinson's Progression Markers Initiative (PPMI) dataset - was used in this dissertation. This dataset concerns a study conducted by Michael J. Fox Foundation from planned trials dated between April 1st of 2010 and September 30th of 2023. This dataset was acquired to identify clinical, imaging and biomarkers of PD progression for clinical trials aspects and further therapies from a multi-centre study to asses progression of clinical features [95]. Moreover, it has already been used in many studies as presented in Chapter 4 Section 4.1 and it provides to develop a preliminary verification study around different stages of PD in a wide range of people.

This dataset concerns around proximally 2000 subjects, male and female age 18 years or older, divided into different groups according to their status as visible in Table 5.1. Hence, this study contemplates around 400 PD patients that participated in the trials with an average age of 45 and a Hoehn & Yahr score of stage I or II at Baseline.

Table 5.1 - PPMI Dataset Subject Characterization [95].

Number of Subjects	Subject Characterization	Main Eligibility Criteria	
400	Parkinson Disease (PD)	Includes: • Patients with at least 2 symptoms such as tremor, bradykinesia, rigidity, etc. • Diagnosis for 2 years • Hoehn and Yahr stage I or II at Baseline. • Non expecting medication within 6 months from Baseline • Male and Female age 30 years or older	Excludes: • Currently taking levodopa, dopamine agonists, MAO-B inhibitors, amantadine or another PD medication.
70	SWEDD	People with PD symptoms but without evidence of dopaminergic deficit	
200	Healthy controls (HC)	Includes: • Male and Female age 30 years or older	Excludes: • Current or active clinically significant neurological disorder • First degree relative with idiopathic PD • MoCA score < 26.
100	Prodromal	All subjects must demonstrate hyposmia and/or RBD, plus must be eligible based on DaTSCAN assessment by imaging core.	
600	Genetic Cohort PD subjects (300 PD patient with mutation (LRRK2, GBA, or SNCA) and 300 unaffected/first degree individuals with those mutations)	Includes (PD): • Patients with at least 2 symptoms such as tremor, bradykinesia, rigidity, etc. • Diagnosis for 7 years • Hoehn and Yahr stage < 4 at Baseline. • Male or female age 18 years or older. Includes (unaffected): • Male or female age 45 years or older with a LRRK2 or GBA mutation	Excludes (PD): • Current treatment with anticoagulants Excludes (unaffected): • Clinical diagnosis of PD
600	Genetic registry PD subjects (300 PD	Includes:	

patient with mutation (LRRK2, GBA, or SNCA) and 300 unaffected/first degree individuals with those mutations)	• Male or female age 18 years or older. • Individual with a LRRK2, GBA or SNCA mutation and/or a first degree relative with a LRRK2, GBA or SNCA mutation

Once registered into the Michael J. Fox Foundation website, the PPMI dataset is reserved and provided a 109 files compilation with all the clinical data available for researchers to download at http://www.ppmi-info.org. From clinical exams to motor and non-motor symptoms, imaging, MOCA tests, medication registration, it was imperative to organize the files with all the data and proceed to the data pre-processing phase.

5.2 – Dataset Organization and Analysis

For the preparation phase, in this specific study, the focus was on organizing all the data and select the PPMI Medical Records within the dataset. Hence, the first challenge was to take on proximally 1000 “features” and organize the medical data that could be used for determining PD. Therefore, giving all the data, category segmentation was considered a crucial first approach. This step consisted of organizing the data into 7 different categories that had similar information, as presented in Table 5.2.

Table 5.2 - Data aggregation into categories.

	Dataset category organization
1	Subject Characteristics
2	Biospecimen
3	Enrolment
4	Imaging
5	Medical History
6	Motor Assessments
7	Non-motor Assessments

All categories cover multiple types of information, for example, the non-Motor Assessments consists in all the measurements that are usually looked at to measure disease progression as the Hoehn and Yahr scale (Table 2.3, Chapter 2). In light of that information, a selection of information could be processed. However, there was an exception concerning the last Imaging Data Results for various reasons, one being imaging was done on too few patients and it would be another field of research concerning the focus of this dissertation.

In this case, the implementation recurred to *Jupyter Notebook (in Anaconda)*, *Python* [96] algorithms were developed concerning the steps presented in the implementation pipeline in Figure 5.1. Accessing the files and concerning a broad set of features at this point, one of the goals was to explore more intuitive and meaningful features for the detection of PD manifestation signals. This pipeline already discriminates the methods selected, which were based on the most common algorithms found in the literature review.

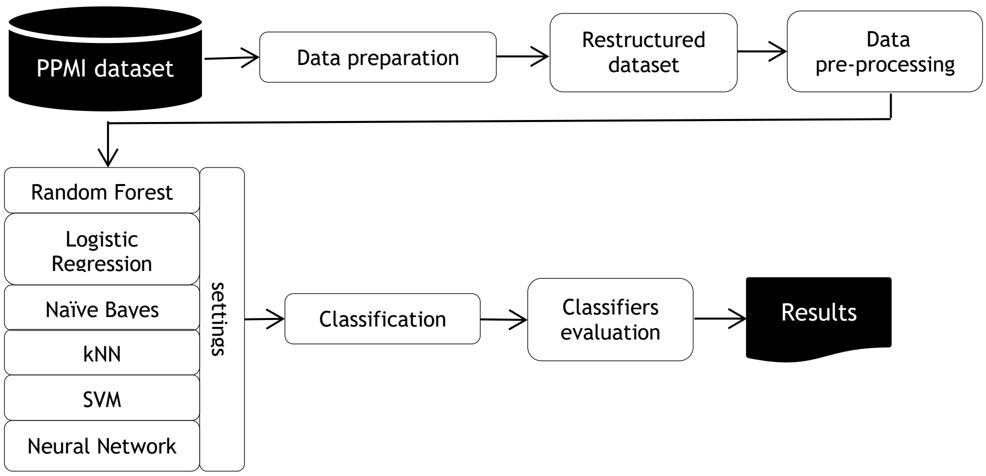

Figure 5.1 - Pipeline of the research activities on this study based on [88].

Taking the files, the goal consisted of gathering all the data in one *DataFrame* with all the information available of each patient along with the study and define the *labels* for the *targets* for the classification phase. The *DataFrame* is provided by the *pandas* class in Python, allowing to implement a two-dimensional size-mutable tabular data structure with labelled axes. This labelling provides a clearer view of the features and the data as an all.

Once focusing on the PPMI Medical Records, a reduction/selection of features from these dataset categories was performed in order to provide easier access and support for further manipulation of the data in upcoming implementation steps. After removing void or incomplete data for all patients, which would conditionate and complicate data processing, it resulted in 370 features. Some examples of the information (features) selected are:

- Motor Assessments: Tremors, Rigidity, Bradykinesia, Gait problems, UPDRS
- Non-Motor Assessments: Olfactory dysfunction (University of Pennsylvania Smell Identification Test (UPSIT)), Rapid Eye Movement, Sleep disorder, Cognitive assessments
- Medication: Type of medication, Dose, Interval between doses

All the acquired measurements are based on the PPMI protocol available on http://www.ppmi-info.org/study-design/research-documents-and-sops/, where all the examination steps for each extracted features of this dataset can be consulted.

5.2.1 – Data Pre-processing

After the data organization, it was important to implement a pre-processing phase, taking the estimators restructured dataset. This allows changing raw feature vectors into a more suitable representation for further processing and implementation phases. This step was developed by applying the *preprocessing* package, class *StandardScaler*, from *Scikit-learn* tool.

The *Standard Scaler* provides a feature standardization by removing the mean and scaling to unit variance. Each feature is independently centred and scaled by computing the relevant statistics in the training set sample, storing mean and standard deviation to the transform method. This type of standardization is a common step in this process since it can help improving estimator's performance [97].

5.3 - Classification

This section presents the main steps that are part of the classification procedures, along with a brief explanation of the main associated concepts. In order to find the best classifiers, several were tested considering different possible combination and variations of its parameters. After having the dataset narrowed, the plan consisted of testing the classifiers presented on sub-section 5.3.2.

5.3.1 - Training and Test Procedures

Considering the numerous features of the dataset, it is crucial to divide it into training data and test data. It was implemented a random split into training and test sets that can be quickly computed. This split between data is performed in a way that the test set covers 30% of the total dataset. This division was inspired by previous reports as well as the application of 10-fold cross-validation [98]. Cross-Validation is the most common method that allows us to compare models with different parameters. In this method, the data is split into K subsets of the same size where a different one is selected for testing and the remainder for training in each iteration (total of K times). Figure 5.2 represents an overall pipeline for this part of the implementation.

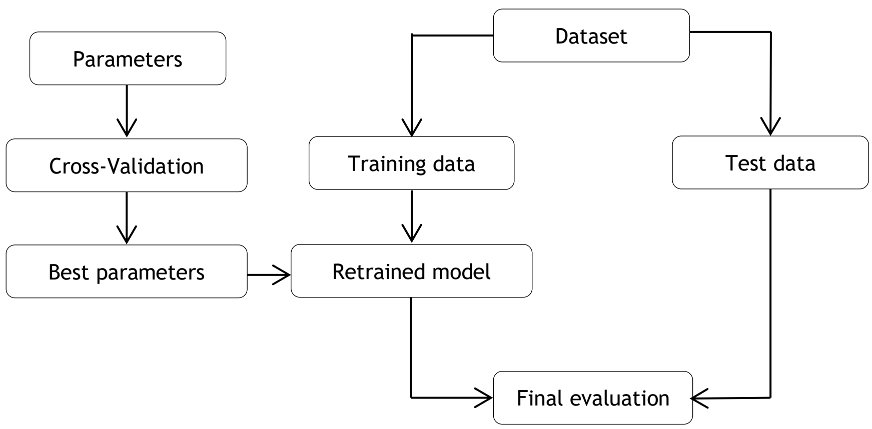

Figure 5.2 - Flowchart of typical cross validation workflow in model training based on [98].

5.3.2 – Classifiers

As for classification, the idea consists of implementing diverse classification methods until maximum possible optimization. Taking previous case studies/reports as an example for classification algorithm selection, a diverse implementation will allow comparing models in order to conclude which will provide the best results. Classifiers are also addressed in Chapter 3, Section 3.6 (Table 3.3).

Since having a well-documented data dictionary, feature selection was based analysing encoded columns names and their definition by what is extracted in the PPMI study. Some examples would be the time since the first PD diagnosis, calculations and padding for patients with very few visits, UPDRS results, so it was possible to have a cohesive data frame with many features to train the classifiers. The methodology implemented follows supervised learning guidance which proposed the use of labels. In this case, the labels used consist on the H&Y scale results, allowing classification of PD stage of each patient as a diagnosis.

Having prior reports, as examples, using model-free machine-learning techniques to diagnose PD, encourage to select models like Random Forest, Support Vector Machines and Neural Networks but also model-based like Logistic Regression (LR). That way it is possible to test simpler classifiers like NB and LR to advanced classifiers like SVM and RF [8], [99]. These classifiers were implemented recurring to a Machine Learning tool in Python *Scikit-learn*, which provides simple yet efficient tools for data mining and data analysis.

- **Random Forest**

Non-parametric algorithm and flexible supervised machine learning method. It consists on an ensemble method built on decision trees randomly restricted as it tries to solve overfitting problems from decision trees and provide accurate/stable results. This is possible due to its capability of aggregating the results of an ensemble of simpler estimators [88], [100].

Due to the simpler estimators' ensemble, increasing the number of splits performed can tend to overfit the data by the decision tree. Hence, RF combines them in order to reduce it by fitting many dataset sub-samples and the averages results for a better classification [100].

- **Logistic Regression**

One of the most frequently used model-based methods, however, besides the name "regression" is more like a linear model for classification (linear regression). Hence, by using a logistics function, the model can describe possible outcomes of single trials and can fit binary, One-vs-Rest, or multinomial logistic regression [99], [101]. It measures the relationship between a response variable and independent variables [102]

- **K-Nearest Neighbors**

The third model is one of the most well-known supervised learning algorithms in pattern classification, kNN, being K=1 its simplest form. So, from its instance-based learning, it will simply store instances of the training data and not construct a general internal model. This method assigns to each query a class represented by the majority label of its kNN in the training by retaining the training set during learning process [65], [71]. ´

KNN presents several advantages, such as simplicity, effectiveness and competitive classification performance. Learns by analogy and search a pattern within the k closest training samples of the unknown. The training samples stored in n-dimensional pattern space and each sample is a point. In order to optimize the model, the K value is extremely important but yet highly data-dependent: bigger value means less distinct classification boundaries but can suppress noise [65], [71], [103].

- **Naïve Bayes**

Naive Bayer algorithm simplifies learning. This algorithm takes the Bayer formula to estimate which instance belongs to which class. By the feature values in that instance the probability of each class is calculated and allocated to the class with the highest probability. Its goal is to match classification examples or samples to a class based on attribute value, by approximating class and attributes joint probability [65], [70].

This model also differs from others since it does a feature independently evaluation from other features [88].

- **Support Vector Machine**

SVM can be referred to linear and polynomials case and be based on splines, radial basis function networks and multilayer perceptron. SVM solution does not depend directly on the dimensionality of the input space and it is mostly applied for pattern classification. Linear patterns are easily distinguishable and separated, contrary to non-linear [65], [69].

In order to have generalized a separable gap and identify the classification, SVM plots the training cases as points in space. In other words, it defines a hyperplane that has the largest distance to the nearest training data point of any class. [104].

- **Neural Network**

In this model, knowledge is automatically acquired by the learning algorithm producing its own implicit rules and acquires knowledge automatically through learning instances [39]. It consists of network nodes of several linked artificial neurons by interconnections and perform linear or nonlinear computational operations and requires a large sample size [88]. Types of neural networks architectures, the most widely used for prediction is the multi-layer feed-forward neural network [65], [67].

Briefly, focusing the implementation of the classification phase it is possible to divide it into two parts, as demonstrated in Figure 5.3: 1) consisting in an in-depth parameter search implementation using the *GridSearchCV* function available in *sklearn.modelselection* Python module and compare to the results obtained by default values for each proposed classifier.

It is important to notice that hyper-parameters are parameters that are not directly learnt within estimators and are passed as arguments to the constructor of the estimator classes in *Scikit-learn*. And 2) classification 2nd trial/approach by implementing function available in the *sklearn.feature_selection* class for further improvement and comparing the differences. The FS phase is based on wrapper methods, providing the subset to score features based on predictive power. A backward selection was implemented since classifiers started by considering all features, then excludes them 1 by 1 until only the most relevant features remain.

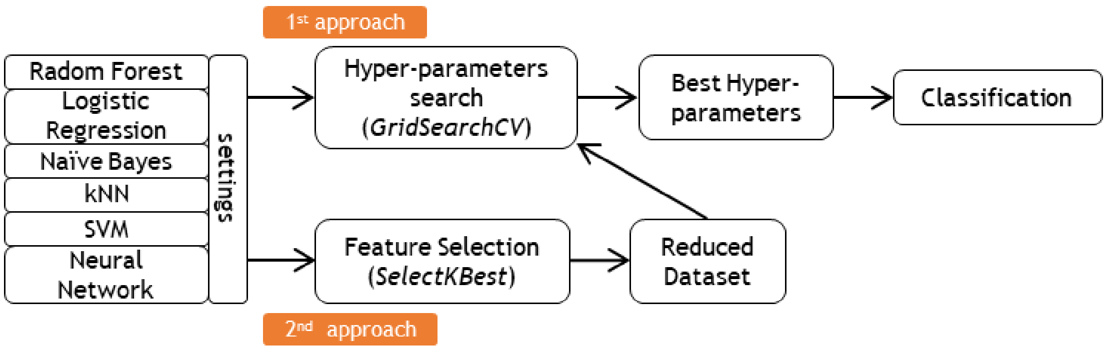

Figure 5.3 - Detailed pipeline of the methodology based the scheme in Figure 5.1.

Considering the classifiers presented on Sub-section 5.3.2, a set of specific parameters were tested and finally, the best parameters values for an estimator are found. Table 5.3 also presents the list of classifiers and respective parameters, which were optimized by a 10-fold cross validation grid-search - best results from the set of parameters tested carried on to evaluation. This step is performed by performance metrics.

Table 5.3 - List of classifiers and parameters tested.

Classifier	Hyperparameters	Range
RF	*n_estimators* *max_depth* max_features	{50,100,150} [1,10] {5,10,15,20,25,30}
LR	*C* *solver* *multi_class*	[10^{-3}, 10^{6}] {*'liblinear', 'saga', 'newton-cg', 'sag', 'lbfgs'*} {*'ovr', 'auto'*}
NB	*var_smoothing*	[10^{-1}, 10^{6}]
kNN	*n_neighbors*	[1,30]
Linear SVM	*C*	[10^{-6},10^{6}]
SVM (RBF)	*C* Gamma *(γ)*	[0.1,2] [10^{-9},1]
NN	*activation* *solver* shuffle and early_stopping	{*'logistic', 'relu'*} {*'adam', 'lbfgs'*} {False, True}

5.4 – Performance Evaluation

At last, as a validation of the implemented method/algorithm, it is important to implement an evaluation recurring to performance metrics. The algorithms of PD detection aim at the optimization of the methods. All the performance metrics used are described in Chapter 3, Section 3.7: Confusion Matrix, Sensibility, Recall and F1-score.

5.5 – Final remarks

This chapter provides an overview of the methodology implemented in this dissertation. Once having all the data prepared for processing, it was possible to gather all the information of each patient and prepare the labels for the implementation of two supervised machine learning classification approaches in Python.

Considering 6 selected models: RF, LR, kNN, NB, SVM and NN, the first approach relies on a fewer steps process on handling data, taking all available features for classification. On the second approach, a feature selection function is implemented in order to explore better classification results by considering only the most relevant features. In both cases, a hyperparameter search is considered for further model optimization. Finally, performance evaluation should be done by some selected metrics for model validation.

Chapter 6

Results and Discussion

This chapter presents the outcomes following the methodology described in Chapter 5 as the results of the main phases of the implemented algorithm. The results are divided essentially into two sections as for each approach of classification, presenting its results. Lastly, the algorithm's performance evaluation, comparing the machine learning algorithms described in Section 5.3. Throughout the presentation of the result, they are also discussed.

Concerning the choice of different classifiers and testing all of them, it is important to remember that there is not one algorithm that is always better than the other. Since both model-based and model-free techniques may be employed for the prediction of specific clinical outcomes or diagnostic phenotypes. Hence, in this dissertation, the goal was to implement the classification phase with different methods, testing with a different set of classifiers and choose the best one to predict PD. Sections 6.1 and 6.2 performs the analysis of the results concerning the classification step. A comparison between the performance of the several models is performed, considering the search for the best parameters of each classifier. However greater focus is given to the classifier with the best results, as presented in section 6.3.

6.1 – Full data classification results

As a first approach or trial, it consisted on taking the entire dataset as each model was trained with the default values of its parameters in order to provide a base idea of the process and consequently a support for comparisons among the models with tested parameters. This way, it would be possible to compare the improvement and conclude if the results got better. Table 6.1 presents the classification results for training and testing with the default values.

Table 6.1 - Training and Test classification results with classifiers' default values.

Method	Classification training (%)	Classification testing (%)
RF	99.3	82.9
LR	94.8	81.1
NB	67.6	60.6
KNN	80.5	76.5
SVM (Linear)	64.9	67.5
SVM (RFB)	100	69.8
NN	81.9	77,5

In order to optimize the process, as stated in the proposed methodology in Chapter 5, a hyperparameter search was made for the different models. Taking the parameters presented in Table 5.3 (Chapter 5), Table 6.2 presents the highest classification results accomplished in this study. Discrimination of the parameters and its values for each classifier will be described.

From an overall look, comparing Tables 6.1 and 6.2, the proposed approach reveals higher scores for the test classification and slightly lower values of training in two models. However, it was possible to reduce some overfitting by adjusting the parameters to the final ones tested. For each classifier, a parameter search and several combinations between them were experienced until the final results.

Table 6.2 - Classification results (train and test) from *GridSearchCV* best parameters search.

Method	Classification training (%)	Classification testing (%)
RF	99.1	87.7
LR	95.5	84.3
NB	71.4	69.2
KNN	78.4	77.5
SVM (Linear)	93.4	82.3
SVM (RFB)	87.4	78.4
NN	84.7	80.5

Generally, the classification of the seven models was noticeably improved when implementing the restructured dataset and final parameters values, raising the classification by an average of 5% proximally. The biggest improvement happens with SVM Linear and RBF, improving 14.8 and 17.6, respectively. Hence, RF had the lowest difference between classification results, proximally 0.6%, which could mean it default values are already well optimized.

RF turn out as the top classifier with the highest test classification results of 87.7%, which is one of the advanced models tested. For these results, the parameter search resumed on testing 3 main parameters from the *GridSearchCV* module, concluding with a n_estimators=150, max_depth=10 and max_features=30. This outcome was expected since the literature presents also great results once implementing this model.

LR returned a proximally 84% classification accuracy, being the second highest score, almost tied with Linear SVM. LR models are discriminative models for classification that produces linear decision boundaries even being less flexible. However, it is a model-based and one of the simpler classifiers tested, comparing to the RF.

Figure 6.1 presents a heatmap plot of the hyperparameter C and Solver values search of this model. A combination of C=0.1 with Solver=*'libilinear'* and C=0.07 with solver='newton-cg' were tested, in a smaller and more specific test, giving the best results with the second approach, considering a balanced class weight.

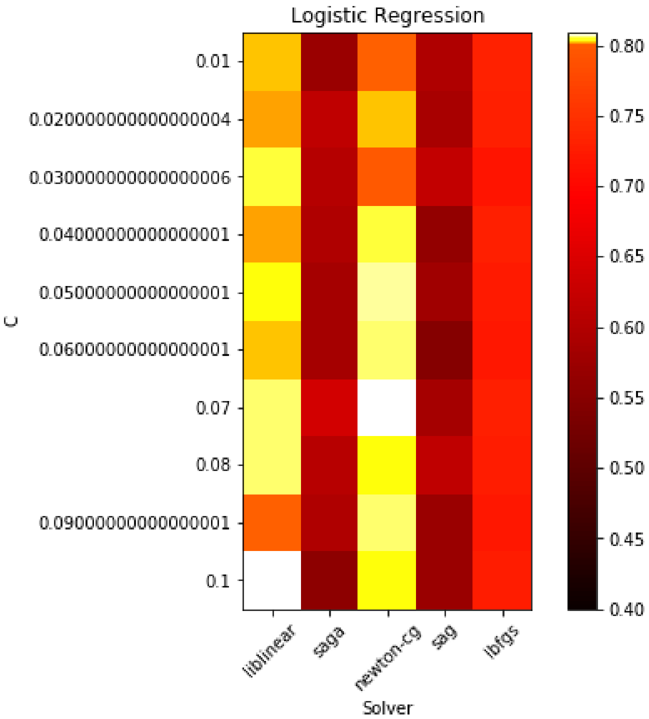

Figure 6.1 - Heatmap plot of LR parameters.

Like LR, NB is also a simple classifier. However, in this case, it provided the classification with the lowest score: 69.2%. Since having a small number of parameters, it is more difficult to adjust to the problem. Despite being considered a simple yet powerful classifier for real-world problems, this model often suffers from a high bias, which in this case did not turn it competitive against the other algorithms based on its scores.

For the kNN, only one parameter was considered to optimize its performance: K. According to the study in [106], K value is normally around 10, nevertheless, regarding different applications, choosing the same optimal k becomes almost impossible. Since it is not an exact criterion, taking once more the *GridSearchCV* module, a search for the optimal value of K was made. The tested values comprehended a wide window of K values from 1 till 30, progressively increasing 5 neighbouring points. The best score was accomplished with the optimal value of

K=20, as shown on the graph of Figure 6.2. However, considering the values tested, the score did present a big variance, being the lowest score of 67.2% when K=1 and the best score of 77.5% when K=20.

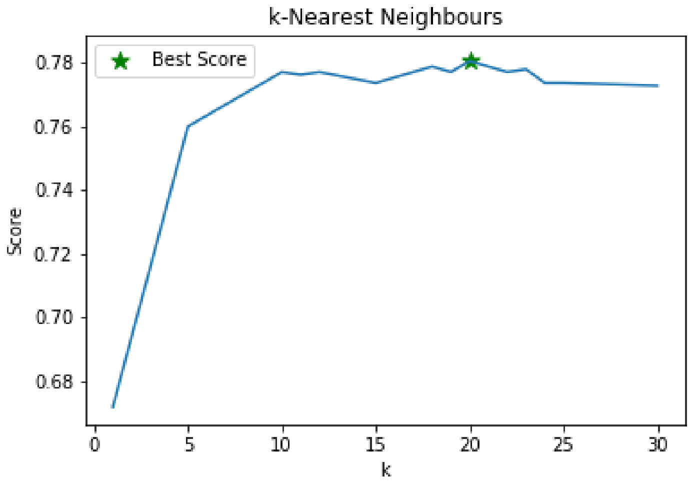

Figure 6.2 - kNN plot of training score for different values of K.

In similarity to RF, SVM is also an advanced and model-free algorithm, which means they both rely on some trial-and-error experience for the classification. Hence, considering the linear kernel, provided the second-best score. Both Linear and RBF kernel were tested, along parameter C and also Gamma for RBF.

The highest score was from Linear SVM corresponding to 82.3% with C= 0.012 as present in Figure 6.3. C value controls the influence of each individual support vector and as this value increases, the less regularization is used in the model, decreasing the cost of misclassification. As seen in the graph the score decreases when C increase. Even though it is not a faraway score from RF, SVM is known as a computationally expensive algorithm which is a weaker point but that presents a high accuracy in general.

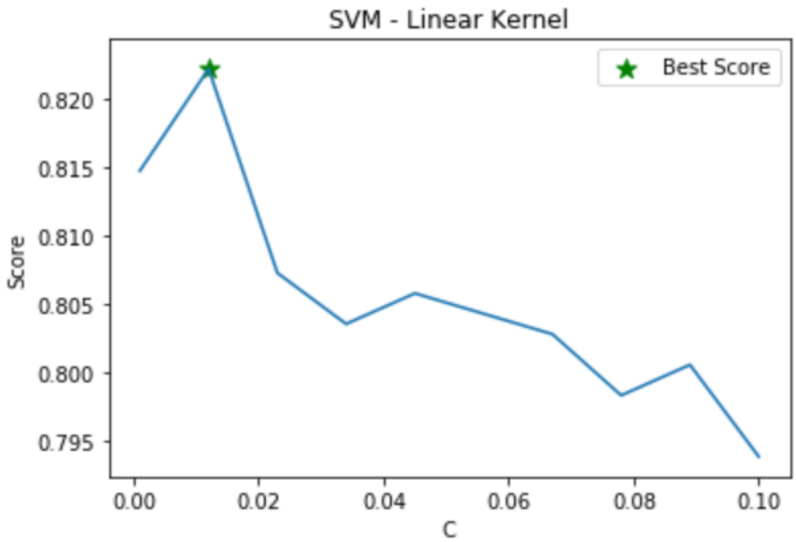

Figure 6.3 - SVM test score according to different values of parameter C.

For the RBF kernel, C and Gamma parameters were tested as presented on Figure 6.4. When concluding a smaller value window for optimization, Figure 6.5. The highest score was 78.4% when testing C=0.88 and γ =1.75x10^{-5}.

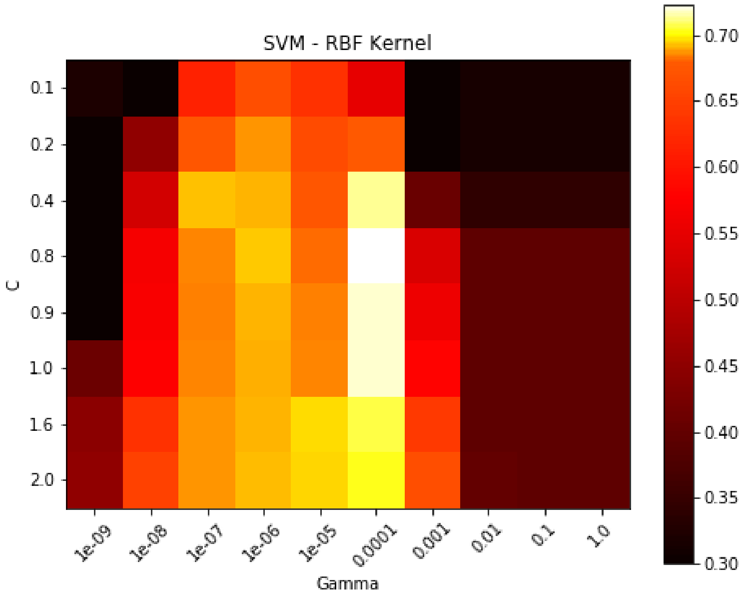

Figure 6.5 - SVM RBF test score heatmap according to different values of parameter C and Gamma.

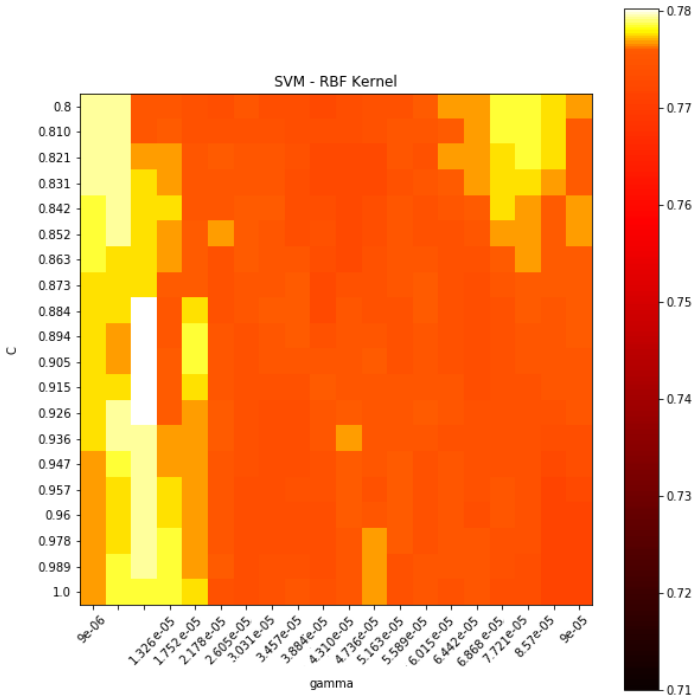

Figure 6.4 - SVM RBF heatmap for a smaller window of values for C and Gamma.

The Neural Network was the last classifiers implemented since it could be interesting to add another model-free classifier to compare with the others. Although one of its weaker points rely on slow learning, it set a score of 80.5%. One of the tested parameters was the *hidden_layer_sizes*, which is used to set the size of the hidden layers. In this study, since there is no standard formula for choosing the number of layers and nodes, after testing some combinations, three layers of 10 nodes each were used to see if it worked the best.

Since the RF outperforms the other methods, presenting the highest score in this accuracy-based approach, feature relevance was studied, as presented in Figure 6.6. From its feature importance plot, it is possible to observe that the measurements acquired in the motor symptoms evaluation performed by the expert are the most relevant. Feature relevancy studies can help improving PD appointment once knowing on which data to focus will examine the patient and reduce computation time.

Most features, as the first two in the graph per example, rely on Part III of the UPDRS which consist of the bigger influence in the score in this scale. These features are identified once their descriptions start with “NP3”. Moreover, other relevant features consist of the time from the first visit and the dose of medication prescribed.

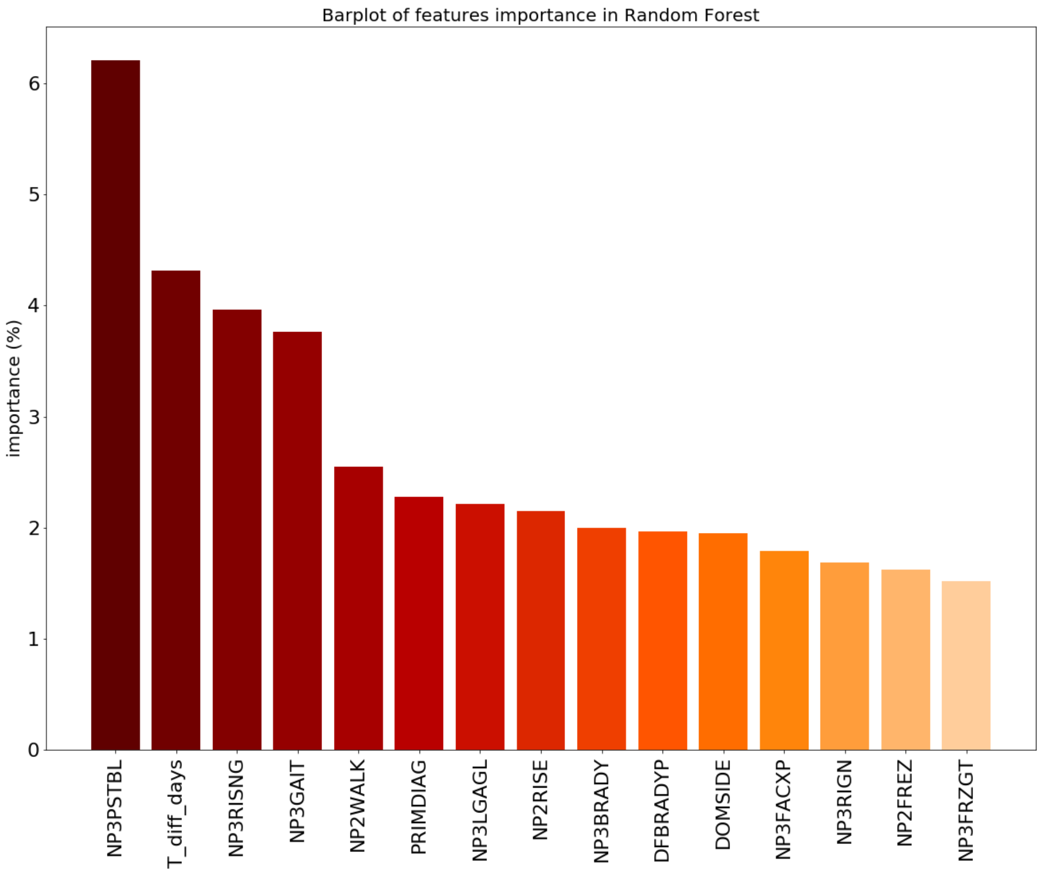

Figure 6.6 - Feature importance for RF classifier.

This approach based on the training and testing classification, of each model, provides an overview of what to expect and which could be more interesting. However, judging the classifier performance based on accuracy is not enough. Nonetheless, a performance evaluation was implemented in order to detail the most promising models for this approach on PD classification.

6.1.1 – Analysis of the performance evaluation

Measuring the performance of the classifier is normal when considering classification problems since the classifiers will predict the class of the instances, as stated in Chapter 3 Section 3.7. In this case, considering a multiclass classification, multiclass confusion matrixes were calculated for each model (Figure 6.7) as well as Precision, Recall and F1-score measurements, Table 6.3.

Table 6.3 – Classification models evaluation by performance metrics.

Method	Accuracy (%)	Precision	Recall	F1-score
RF	87.7	0.86	0.88	0.86
LR	84.3	0.82	0.83	0.82
Gaussian NB	69.2	0.62	0.69	0.65
KNN	77.5	0.69	0.78	0.72
Linear SVM	82.3	0.81	0.82	0.81
SVM (RFB)	78.4	0.71	0.77	0.71
NN	80.5	0.77	0.79	0.73

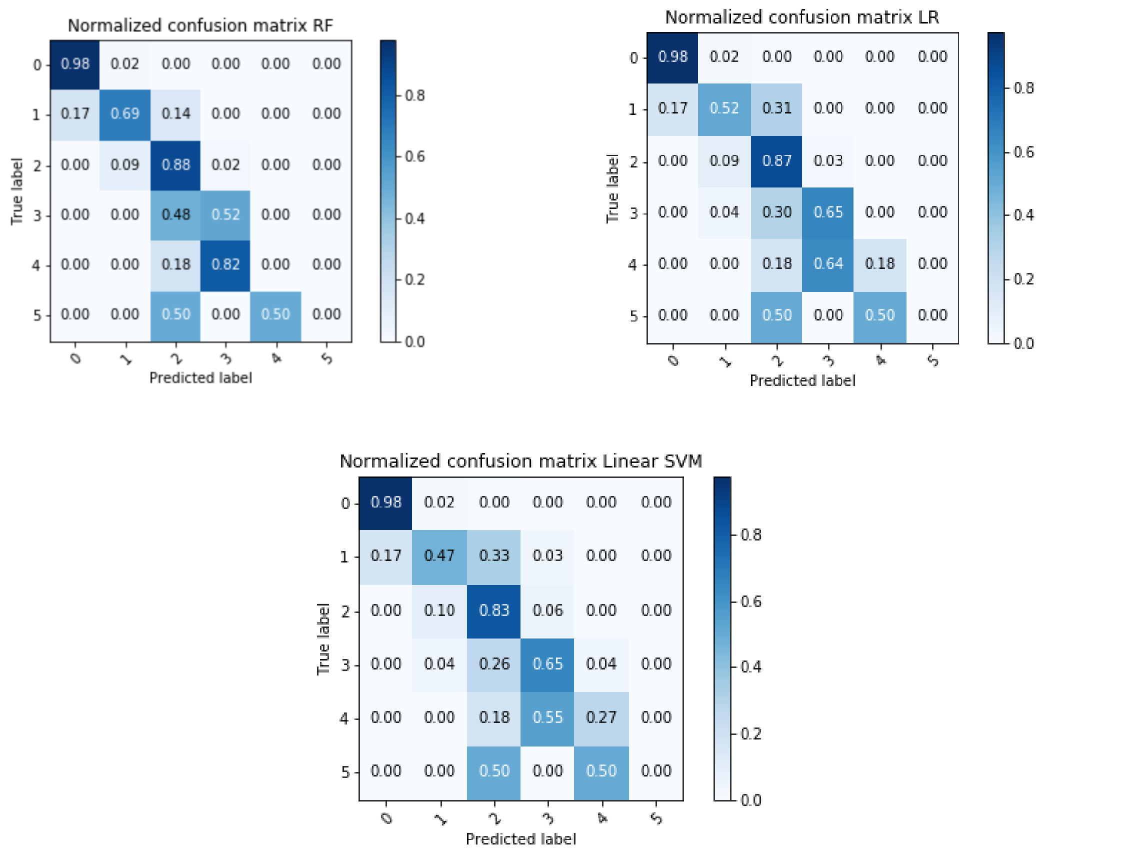

Figure 6.7 - Confusion Matrix from most promising classifiers: RF, LR and Linear SVM.

Bearing in mind the results of the performance metrics examination, the models RF, LR and Linear SVM continue to be the most cohesive. That is considering they present the best precision values indicates how often the prediction is correct, giving scores of 86% for RF and 82% for LR. Even high values of recall, from 88% to 82%, that correspond to the measurement of the correctly predicted rate of the actual samples for a given class.

Analyzing the confusion matrix is possible to conclude different points, starting with an overall look at the diagonal values. Since the diagonal cells show the number and percentage of correct classifications by the trained classifier, a good performance would be checked if that value corresponded to the full number of subjects in that class. In the other hand, off-diagonal cells represent the misclassified predictions. Comparatively, RF continues to present a better performance overall, mainly on class 0 to 2, which means a better precision on early PD cases. Nonetheless, LR and SVM do not perform much worse, presenting similar results.

As for the more severe cases (class 4 and 5), none perform well. However, it can be due to the lack of data to train the algorithms. So, these models present better results when considering early phases of PD, having the downside of misclassifying severe cases which could lead to depreciate symptoms and treatments.

Briefly, it is possible to observe that all the classifiers performed reasonably well with satisfactory results. However, in an accuracy-based judgment, it is important to stand out the classifiers with better performance: RF, LR and Linear SVM, respectively. The best classification score was accomplished when using FR classifier giving the best performance with 87.7% accuracy, 86% precision and 88% recall. Thus, it is also the method with the best overall metric score, including F1-score (0.86), the closest score to the ideal value (1) comparing to the rest.

In comparison to the related works, the performance results are lower, however, they mainly focus on detecting early PD (Hoehn and Yahr scale 1 to 2) and not expand the diagnosis for more PD stages (Hoehn and Yahr scale from 0 to 5). This can be described as a limitation to this study, as pursuing better results, since it included from healthy to early PD to severe cases to subjects at risk of PD but with small symptoms alerts, meaning a more difficult classification process.

6.2 – Classification 2nd trial focusing on previous best results

This section is focused on taking this study to another level as an extension of the classification. This could be done by taking the previous classifiers and address the problem by exploring other *Scikit-learn* tool and more information around its classification process.

As not all the features in the clinical record are strongly associated with each specific response, after implementing a methodology concerning all data available, now the goal is to improve the classification results by identifying some important critical features from the selected methods. In the beginning, all classifiers went through a selection of the 100 best features in order to improve computational running time by implementing the *SelectKBest* class and *f_classif* function. In order to improve the classification scores, this module can be used for feature selection or dimensionality reduction on data sets. From this approach, it was possible to obtain the results presented in Table 6.4.

Table 6.4 - Classification results (train and test) from best features selection and hyperparameters.

Method	Classification training (%)	Classification testing (%)
RF	98.1	88.7
LR	90.2	85.5
NB	80.8	79.7
KNN	83.3	82.5
SVM (Linear)	87.1	86.3
SVM (RFB)	93.9	84.3
NN	99.7	85.6

Through this process, it was possible to plot the influence of the number of features used in the training dataset on the classifiers scores, as presented in Figure 6.8. The plots allow concluding that most methods have their best scores using less than 70 features.

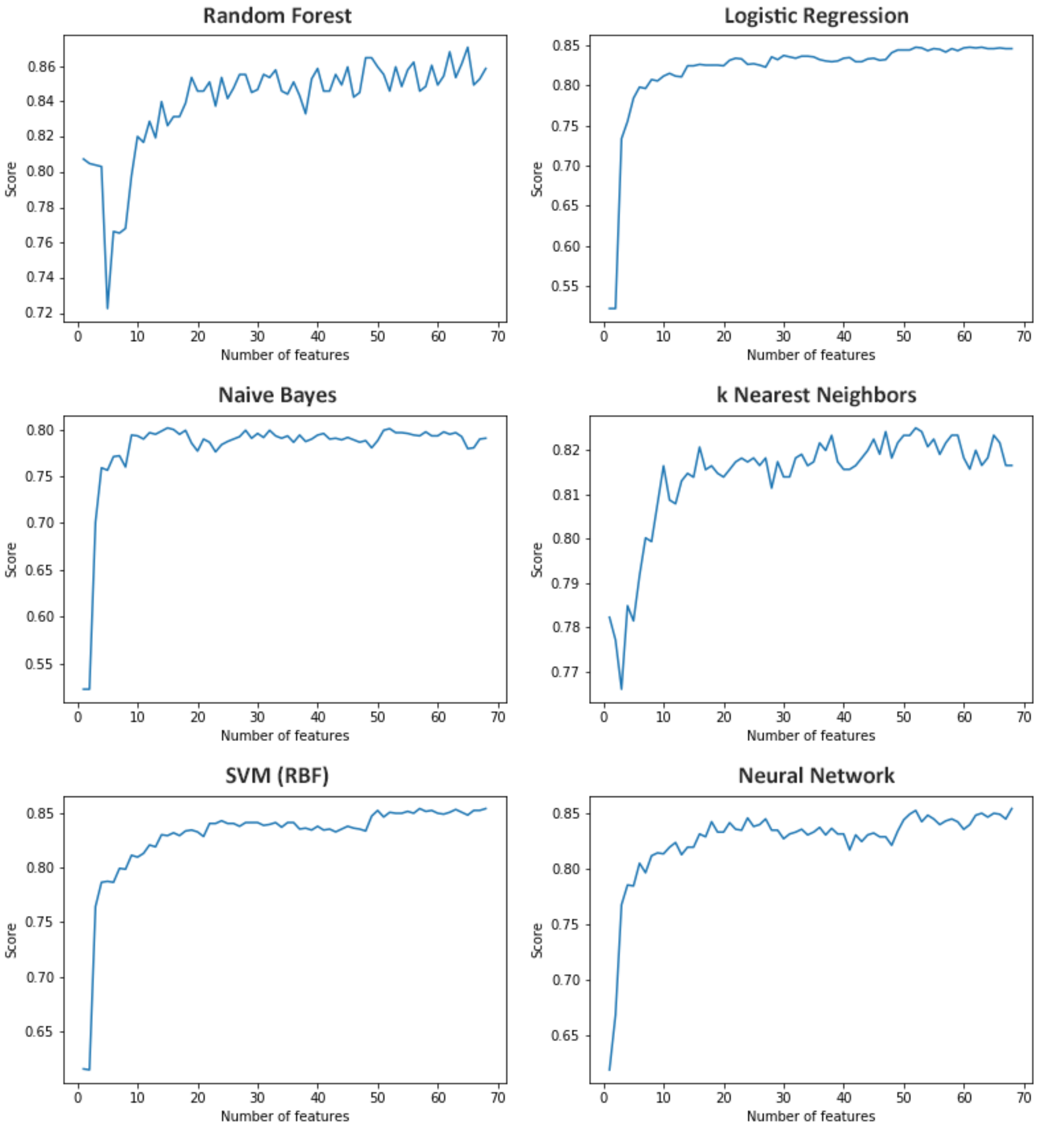

Figure 6.8 - Score results depending on the number of features used.

Since RF outperformed the other methods on Section 6.1, it was the first method tested in this approach promoting the first feature reduction based on the 69 features it declared as most relevant. These reported variable importance results may be useful for selecting features that may be important biomarkers to help clinicians' examinations. It is also visible that the score quickly increases once considering around 10-15 features in all classifiers.

Moreover, focusing on some examples, NB performs its best with only 15, with a score of 79.7%, revealing a worse performance once considering more features of the dataset. However, it also reveals it needed FS step since it improved by 10% comparing to previous results. On the other hand, RF is the model with the highest accuracy of proximally 88.7% at 46 features, followed by Linear SVM with 86.3% and NN with 85.6% accuracy and. Moreover, from these results is possible to compare with the previous approach that contemplates all data, concluding that the RF and Linear SVM gave the best accuracy-based results in both situations. LR also had a similar score with 85.5%.

6.2.1 – Analysis of the performance evaluation

Once more, performance evaluation and analysis were needed since accuracy-based judgment alone is not enough and metrics like Precision and Recall provide a better perception of the classification. Ideally, is important to seek a classifier with high recall and high precision. Table 6.5 presents the mean results of these metrics.

Table 6.5 - Classification models evaluation by performance metrics.

Method	Accuracy (%)	Precision	Recall	F1-score
RF	88.7	0.87	0.87	0.87
LR	85.5	0.84	0.85	0.84
Gaussian NB	79.7	0.83	0.80	0.81
KNN	82.5	0.80	0.83	0.81
Linear SVM	86.3	0.85	0.84	0.83
SVM (RFB)	84.3	0.82	0.84	0.82
NN	85.6	0.85	0.86	0.85

From these results of the performance metrics examination, the model RF continues to be the most cohesive, once again. Nonetheless, RF stands out not only for the highest accuracy (88.7%) but also for being the model presenting more precision and recall. Since both present high and equal values, F1-score (being interpreted as a weighted average of those metrics) is also a reasonable value once is close to 1. When F1-score equals 1, it corresponds to the best-case scenario. Although giving the second-best score, Linear SVM presents worse metrics values with an F1-score of 0.83.

Considering RF results, it presents the best precision values indicates how often the prediction is correct, giving a score of 87% for RF, followed by 85% for NN. Even focusing on the recall values, from 88% to 81%, corresponding to satisfactory measurements of the correctly predicted rate of the actual samples for a given class. Nonetheless, the important is to have a nice balanced between precision and recall, ideally both with high scores, which is the case of RF. Besides both NN and SVM having high classification speed but slow learning and high accuracy in general, SVM is also more computationally expensive to implement and since RF provides higher accuracy and precision, it would be the most indicated model to implement in this study context. Figure 6.9 presents these classifiers Confusion Matrixes for an overall perception of its performance.

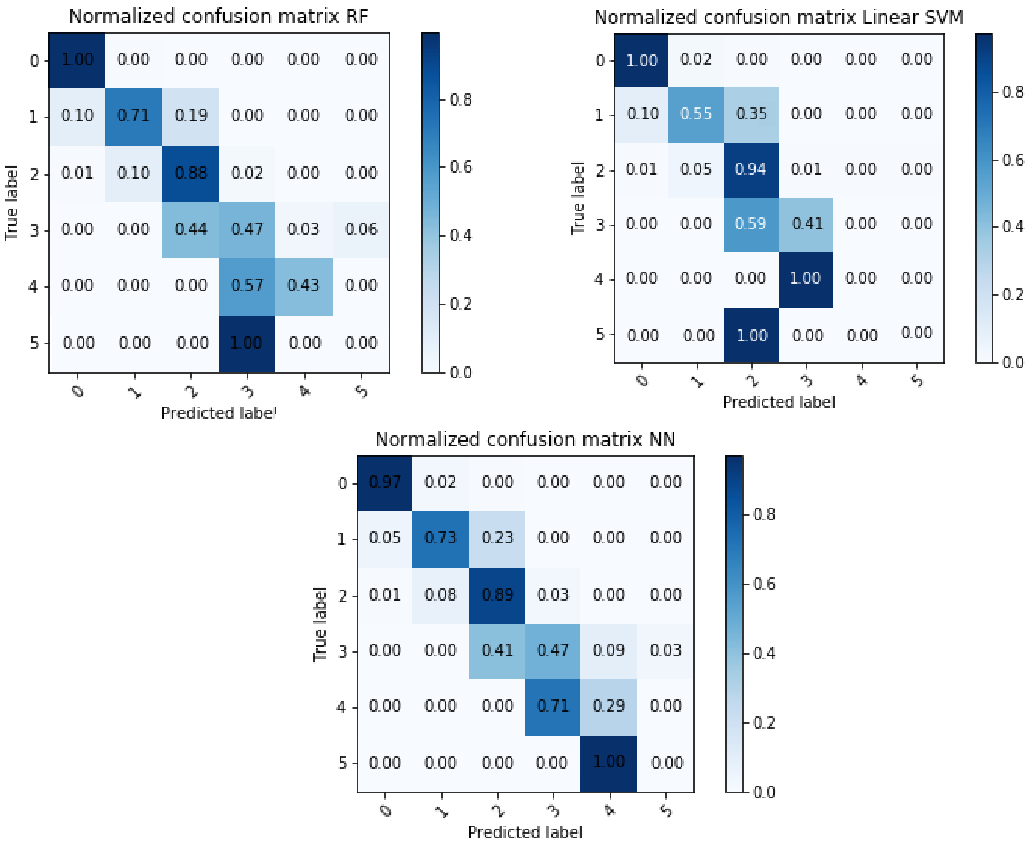

Figure 6.9 - Confusion Matrix for best classifiers (RF, Linear SVM and NN).

Comparing the tree confusion matrixes, it is possible to notice that, once more, the tree models perform worse once classifying severer cases, what could lead to underestimating the patient wellbeing. The most compatible predictions rely on stages 0 (healthy) and 2 (bilateral disease), revealing some struggles once classifying early stage 1 which consists of unilateral manifestation of PD. SVM provides the least favourable diagonal distribution, while RF and NN performed better. Hence, focusing on the matrix diagonals, RF and NN are similar, having fewer cases outside the diagonal, which indicates more actual accurate classification than the other two methods. However, RF presents a more precise classification on early cases.

6.3 - Comparing both trials

Finally, taking an overview of the results, Figure 6.10 provides a general perspective of the results obtained through the proposed methodology. Since performing two trials, one using all data and another focusing on feature reduction, it is possible to compare the final accuracy results. On the graph, the first bar corresponds to the first trial or approach (Section 6.1) and the second bar to the second approach (Section 6.2). As stated before, it is noticeable an overall improvement of the classifier's performance.

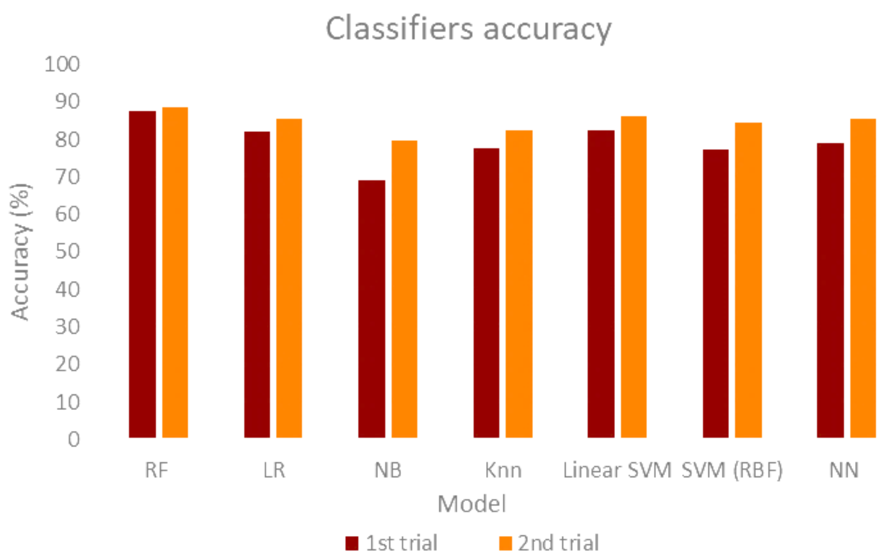

Figure 6.12 - Comparing accuracy scores for every method by the proposed approaches.

All trials were run on a system with an Intel® CoreTM i7-4700MQ CPU @ 2.40 GHz, with 12 GB of RAM and NVIDIA GeForce 840M with 8GB of RAM, therefore the robustness of the algorithm to different environments was not tested. Doing an overall perspective, by implementing feature dimension reduction, an impact on computational run time is stated, which allows comparing its difference. Since the hyper-parameter search was implemented in both trials, it is possible to compare how much time it spent on this process.

Ruining all classifiers, the first trial took about 54 minutes while the second one took proximally 13 minutes, resulting in a 75%-time reduction. However, when tested already with the hyper-parameters, run times decreases to 3 minutes and 10 seconds, respectively. On both attempts, LR and SVM (RBF) were the slowest models, opposing to RF.

However, it important to notice that, as expected, besides the 2nd approach taking less time testing, the results were also superior. Even not accomplishing perfect scores, this approach could be implemented as a support tool for expert performing PD diagnosis.

6.4 – Final remarks

The approaches conducted over the course of this chapter show good results. Since accuracy alone does not provides a complete conclusion on the algorithm performance, both approaches were tested and evaluated. In the first one, using all the data available in the *dataframe*, methods like RF, linear SVM and LR showed the best results. On the other hand, adding the feature selection step, the best classifiers were also RF and linear SVM, adding NN.

Comparing both it was possible to conclude that the best classification results are provided by RF with an 88.7% accuracy and 0.87 F1-score, on the second approach. Since it performs on a smaller dataset, containing the most significant features, its run time decreases, and its performance evaluation showed better results. Hence, an overall improvement between approaches is visible.

Nonetheless, the PPMI is currently the most noticeable comprehensive multicenter, international study for PD, embracing a large variety of subjects. Since being an overwhelmed dataset that includes a large variety of information and PD biomarkers, it is important and interesting to explore different approaches on how to treat the data. Moreover, training and testing several methods provide to perform the classification either to improve estimators' accuracy scores or to boost their performance on very high-dimensional datasets. Furthermore, performance improvement becomes an extremely important step once focusing on medical dataset.

Chapter 7

Conclusions and Future Work

The last chapter of this dissertation concerns an overall analysis of all the information gathered in this document, from theoretical concepts of PD diagnosis and Machine Learning approaches to this dissertation approach and the final thoughts concerning the obtained results. Finally, a perspective of future work.

7.1 – Conclusions

Early and precise diagnosis is one ambitious objective of modern systems when applied to the medical fields. Currently, most common methods for PD evaluation and diagnosis require and rely essentially on physicians or experts in the area. However, at the time of detection, proximally 60% of the dopaminergic neurons have been already lost. Nonetheless, it is necessary to invest in diagnosis improvement taking experts systems into account as becoming an expert doctor can mean automatizing diagnosis processes. This can lead to human errors or difficulties on multi features correlations can occur.

This dissertation presented a supervised machine learning methodology to detect/classify PD patients from a wide range of data provided by the PPMI study from the Michael J. Fox Foundation. In the beginning, following most literature review studies, the primordial planned goal consisted of binary classification between healthy and parkinsonian diagnosis. However, giving the variety of subject enrol in the used dataset, a full spectrum diagnosis was approached, turning to a multiclass classification process. This allowed a PD detection or diagnosis in different stages of the disease, from early to severe.

Starting from organizing all the available data and excluding impartial/incomplete information, it was possible to gather a work-ready dataset to test the proposed methodology. After normalizing the data, the focus consisted of putting several models or classifiers to the test and see which machine learning approach could provide the best classification. This study was implemented considering 672 people with no signs of PD, within the 1674 involved. Therefore, the RF, LR, NB, kNN, SVM (Linear and RBF) and NN were selected, tested, optimized by a hyperparameter search and finally evaluated by performance metrics. Passing the data organization and pre-processing, the classification phase allowed to compare the behaviour of each model and its performance facing the problem.

Focusing on the final results, it is possible to conclude that considering all data available on the dataset, RF, LR and Linear SVM promoted the best performance. Yet, on a second approach, narrowing down the features, the final selected models were RF, Linear SVM and NN. Besides RF and Linear SVM being present at both approaches, RF outperformed the rest by presenting the highest scores for accuracy, precision and recall. Since SVM are a known computational expensive model compare to others, RF becomes more promising by also performing a faster classification in this study. Taking everything into consideration, the best solution for the proposed problem consists of implementing a RF classification model with feature dimensional reduction, considering the features with more relevance. This option presented an 88.7% accuracy, 87% precision and 87% recall, outperforming the 87% accuracy of the full data approach and decreasing computational run time.

Overall, the motivation for this proposed approach on the problem, besides the wellbeing and diagnosis of PD patients, was to explore classification models which could provide a support mechanism for doctors in the near future. Being the main goal to apply machine learning concepts to PD diagnosis, it is possible not only to help on the diagnosis process but also to help a more complete follow up of the patients. Based on the literature, most present high-performance solutions, test different algorithms and classifiers looking for the most accurate, efficient and computably assessable system, which motivates the continuous investment on PD diagnosis improvement.

Finally, this dissertation allowed to understand that besides several efforts and generally satisfactory results when testing machine learning algorithm, there is no single method that can be generalized and straightforward for the study of PD and the proposed methodology provided yet another take on the subject. Therefore, by improving the classification and diagnosis with this type of approach, there is the hope of promoting an extra investment in technology in healthcare. Not only could it improve diagnosis but also help on scheduling the next visits and optimizing treatment doses. This can be done by assessing the patient's state based on the classification result and even bring advantages to hospitals as it can enhance the management process and anticipation of their next appointment, reducing costs.

7.2 – Future work

As future work, it would be interesting to continuing improving the classification and even advance for more computational powerful tools. It could also be interesting to take each part of the database, focusing on only motor symptoms per example and do a more in-depth study, exploring a solution for more promising models with better performance. Although this methodology could not replace a doctor/expert diagnosis, better results could promote an extra investment in technology in healthcare promising results for PD analysis and diagnosis through Machine Learning and Data Mining techniques application. However, this still consists of a major challenge and concern as many of the methods still offer limited solutions.

References

[1] A. Schapira, *Parkinson's Disease*. London: Oxford University Press, 2010.

[2] R. Junior, "Doença de Parkinson vai aumentar em Portugal," 2017. [Online]. Available: https://www.dn.pt/sociedade/interior/doenca-de-parkinson-vai-aumentar-em-portugal-com-o-envelhecimento-da-populacao-5784080.html. [Accessed: 20-Apr-2018].

[3] J. J. Ferreira, "Prevalence of Parkinson's disease: a population-based study in Portugal," *Eur. J. Neurol.*, vol. 24, no. 5, pp. 748–750, 2017.

[4] "Doença de Parkinson @ www.saudecuf.pt." [Online]. Available: www.saudecuf.pt. [Accessed: 01-May-2018].

[5] M. Achey, "The past, present, and future of telemedicine for Parkinson's disease," *Mov. Disord.*, vol. 29, no. 7, pp. 871–883, 2014.

[6] W. H. Organization, "Neurological disorders associated with malnutrition," *Neurol. Disord. public Heal. challenges*, pp. 140–150, 2006.

[7] A. Gustavsson *et al.*, "Cost of disorders of the brain in Europe 2010," pp. 718–779, 2011.

[8] R. Prashanth, S. Dutta, P. K. Mandal, and S. Ghosh, "High-Accuracy Detection of Early Parkinson ' s Disease through Multimodal Features and Machine Learning," *Int. J. Med. Inform.*, vol. 90, pp. 13–21, 2016.

[9] F. Bill and M. G. Foundation, "Articles Global , regional , and national burden of Parkinson ' s disease , 1990 – 2016 : a systematic analysis for the Global Burden of Disease Study 2016," pp. 939–953, 2016.

[10] Scopus, "Form @ Www.Scopus.Com." [Online]. Available: http://www.scopus.com/. [Accessed: 25-Jan-2019].

[11] K. Theodoridis and S. Koutroumbas, *Pattern Recognition*, 4th ed. Elsevier Academic Press, 2003.

[12] I. Kononenko, "Machine learning for medical diagnosis: History, state of the art and perspective," *Artif. Intell. Med.*, vol. 23, no. 1, pp. 89–109, 2001.

[13] J. Massano, "Doença de parkinson: Actualização clínica," *Acta Med. Port.*, vol. 24, no. SUPPL.4, pp. 827–834, 2011.

[14] S. Francis and E. Medical, "James Parkinson's Essay on the Shaking Palsy Shaking," pp. 84–86, 2015.

[15] A. B. Pimentel, "Algorithm for the Parkinson ' s Disease Behavioural Models Characterization using a Biosensor," 2012.

[16] E. C. Hirsch, "Biochemistry of parkinson's disease with special reference to the dopaminergic systems," *Mol. Neurobiol.*, vol. 9, no. 1–3, pp. 135–142.

[17] B. Fischer, “Efeitos do treinamento de potência na força muscular de indivíduos com doença de parkinson,” 2014.

[18] S. Von Campenhausen *et al.*, “Prevalence and incidence of Parkinson ’ s disease in Europe B,” vol. 15, pp. 473–490, 2005.

[19] Lazaros C. and Triarhou, *Dopamine and Parkinson’s Disease*. Landes Bioscience, 2013.

[20] J. Clark, “Exploring Degenerative Diseases,” 2017. .

[21] F. . Moisan and A. Elbaz, “Update in the epidemiology of Parkinson’s disease,” *Curr Opin Neurol*, vol. 21, no. 4, pp. 454–460, 2008.

[22] JMS, “Cuf - Doença de Parkinson,” 2017. [Online]. Available: https://www.saudecuf.pt/mais-saude/doencas-a-z/parkinson. [Accessed: 10-May-2018].

[23] “ECP, um tratamento cirúrgico para a doença de Parkinson,” 2016. .

[24] T. Pringsheim, N. Jette, A. Frolkis, and T. D. L. Steeves, “The Prevalence of Parkinson ’ s Disease : A Systematic Review and,” vol. 29, no. 13, pp. 1583–1590, 2014.

[25] J. Jankovic, “Parkinson ’ s disease : clinical features and diagnosis,” no. 1957, pp. 368–376, 2008.

[26] J. M. S. Sousa, “Análise da Conectividade Estrutural na Doença de Parkinson.”

[27] American Parkinson Disease Association, “SYMPTOMS OF PARKINSON’S.” [Online]. Available: https://www.apdaparkinson.org/what-is-parkinsons/symptoms/. [Accessed: 20-Dec-2018].

[28] E. C. Wolters, “Variability in the clinical expression of Parkinson ’ s disease,” vol. 266, pp. 197–203, 2008.

[29] American Parkinson Disease Association, “Diagnosis of Parkinson ’s.” [Online]. Available: https://www.apdaparkinson.org/what-is-parkinsons/diagnosing/.

[30] E. Rovini, C. Maremmani, F. Cavallo, and G. Mirabella, “How Wearable Sensors Can Support Parkinson ’ s Disease Diagnosis and Treatment : A Systematic Review,” vol. 11, no. October, 2017.

[31] J. Kassubek, “Diagnostic procedures during the course of Parkinson ’ s Disease,” *Biochem. Pharmacol.*, pp. 2–5, 2014.

[32] C. G. Goetz *et al.*, “Movement Disorder Society-Sponsored Revision of the Unified Parkinson ’ s Disease Rating Scale (MDS-UPDRS): Scale Presentation and Clinimetric Testing Results,” vol. 23, no. 15, pp. 2129–2170, 2008.

[33] C. G. Goetz, “MDS-UPDRS,” 2008.

[34] R. Graça, “ParkDetect Early Diagnosing Parkinson ’ s Disease,” University of Porto, 2013.

[35] C. G. Goetz *et al.*, “Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: Status and recommendations,” *Mov. Disord.*, vol. 19, no. 9, pp. 1020–1028, 2004.

[36] American Parkinson Disease Association, “Treatment & Medication.” [Online]. Available: https://www.apdaparkinson.org/what-is-parkinsons/treatment-medication/. [Accessed: 22-Dec-2018].

[37] C. C. Bennett and K. Hauser, “Artificial intelligence framework for simulating clinical decision-making : A Markov decision process approach,” *Artif. Intell. Med.*, vol. 57, no. 1, pp. 9–19, 2013.

[38] C. F. Tan, L. S. Wahidin, S. N. Khalil, N. Tamaldin, J. Hu, and G. W. M. Rauterberg, “the Application of Expert System : a Review of Research,” vol. 11, no. 4, pp. 2448–2453, 2016.

[39] H. Tan, “A brief history and technical review of the expert system research,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 242, no. 1, 2017.

[40] V. L. Patel *et al.*, “The coming of age of artificial intelligence in medicine,” *Artif. Intell. Med.*, vol. 46, no. 1, pp. 5–17, 2009.

[41] A. N. Ramesh, C. Kambhampati, J. R. T. Monson, and P. J. Drew, “Artificial intelligence in medicine,” *Ann. R. Coll. Surg. Engl.*, vol. 86, no. 5, pp. 334–338, 2004.

[42] E. Keravnou and N. Lavrač, “AIM portraits: Tracing the evolution of artificial intelligence in medicine and predicting its future in the new millennium,” *Artif. Intell. Med.*, vol. 23, no. 1, pp. 1–4, 2001.

[43] S. Navigation and N. S. Route, *Data Mining Practical Machine Learning Tools and Techniques*. .

[44] N. J. Nilsson, “Introduction to Machine Learning,” 1997.

[45] Z. Obermeyer and E. J. Emanuel, “Predicting the Future — Big Data, Machine Learning, and Clinical Medicine,” vol. 375, no. 13, pp. 1216–1219, 2016.

[46] M. Stefanelli, “The socio-organizational age of artificial intelligence in medicine,” *Artif. Intell. Med.*, vol. 23, no. 1, pp. 25–47, 2001.

[47] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, “Machine Learning and Data Mining Methods in Diabetes Research,” *Comput. Struct. Biotechnol. J.*, vol. 15, pp. 104–116, 2017.

[48] B. G. Silverman, N. Hanrahan, G. Bharathy, K. Gordon, and D. Johnson, “A systems approach to healthcare: Agent-based modeling, community mental health, and population well-being,” *Artif. Intell. Med.*, vol. 63, no. 2, pp. 61–71, 2015.

[49] D. D. Luxton, “Recommendations for the ethical use and design of artificial intelligent care providers,” *Artif. Intell. Med.*, vol. 62, no. 1, pp. 1–10, 2014.

[50] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, “Supervised Machine Learning: A Review of Classification Techniques,” *Artif. Intell. Rev.*, vol. 26, pp. 159–190, 2006.

[51] D. Erhan, A. Courville, and P. Vincent, “Why Does Unsupervised Pre-training Help Deep Learning ?,” vol. 11, pp. 625–660, 2010.

[52] M. Martiarena and S. Matteucci, “Semi-Supervised Learning Literature Survey,” *Fac. Arquit.*, vol. Master, p. 179, 2007.

[53] Alpaydin E., *Introduction to machine learning*. Cambridge Massachusetts London England: The MIT Press, 2004.

[54] M. Kornitzer *et al.*, “Medical Image Feature, Extraction, Selection And Classification,” *Lancet*, vol. 321, no. 8333, pp. 1066–1070, 1983.

[55] Ö. E. Yetgin and Ö. N. Gerek, “Feature extraction, selection and classification code for power line scene recognition,” *SoftwareX*, 2017.

[56] D. Lafond, M. Duarte, and F. Prince, “Comparison of three methods to estimate the center of mass during balance assessment,” vol. 37, pp. 1421–1426, 2004.

[57] K. J. Cios and G. W. Moore, “Uniqueness of medical data mining.pdf,” vol. 26, pp. 1–24, 2002.

[58] O. Maimon and L. Rokach, *Data Mining and Knowledge Discovery*, vol. 26, no. 1. 1999.

[59] Ted. Shah, “How AI is making it easier to diagnose a disease @ Www.Ted.Com.” [Online]. Available: https://www.ted.com/talks/jane_chen_a_warm_embrace_that_saves_lives/up-next#t-153523. [Accessed: 20-Dec-2018].

[60] T. Z. Phyu and N. N. Oo, “Performance Comparison of Feature Selection Methods,”

MATEC Web Conf., vol. 42, p. 06002, 2016.

[61] L. Ladha and T. Deepa, “Feature Selection Methods and Algorithms,” *Int. J. Comput. Sci. Eng.*, vol. 3, no. 5, pp. 1787–1797, 2011.

[62] “What is the difference between filter, wrapper, and embedded methods for feature selection? @ sebastianraschka.com.” [Online]. Available: https://sebastianraschka.com/faq/docs/feature_sele_categories.html?fbclid=IwAR2cALUhPfYwY7Re-KhVlOGsUiMIRC4yOlKrdZnFfoSJ-7K-eeMxBgZUHgs. [Accessed: 02-Jan-2019].

[63] S. F. Cotter, K. Kreutz-Delgado, and B. D. Rao, “Backward sequential elimination for sparse vector subset selection,” *Signal Processing*, vol. 81, no. 9, pp. 1849–1864, 2001.

[64] A. Ozcift, “SVM feature selection based rotation forest ensemble classifiers to improve computer-Aided diagnosis of Parkinson disease,” *J. Med. Syst.*, vol. 36, no. 4, pp. 2141–2147, 2012.

[65] A. Suganya, “Lung Nodule Classification Techniques for Low Dose Computed Lung Nodule Classification Techniques for Low Dose Computed Tomography (LDCT) Scan Images as Survey,” no. December, 2015.

[66] L. Parisi, N. RaviChandran, and M. L. Manaog, “Feature-driven machine learning to improve early diagnosis of parKinson’s disease,” *Expert Syst. Appl.*, vol. 110, pp. 182–190, 2018.

[67] R. Das, “A comparison of multiple classification methods for diagnosis of Parkinson disease,” *Expert Syst. Appl.*, vol. 37, no. 2, pp. 1568–1572, 2010.

[68] I. Maglogiannis, K. Karpouzis, M. Wallace, and J. Soldatos, *Emerging Artificial Intelligence Applications In Computer Engineering*. IOS Press.

[69] G. Orrù, W. Pettersson-Yeo, A. F. Marquand, G. Sartori, and A. Mechelli, “Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review,” *Neurosci. Biobehav. Rev.*, vol. 36, no. 4, pp. 1140–1152, 2012.

[70] N. Williams, S. Zander, and G. Armitage, “A preliminary performance comparison of five machine learning algorithms for practical IP traffic flow classification,” *ACM SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 5, p. 5, 2006.

[71] C. Paper and P. Brazdil, “Artificial Intelligence in Theory and Practice,” vol. 217, no. August 2006, pp. 0–10, 2006.

[72] M. Sunasra, “Performance Metrics for Classification problems in Machine Learning @ medium.com,” 2017. [Online]. Available: https://medium.com/greyatom/performance-metrics-for-classification-problems-in-machine-learning-part-i-b085d432082b. [Accessed: 23-Dec-2018].

[73] A. Guanga, “Understand Classification Performance Metrics @ becominghuman.ai,” 2018. [Online]. Available: https://becominghuman.ai/understand-classification-performance-metrics-cad56f2da3aa. [Accessed: 23-Dec-2018].

[74] A. Mishra, “Metrics to Evaluate your Machine Learning Algorithm @ towardsdatascience.com,” 2018. [Online]. Available: https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234. [Accessed: 23-Dec-2018].

[75] P. B. Cerrito, *Introduction to data mining using SAS enterprise miner*. Cary, NC, USA, 2006.

[76] J. W. M. D. Souza, S. S. A. Alves, E. D. S. Rebouças, J. S. Almeida, and P. P. Rebouças Filho, “A New Approach to Diagnose Parkinson’s Disease Using a Structural Cooccurrence

Matrix for a Similarity Analysis," *Comput. Intell. Neurosci.*, vol. 2018, 2018.

[77] H. Thimbleby, "Technology and the future of healthcare," vol. 2, 2013.

[78] Pubmed, "Pubmed @ www.ncbi.nlm.nih.gov." [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/. [Accessed: 10-Feb-2018].

[79] ScienceDirect, "Index @ Www.Sciencedirect.Com," 2560. [Online]. Available: http://www.sciencedirect.com/.

[80] Mendeley, "Machine Learning Parkinson's Disease @ www.mendeley.com." [Online]. Available: https://www.mendeley.com/research-papers/?query=Machine learning Parkinson. [Accessed: 11-Feb-2018].

[81] M. Contin, R. Riva, A. Baruzzi, F. Albani, and S. Macri, "Postural Stability in Parkinson ' s Disease : the Effects of Disease Severity and Acute Levodopa Dosing," vol. 2, no. 1, pp. 29–33, 1996.

[82] R. de M. Roiz, E. W. A. Cacho, M. M. Pazinatto, J. G. Reis, A. Cliquet Jr, and E. M. A. Barasnevicius-Quagliato, "Gait analysis comparing Parkinson's disease with healthy elderly subjects," *Arq. Neuropsiquiatr.*, vol. 68, no. 1, pp. 81–86, 2010.

[83] H. hazfiza Md Tahir, Nooritawati,Manap, "Parkinson's Disease Gait Classification based on Machine Learning Approach.pdf." 2012.

[84] E. Abdulhay, N. Arunkumar, K. Narasimhan, E. Vellaiappan, and V. Venkatraman, "Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease," *Futur. Gener. Comput. Syst.*, vol. 83, pp. 366–373, 2018.

[85] C. R. Pereira *et al.*, "A new computer vision-based approach to aid the diagnosis of Parkinson's disease," *Comput. Methods Programs Biomed.*, vol. 136, pp. 79–88, 2016.

[86] I. Mandal and N. Sairam, "Accurate telemonitoring of Parkinson's disease diagnosis using robust inference system," *Int. J. Med. Inform.*, vol. 82, no. 5, pp. 359–377, 2013.

[87] T. Khan, J. Westin, and M. Dougherty, "Classification of speech intelligibility in Parkinson's disease," *Biocybern. Biomed. Eng.*, vol. 34, no. 1, pp. 35–45, 2014.

[88] S. A. Mostafa *et al.*, "Examining Multiple Feature Evaluation and Classification Methods for Improving the Diagnosis of Parkinson's Disease," in *Cognitive Systems Research*, vol. 54, Elsevier B.V., 2018, pp. 90–99.

[89] F. Segovia, J. M. Górriz, J. Ramírez, I. Álvarez, J. M. Jiménez-Hoyuela, and S. J. Ortega, "Improved Parkinsonism diagnosis using a partial least squares based approach," *Med. Phys.*, vol. 39, no. 7, pp. 4395–4403, 2012.

[90] I. I.A. *et al.*, "Automatic assistance to Parkinson's disease diagnosis in DaTSCAN SPECT imaging," *Medical physics*, vol. 39, no. 10. pp. 5971–5980, 2012.

[91] R. Armañanzas, C. Bielza, K. R. Chaudhuri, P. Martinez-Martin, and P. Larrañaga, "Unveiling relevant non-motor Parkinson's disease severity symptoms using a machine learning approach," in *Artificial Intelligence in Medicine*, vol. 58, no. 3, 2013, pp. 195-202.

[92] K. Marek *et al.*, "The Parkinson Progression Marker Initiative (PPMI)," *Prog. Neurobiol.*, vol. 95, no. 4, pp. 629–635, 2011.

[93] R. Prashanth, S. Dutta Roy, P. K. Mandal, and S. Ghosh, "Automatic classification and prediction models for early Parkinson's disease diagnosis from SPECT imaging," *Expert Syst. Appl.*, vol. 41, no. 7, pp. 3333–3342, 2014.

[94] M. Laal, "Technology in Medical Science," *Procedia - Soc. Behav. Sci.*, vol. 81, pp. 384-388, 2013.

[95] K. Marek, *The Parkinson's Progression Markers Initiative (PPMI)*. 2017.

[96] Python Software Foundation, “ython Language Reference, version 2.7,” 2015. [Online]. Available: https://www.python.org/.
[97] “Sklearn StandardScaler @ Scikit-Learn.Org.” [Online]. Available: http://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html. [Accessed: 05-Apr-2019].
[98] “Cross_Validation @ Scikit-Learn.Org.” [Online]. Available: https://scikit-learn.org/stable/modules/cross_validation.html#. [Accessed: 15-Apr-2019].
[99] C. Gao *et al.*, “Model-based and model-free machine learning techniques for diagnostic prediction and classification of clinical outcomes in Parkinson’s disease,” *Sci. Rep.*, vol. 8, no. 1, pp. 1–21, 2018.
[100] J. VanderPlas, *Python Data Science Handbook*. O’Reilly Media, 2016.
[101] “Logistic regression @ scikit-learn.org.” [Online]. Available: https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression. [Accessed: 26-Mar-2019].
[102] B. Çığşar and D. Unal, “Comparison of Data Mining Classification Algorithms Determining the Default Risk,” *Sci. Program.*, vol. 2019, 2019.
[103] “Nearest Neighbors @ scikit-learn.org.” [Online]. Available: https://scikit-learn.org/stable/modules/neighbors.html#classification. [Accessed: 18-Apr-2019].
[104] “Support Vector Machines @ scikit-learn.org.” [Online]. Available: https://scikit-learn.org/stable/modules/svm.html#svm-mathematical-formulation. [Accessed: 25-Mar-2019].
[105] “neural_networks_supervised @ scikit-learn.org.” [Online]. Available: https://scikit-learn.org/stable/modules/neural_networks_supervised.html#neural-networks-supervised. [Accessed: 10-May-2019].
[106] C.-M. Ma, W. Yang, and B. Cheng, “How the Parameters of K-nearest Neighbor Algorithm Impact on the Best Classification Accuracy: In Case of Parkinson Dataset,” *Sci. Alert*, vol. 14, no. 2, pp. 171–176, 2014.