Multimodal Assessment of Parkinson's Disease Patients Using Information from Speech, Handwriting, and Gait

Juan Camilo Vásquez Correa

Tesis doctoral presentada para optar al título de Doctor en Ingeniería Electrónica y de Computación

Directores
Juan Rafael Orozco Arroyave, Doctor (PhD) en Ingeniería
Elmar Nöth, Doctor (PhD) en Ingeniería

Universidad de Antioquia
Facultad de Ingeniería
Doctorado en Ingeniería Electrónica y de Computación
Medellín, Antioquia, Colombia
2023

Multimodal Assessment of Parkinson's Disease Patients Using Information from Speech, Handwriting, and Gait

Multimodale Verarbeitung von Parkinson-Krankheit unter Verwendung von Informationen aus Sprache, Handschrift und Gang

Der Technischen Fakultät der
Friedrich-Alexander-Universität Erlangen-Nürnberg

und

Der Technischen Fakultät der
Universidad de Antioquia

zur Erlangung des Grades

DOKTOR-INGENIEUR

vorgelegt von

Juan Camilo Vásquez-Correa
Erlangen, Deutschland

Multimodal Assessment of Parkinson's Disease Patients Using Information from Speech, Handwriting, and Gait

Multimodale Verarbeitung von Parkinson-Krankheit unter Verwendung von Informationen aus Sprache, Handschrift und Gang

Als Dissertation genehmigt
von der Technischen Fakultät
der Friedrich-Alexander-Universität Erlangen-Nürnberg,
Erlangen, Deutschland

un der

Technischen Fakultät
der Universidad de Antioquia, Medellín, Colombia

Tag der mündlichen Prüfung:	07.10.2022
Gutachterinnen:	Prof. Dr.-Ing. habil. Elmar Nöth
	Prof. Dr.-Ing. Juan Rafael Orozco-Arroyave
	Prof. Dr. Julie Mauclair
	Prof. Dr. Meinard Müller

Abstract

The automatic analysis of different bio-signals from patients with Parkinson's disease is a highly relevant topic that has been addressed by the research community within several years. Identifying bio-markers for early and differential diagnosis, severity assessment, and response to therapy is a primary goal of the research on Parkinson's disease today. There are important contributions of these topics considering different bio-signals individually. Multimodal analyses, i.e., considering information from different sensors, have not been extensively studied. Although many improvements have been shown in several tasks, there is still an absence of a multimodal system able to deliver an accurate prediction of the disease severity and to monitor the disease progression. The aim of this thesis is to develop robust models for the accurate diagnosis of Parkinson's disease and to evaluate the disease severity of patients using different bio-signals such as speech, online handwriting, gait (using inertial sensors), and those signals collected from smartphones. The proposed models are evaluated in three application scenarios: (1) The automatic classification of healthy subjects and Parkinson's patients. (2) The evaluation of the disease severity of the patients based on a clinical scale, including both the motor-state severity and the dysarthria level of the subjects. (3) The classification of patients into different groups according to their disease severity e.g., mild, moderate, and severe. The experiments covered both traditional pattern recognition and novel deep learning models.

Three approaches are introduced to model the speech of Parkinson's patients: (1) phonological analysis of speech, which is more interpretable for clinicians by directly modeling information about the mode and manner of articulation. (2) Representation learning strategies using recurrent autoencoders, which have the potential to extract more abstract and robust features than those traditionally computed. Finally, (3) convolutional neural networks trained to process time-frequency representations of the speech of the patients. Regarding handwriting analysis, the proposed approach involves the computation of traditional kinematic features, combined with novel approaches based on geometric, and in-air features. Deep learning models based on convolutional neural networks are also proposed to evaluate both raw online handwriting data, and the reconstructed offline images created by the patients. The proposed approaches for gait analysis involve the computation of traditional kinematic and spectral features, combined with novel approaches based on non-linear dynamics. A deep-learning approach combining convolutional and recurrent neural networks is also introduced to model the gait signals from the patients. Finally, this thesis covered a multimodal analysis of the speech, handwriting, and gait signals collected from the patients. The addressed experiments are carried out using both early and late fusion strategies. The proposed methods are also evaluated in two scenarios: (1) high quality sensors, which can be available in medical centers for the assessment of patients, and (2) data collected via smartphones, which can be used for continuous monitoring of patients at home. The results indicate that the combined results outperformed those obtained with each bio-signal separately, both for the automatic classification of the disease and the evaluation of the disease severity. In addition, the proposed models are robust to be applied both on signals collected with high-quality sensors and smartphones.

Resumen

El análisis automático de diferentes bio-señales en pacientes con enfermedad de Parkinson es un tema altamente relevante y que ha sido abordado por la comunidad científica durante varios años. La identificación de bio-marcadores para la detección temprana y diferencial, evaluación de la severidad, y respuesta a la terapia es un objetivo primordial en la investigación actual de la enfermedad de Parkinson. Existen importantes contribuciones en estos temas considerando diferentes bio-señales de manera individual. Sistemas multimodales que consideren información de diferentes sensores no han sido ampliamente estudiados. A pesar de que muchas contribuciones se han propuesto para diferentes tareas, aún existe una ausencia de un sistema multimodal capaz de entregar una predicción acertada de la severidad de la enfermedad y monitorear el progreso de la misma. El objetivo de esta disertación es desarrollar modelos para apoyar el diagnóstico y evaluar la severidad de la enfermedad de Parkinson por medio de diferentes bio-señales como la voz, la escritura online, y la marcha (usando sensores inerciales), además de señales capturadas con teléfonos inteligentes. Los modelos propuestos se evalúan en tres escenarios de aplicación: (1) clasificación automática de personas sanas y pacientes con Parkinson. (2) Evaluación de la severidad de le enfermedad basada en una escala clínica tanto para la severidad motora y el nivel de disartria de los pacientes. (3) Clasificación de pacientes en diferentes grupos de acuerdo con su estado de severidad, por ejemplo, inicial, moderado, y severo. Los experimentos realizados cubren tanto esquemas tradicionales de reconocimiento de patrones además de modelos novedosos de aprendizaje profundo.

Se proponen tres enfoques para modelar la voz de pacientes con Parkinson: (1) análisis fonológico de la voz, el cual es más interpretable para los médicos al modelar directamente la información acerca del modo y manera de articulación. (2) Estrategias de aprendizaje por representación utilizando autoencoders recurrentes, los cuales tienen el potencial de extraer características más abstractas y robustas que aquellas tradicionalmente calculadas. Finalmente, (3) redes neuronales convolucionales entrenadas para procesar representaciones tiempo-frecuencia de la voz de los pacientes. Para el análisis de escritura, el enfoque propuesto envuelve el cálculo de características cinemáticas tradicionales, combinado con enfoques novedosos basados en características geométricas y en-aire. Modelos de aprendizaje profundo basados en redes neuronales convolucionales también se proponen para evaluar tanto las señales brutas de escritura online, así como las imágenes offline reconstruidas creadas por los pacientes. Los enfoques propuestos para el análisis de marcha envuelven el cálculo de características cinemáticas y espectrales, combinadas con un enfoque novedoso basado en dinámica no lineal. Un enfoque basado en aprendizaje profundo combinando redes neuronales convolucionales y recurrentes también se considera para modelar las señales de marcha de los pacientes. Finalmente, esta disertación cubre un análisis multimodal de señales de voz, escritura, y marcha. Los experimentos realizados consideran estrategias de fusión temprana y tardía. Los métodos propuestos son evaluados en dos escenarios: (1) usando sensores de alta calidad, y que pueden estar disponibles en centros clínicos para la evaluación de los pacientes, y (2) datos obtenidos de teléfonos inteligentes, los cuales pueden ser usados para el monitoreo continuo de los pacientes en casa. Los resultados indican que el análisis conjunto de diferentes bio-señales mejora aquellos obtenidos individualmente, tanto para la clasi-

ficación automática de la enfermedad y la evaluación de la severidad de los pacientes. Adicionalmente, los modelos propuestos son robustos para ser aplicados tanto en señales de alta calidad, como de aquellas obtenidas de teléfonos inteligentes.

Kurzdarstellung

Die automatische Analyse verschiedener Biosignale von Patienten mit Parkinson-Krankheit ist ein hochaktuelles Thema, das seit mehreren Jahren in der Forschungsgemeinschaft behandelt wird. Die Identifizierung von Biomarkern für die Früh- und Differenzialdiagnose, die Bewertung des Schweregrads und das Ansprechen auf die Therapie ist heute ein primäres Ziel der Forschung zur Parkinson-Krankheit. Es gibt wichtige Beiträge zu diesen Themen, die verschiedene Biosignale einzeln betrachten. Multimodale Analysen, d.h. die Berücksichtigung von Informationen verschiedener Sensoren, wurden nicht umfassend untersucht. Obwohl bei mehreren Aufgaben viele Verbesserungen gezeigt wurden, fehlt noch immer ein multimodales System, das in der Lage ist, eine genaue Vorhersage der Schwere der Erkrankung zu liefern und den Krankheitsverlauf zu überwachen. Das Ziel dieser Dissertation ist es, robuste Modelle für die genaue Diagnose der Parkinson-Krankheit zu entwickeln und die Schwere der Erkrankung von Patienten anhand verschiedener Biosignale wie Sprache, Online-Handschrift, Gang (mit Inertialsensoren) und der von Smartphones gesammelten Signale zu bewerten. Die vorgeschlagenen Modelle basieren auf drei Anwendungsszenarien: (1) Die automatische Klassifizierung von gesunden Probanden und Parkinson-Patienten. (2) Die Bewertung der Schwere der Erkrankung der Patienten basierend auf einer klinischen Skala, die sowohl die Schwere des motorischen Zustands als auch den Grad der Dysarthrie der Probanden umfasst. (3) Die Einteilung von Patienten in verschiedene Gruppen entsprechend ihrer Schwere der Erkrankung, z. B. leicht, mittelschwer und schwer. Die Experimente umfassten sowohl traditionelle Mustererkennung als auch neuartige Deep-Learning-Modelle.

Drei Ansätze werden vorgestellt, um die Sprache von Parkinson-Patienten zu modellieren: (1) phonologische Analyse der Sprache, die für Kliniker besser interpretierbar ist, indem Informationen über die Art und Weise der Artikulation direkt modelliert werden. (2) Repräsentation Lernstrategien unter Verwendung wiederkehrender Autoencoder, die das Potenzial haben, abstraktere und robustere Merkmale zu extrahieren als die herkömmlich berechneten. Schließlich (3) konvolutionale neuronale Netze, die darauf trainiert sind, Zeit-Frequenz-Darstellungen der Sprache der Patienten zu verarbeiten. In Bezug auf die Handschriftanalyse umfasst der vorgeschlagene Ansatz die Berechnung traditioneller kinematischer Merkmale, kombiniert mit neuartigen Ansätzen, die auf geometrischen und in der Luft befindlichen Merkmalen basieren. Deep-Learning-Modelle auf der Grundlage von Convolutional Neural Networks werden ebenfalls vorgeschlagen, um sowohl rohe Online-Handschriftsdaten als auch die von den Patienten erstellten rekonstruierten Offline-Bilder auszuwerten. Die vorgeschlagenen Ansätze zur Ganganalyse beinhalten die Berechnung traditioneller kinematischer und spektraler Merkmale, kombiniert mit neuartigen Ansätzen, die auf nichtlinearer Dynamik basieren. Ein Deep-Learning-Ansatz, der konvolutionelle und rekurrente neuronale Netze kombiniert, wird ebenfalls eingeführt, um die Gangsignale der Patienten zu modellieren. Schließlich befasste sich diese Arbeit mit einer multimodalen Analyse der von den Patienten gesammelten Sprach-, Handschrift- und Gangsignale. Die angesprochenen Experimente werden sowohl mit frühen als auch mit späten Fusionsstrategien durchgeführt. Die vorgeschlagenen Methoden werden auch in zwei Szenarien evaluiert: (1) hochwertige Sensoren, die in medizinischen Zentren zur Beurteilung der Patienten zur Verfügung stehen können, und (2) über

Smartphones gesammelte Daten, die zur kontinuierlichen Überwachung der Patienten verwendet werden können zu Hause. Die Ergebnisse zeigen, dass die kombinierte Analyse verschiedener Biosignale die mit jedem Biosignal erhaltenen sowohl bei der automatischen Klassifizierung der Krankheit als auch bei der Bewertung der Schwere der Erkrankung der Patienten übertraf. Darüber hinaus sind die vorgeschlagenen Modelle robust, um sowohl auf Signale angewendet zu werden, die mit hochwertigen Sensoren als auch auf Smartphones gesammelt wurden.

Acknowledgment

Many people contributed to the development of this work or made it possible at all. First, I want to thank my supervisors during this process, Prof. Dr.-Ing. Habil. Elmar Nöth, and Prof. Dr.-Ing. Juan Rafael Orozco-Arroyave, who has supported me within all these years, and highly contributed and encourage my scientific career. They are also responsible for several ideas and the quality of most of my papers and this dissertation.

Many thanks to all my colleagues in the speech processing group at FAU, especially to Tomas Arias, Philipp Klumpp, Paula Perez, Sebastian Bayerl, Christian Bergler, Hendrik Schröter for the nice discussions about different topics related to speech processing and deep learning, which help a lot to improve the quality of this dissertation. Thanks especially to Tomas, Philipp and Paula for their friendship and the fun times we spend together during this process. Thanks a lot to the colleagues and students of the GITA research group of the University of Antioquia in Colombia, especially to Cristian Rios and Daniel Escobar for their friendship and close collaboration always to perform experiments together.

Thanks to all my colleagues within the Training Network on Automatic Processing of PAthological Speech (TAPAS) for the great conversations we made during the training events, which allow to polish the ideas that are implemented in this dissertation. A special mention to Julian Fritsch and Viviana Mendoza for the work we have done together on modeling dysarthric speech. Special thanks to Prof. Maria Schuster from the Department of Otorhinolaryngology at Ludwig-Maximilians-University of Munich for her great feedback and discussions about the clinical aspects of my work. Her guidance helped a lot to improve the quality of this dissertation. Thanks also to Mathew Magimai-Doss to let me participate during a secondment at IDIAP in Switzerland, and for the great discussions we have done together.

My gratitude to all the students I have the opportunity to supervise during this process: Cristian Rios, Daniel Escobar, Paula Perez, Felipe Gomez, and Orlando Lopez at UdeA, and Martin Strauss, Souvik Tewari, Michael Küpfer, and Gabriel Miller at FAU. They contributed a lot in my formation as a leader, supervisor, and other aspects that are very important to pursue my scientific career.

My deepest gratitude to my family. They have been always with me, sharing successes and failures, supporting me. My special gratitude to my parents, Dora Luz and Luis Guillermo, without their continuous support this work would not have been possible. I am also very grateful to my sister Laura for the great support always. Finally, I owe my loving thanks to my wife Eliana, who supported me in so many ways. Thank you for your understanding over the last years due to the difficulties this process brought to us, and for make me always happy. Thank you for being you, and for being on earth to me. Thanks also for proof-reading the first chapter of this dissertation and give me feedback about the clinical aspects and motivation behind this work.

Juan Camilo Vasquez-Correa

Contents

Chapter 1

Introduction

1.1 Motivation

Parkinson's disease (PD) is a neurological disorder characterized by the progressive loss of dopaminergic neurons in the midbrain, producing several motor and non-motor impairments. PD affects approximately 10 million people worldwide, with a doubling of the global burden over the past 25 years because the increase in longevity of people and a longer disease duration thanks to modern medicine methods [Dors 18b]. In Europe, the prevalence of PD was estimated to be between 1280 and 1500 per 100,000 inhabitants by 2005 [Camp 05]. According to the Global Burden of the disease study, PD is the fastest growing neurological disease in terms of age-standardized rates of prevalence, disability, and deaths [Feig 17]. It is expected that by 2040 the incidence of PD will exceed 17 million patients in the world because the increased longevity and other factors [Feig 19].

The International Parkinson and Movement Disorder Society (MDS) defined a diagnostic criterion based on the presence of bradykinesia, resting tremor, and rigidity [Post 18]; however, these symptoms appear when the dopaminergic neurons in the striatum have been reduced by about 80% [Iran 18] and after roughly 50% of neurons in the substantia nigra have been irrevocably damaged [Duff 13]. In addition, it can take more than 20 years for the motor impairments to appear [Fere 19]. These reasons lead to a need for an early diagnosis of the patients, in order to provide them appropriate treatment before losing such high amount of neurons. The traditional assessment and diagnosis of the disease depends to some extent on the experience of the clinician performing the screening. This fact makes the determination of the exact type of disease as well as its degree of severity difficult. The rate of misdiagnosis of PD is high, especially when it is performed by a non-specialist neurologist. The probability of an inaccurate diagnosis can be up to 20%. This is particularly problematic for patients in the early stages of disease [Rizz 16]. These facts highlight the importance of being able to identify the earliest symptoms of PD in order to be able to treat the disease in the prodromal phase. It also requires the ability to evaluate how severe are the symptoms of a given patient by an accurate and consistent quantification method.

According to the Royal College of Physicians in London [Nati 06], in order to relieve the impact of the disease, PD patients should have access to specialized nurs-

ing care, physiotherapy, and speech and language therapy, in addition to the pharmacological treatment administered by clinicians [Wort 13]. All of these PD-related treatments exceed $US 303,000 per patient during the 12.8 years after diagnosis. The economic burden of PD could be significantly decreased if the disease progression is slowed down by at least 20% [John 13]. Moreover, many PD patients do not see a neurologist [Will 11]. Even in developed countries, the doctor appointments are once or twice per year e.g., in Sweden 1.7 times/year with regional variation between 1.1 and 2.1 [Lokk 11]. Additionally, accessibility to healthcare services is worse in rural regions [Will 11]. For all these reasons, identifying accurate bio-markers for early and differential diagnosis, severity, and response to therapy is a primary goal of the research on PD today. A systematic approach for continuous monitoring of the state of the patients will help in slowing down the impact of PD, and to improve the quality of life of patients.

By 2040, it is expected that prodromal evaluations will be incorporated into active neuroprotective treatment programs, followed by early treatment to slow down the incidence of the disease [Berg 18]. In order to achieve this goal, it is important to develop novel biomarkers using signal processing and machine learning approaches, in addition to the integration of smartphone and wearable technologies to develop progression biomarkers in early stages of the disease [Berg 18, Oroz 20b].

We believe that within the next decade, monitoring of motor and non-motor symptoms of PD patients will gradually shift from the clinic to at-home, where a continuous and non-intrusive monitoring can be performed. This monitoring is going to be addressed with wearable sensors and smartphone technology to monitor different symptoms of the disease, including motor impairments in the upper limbs, lower limbs, and in the speech production, in addition to non-motor impairments such as depression or sleep disorders. The developed technology is going to interact with the electronic health record of the patients in the clinic, and with platforms for population health analysis. The developed technologies are going to generate alerts if some specific behavior appears in the collected data, thus an expert neurologist will be able to prescribe a better and personalized treatment for each patient. This vision is summarized in Figure 1.1.

1.2 Parkinson's Disease

PD was first described by Dr. James Parkinson in [Park 17]. Dr. Parkinson defined the disease as "Involuntary tremulous motion, with lessened muscular power, in parts not in action and even when supported; with a propensity to bend the trunk forward, and to pass from a walking to a running pace: the senses and intellects being uninjured" [Park 17]. PD is a neuro-degenerative disorder that produces different motor and non-motor symptoms in the patients. Motor symptoms include tremor, slowed movement, rigidity, bradykinesia, lack of coordination, among others. Non-motor symptoms include depression, anxiety, sleep disorders, and olfactory dysfunctions, among others [Horn 98]. Approximately 70-90% of PD patients develop a multidimensional speech impairment called hypokinetic dysarthria [Loge 78], which manifests typically in the imprecise articulation of consonants and vowels, monoloudness, monopitch, inappropriate silences and rushes of speech, dysrhythmia, reduced

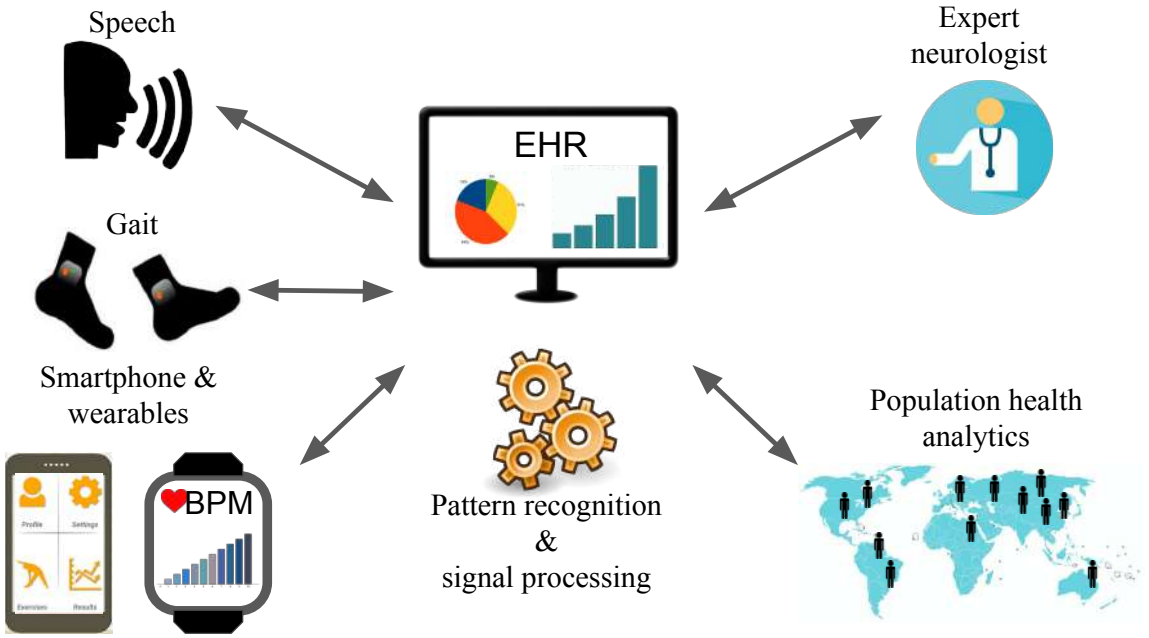

Figure 1.1: The future of digital medicine for monitoring of PD patients. **EHR**: electronic health records. **BPM**: beats per minute

vocal loudness, and harsh or breathy vocal quality. Usually, the motor symptoms are more or less evident from the onset of the disease. However, these motor impairments are different among the patients. This leads to one of the most challenging aspects to manage the disease progression and the treatment: each PD patient experiences the symptoms and the response to the treatment in a different manner.

The underlying cause for these symptoms is the spread of the protein α-synuclein throughout the peripheral and central nervous systems. The function of α-synuclein is to help regulate the release of dopamine, a type of neurotransmitter critical to control the start and stop of voluntary and involuntary movements [Horn 98]. For the case of PD patients, α-synuclein begins to accumulate along the peripheral and central nervous systems, being toxic to different cells. This ultimately leads to a loss of neural population, in particular, dopaminergic neurons in the substantia nigra and contralateral striatum; both being structures in the basal ganglia largely responsible to regulate both the reward network and motor systems [Rodr 09]. In the last two decades it also has been observed that the presence of PD is associated with a plethora of gastrointestinal symptoms produced in the gut and its associated neural structures [Sche 18]. This is of particular interest because the evidence suggests that these symptoms precede the motor impairments and the diagnosis of PD by several years, at least for a subgroup of patients [Sche 18]. This fact gives important insights about the origin of the disease and the possible detection of prodromal stages, which will help to develop novel neuroprotective therapies to halt or slow down the disease progression in early stages.

Despite the fact PD is not an infectious disorder, the disease exhibits many of the characteristics of a pandemic [More 09, Dors 18a], similar to what we are seeing now during the COVID-19 crisis. Pandemics extend over large geographic areas, and

PD is increasing in every major region of the world [Feig 17, Dors 18a]. Pandemics also tend to migrate, and PD appears to be shifting in response to changes in aging and industrialization [Dors 18a]. Similar to other pandemics, PD is experiencing exponential growth, and no one is immune to the condition [Dors 18a]. The PD pandemic seems to be fueled by several factors, including aging populations, increasing longevity and by the use of chemical industrial products [Dors 18a]. Particularly, it has been observed that the contact with numerous products such as specific pesticides, solvents, and heavy metals increase the incidence of PD [Gold 14]. Countries with the most increased industrialization in recent years like China have observed also the greatest increase in incidence of PD. For China, the growing rate of the disease was higher than for any other country in the world between 1990 and 2016 [Feig 19]. In addition, the use of specific pesticides linked to PD such as paraquat still persists in the world. For instance, although 32 countries have banned the use of paraquat, it is still exported and used in several countries like Brazil, Colombia, or the United States [Haki 16].

Besides the environmental factors described previously as a possible cause for PD, there is also evidence of some genetic mutations that make a person more susceptible to acquire the disease. The Leucine-Rich Repeat Kinase 2 (LRRK2) [Gilk 05] and the Parkin (PRKN) [Arki 18] are the most recognized genes associated with increasing a person risk for developing PD. Particularly, there is a well known mutation in the PRKN gene, which particularly produced one of the most extensive genetic clusters of early onset PD in a population in the countryside of Antioquia (Colombia) [Pine 06]. The genetic mutation seemed to be introduced to Antioquia by Spanish immigrants during the colonial times in the 16th century [Pine 06].

At the moment, there is no treatment to halt or to slow down the progression of PD, although there are several pharmacotherapeutic and neurosurgical options available to alleviate certain symptoms. Pharmacotherapeutic treatments include Levodopa as one of the most used medication to alleviate the motor symptoms; however, it seems that the medication is only effective in the early stages of the disease. The neurosurgical option is the deep brain stimulation, which consists of the implantation of electrodes in the brain. These electrodes are connected by wires to a type of pacemaker device (called an implantable pulse generator) placed under the skin of the chest, which creates electrical pulses to stimulate continuously the area of the brain that produces the motor symptoms. There are three areas in the brain that are used for deep brain stimulation in PD patients: the subthalamic nucleus, the globus pallidus internus, and the ventral intermediate nucleus of the thalamus. Each area plays a role in the brain's circuitry responsible for the control of movements. For many patients, the response to deep brain stimulation is similar to Levodopa but without the secondary symptoms associated to the medication such as dyskinesia.

1.3 Contribution to the Progress of Research

The vision about the use of technology based on signal processing, machine learning, and mobile computing for the assessment of PD motivates the development of this work. In particular we aimed to develop robust models for the accurate diagnosis and for the evaluation of the disease severity using different bio-signals such as speech,

online handwriting, gait, and those signals collected from smartphones. This work also aimed to model and understand different phenomena in the speech, handwriting, and general movement of patients affected with PD. In order to contribute to this aim, the following are the main outcomes of this work.

1. A multimodal corpus with speech, handwriting, and gait signals collected from 106 PD patients and 105 HC subjects, age- and gender-balanced was built. Additionally, the longitudinal corpus was built with a subset of 9 PD patients, who were recorded in up to 7 different session between 2012 and 2020 in order to evaluate the progress of the disease in long-term time intervals. At the same time, seven of the PD patients were included in the At-home corpus to monitor the progress of the speech deficits of PD patients in short-term periods of time.

2. The modified-Frenchay dysarthria assessment (m-FDA) is introduced as an alternative scale to evaluate the dysarthria severity of PD patients. The scale can be administered without the physical presence of the examiner by considering only speech recordings of the patients. The assessment of the dysarthria severity of the patients can be fully automated with the application of this scale, especially within short time intervals, where a phoniatrician is not available for the patients.

3. Two novel approaches are proposed to evaluate the speech of PD patients, both to accurately discriminate between PD and HC subjects and to predict the dysarthria severity of the patients. The two novel methods include: (1) phonological posterior features to model the pronunciation of different groups of phonemes based on the mode and manner of articulation of the Spanish language. The phonological features are created using recurrent neural networks with a multi-task learning strategy. (2) An unsupervised representation learning strategy using autoencoders to encode the most important information of the speech of PD patients.

4. Novel deep learning strategies are considered to model the speech of PD patients, both to detect the presence of the disease, and to evaluate the dysarthria severity of the patients.

5. Novel deep learning strategies are proposed to model the online handwriting and gait signals collected from PD patients, both to detect the presence of the disease, and to evaluate the neurological state of the patients.

6. Different fusion strategies were tested to combine the information from speech, online handwriting, and gait from PD patients. This leads to get more accurate models to detect the presence of the disease and to evaluate the severity of the patients.

7. The longitudinal evaluation of the disease progression of the patients is addressed following the methodology introduced in [Aria 18a], and which is based on GMM-UBMs to model features extracted from speech signals. The considered methodology is extended in this thesis to process additional features extracted from speech, handwriting, and gait signals.

8. I participated in the development of the mobile application *Apkinson*, which was developed to collect speech and movement data from PD patients, and to be used to monitor continuously the state of the patients using information from speech, hand movements, and fine-motor skills. The app was the main result of a Colombian - German project, financed by BMBF and COLCIENCIAS, in which 16 young researchers from both countries participated.

1.4 Structure of this Work

Chapter 2 describes the fundamentals of classical pattern recognition and novel deep learning techniques used in the scope of this work to train the classification and regression models based on speech, online handwriting, gait, and smartphone data collected from PD patients.

Chapter 3 starts with the description of the scales used to evaluate the severity of the PD patients, including the movement disorder society-unified Parkinson's disease rating scale (MDS-UPDRS) and the proposed m-FDA scale. Then it describes a list of existing databases from the state-of-the-art to model speech, handwriting, and gait of PD patients. Finally the chapter describes the different corpora considered for the experiments, including the multimodal, longitudinal, and at-home data; and the data collected using the Apkinson app.

Chapter 4 describes the methods considered to model speech signals from PD patients. The chapter includes a state-of-the-art review on methods for speech analysis of PD patients. Then, it describes classical acoustic analyses based on phonation, articulation, and prosody to model speech of PD patients, followed by the novel analysis based on phonological analysis of speech, and the proposed unsupervised representation learning strategy using autoencoders. The chapter finishes with the description of the end-to-end deep learning systems to model the speech signals of PD patients.

Chapter 5 describes the methods considered to model online handwriting signals from the patients. The chapter starts with an overview of current techniques to model handwriting data from PD patients. Then it describes methods based on kinematics, geometrical, and in-air analyses of online handwriting, followed by the description of end-to-end deep learning systems to model the handwriting data.

Chapter 6 describes the methods considered to model gait signals from the patients. The chapter starts with a review of techniques to model gait signals from PD patients using inertial sensors. Then it describes methods based on kinematics, spectral, and non-linear dynamics analyses of gait, followed by the description of end-to-end deep learning systems to model the gait data.

Chapter 7 provides a review on asynchronous multimodal systems to model the state of PD patients, followed by a set of early and late fusion techniques to combine the different speech, handwriting, and gait methods, described in the previous chapters.

Chapter 8 starts by describing the existing applications and technologies to model PD using smartphones. Then it includes a detailed explanation about the development of Apkinson and the main features included in the app.

Chapter 9 includes the details of the experiments that are addressed to evaluate the capability of the proposed methods to discriminate between PD and HC speakers, to evaluate the dysarthria severity of the speakers based on speech signals, and the evaluation of the neurological state using all bio-signals. At the end of this chapter there is an extensive discussion about the performed experiments and their results.

Chapter 10 presents an outlook on future research in the area of PD assessment using different bio-signals such as speech, handwriting, gait, and those collected with smartphone devices.

Chapter 11 summarizes the main insights about PD analysis using speech, handwriting, gait, and smartphone-based signals, and about the main experimental results.

Chapter 2

Theoretical Background

The suitability of the proposed models based on speech, handwriting, gait, and smartphone data is evaluated in three main scenarios: (1) to discriminate between healthy subjects and PD patients, (2) to predict the disease severity of the patients, based on a clinical scale, and (3) the classification of PD patients into different groups according to their disease severity e.g., mild, intermediate, and severe. For these applications, different pattern recognition methods for automatic classification and regression have to be considered. Two different strategies are considered to solve the classification and regression problems: (1) a classical pattern recognition approach using support vector machines (SVM) for classification and regression, and Gaussian mixture models (GMM) for longitudinal monitoring of the disease progression; and (2) a novel approach based on deep learning methods for an end-to-end analysis of the data collected from the patients. These two strategies are explained in detail in the following sections. The end of the chapter describes practical aspects addressed in this work such as cross-validation and hyper-parameter optimization strategies, which are used to validate and select the most robust models to evaluate the proposed approaches.

2.1 Classical Methods of Pattern Recognition

Traditional machine learning methods are considered to model extracted features from speech, handwriting, gait, and smartphone data collected from PD patients with the aim to solve different classification and regression problems. The classification problems consist of discriminating PD patients and HC subjects, and classifying PD patients in different stages of the disease based on their clinical evaluation. These classification problems are solved with support vector machine (SVM) classifiers. The regression problems consist of predicting the neurological scale of the patients based on the MDS-UPDRS-III score assigned by an expert neurologist, and predicting the level of dysarthria of the subjects according to the proposed m-FDA scale assigned by expert phoniatricians. The regression problems are solved using support vector regression (SVR) and GMMs adapted from universal background models (UBMs). Particularly, the GMM-UBM systems are considered to model the disease progression of the patients using longitudinal data. The following subsections explain in detail each of the addressed techniques.

2.1.1 Support Vector Machines

The SVM is a method introduced by the computer science community in the 1990s to solve classification problems [Bose 92]. SVMs have been shown to perform well in a variety of scenarios, especially for medical applications like pathological speech detection, where the high amount of data is often scarce. Despite the fact SVMs were intended for binary classification problems, they can be extended to solve problems with more than two classes i.e., such as the classification of patients in several stages of the disease, or adapted to solve a regression problem when the target variable is continuous.

Support Vector Machines for Classification

The classification of PD patients and HC subjects is evaluated with an SVM, which assigns the feature vector from a training set $\mathbf{x_i} \in \mathbb{R}^d$ with d number of features into one of the available training labels $y_i \in \{-1, +1\}$ i.e., PD or HC. The main aim of the SVM is to find the optimal separating hyper-plane to maximize the separability between the two classes. When the two classes are linearly separable we have the case of a hard-margin SVM. For this first case, the separating hyper-plane satisfies the following constraints:

$$b + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_d x_{id} \geq 1 \text{ if } y_i = +1, \tag{2.1}$$

and

$$b + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_d x_{id} \leq -1 \text{ if } y_i = -1. \tag{2.2}$$

Equivalently, a separating hyper-plane has the property from Equation 2.3 for all $i = 1, \cdots, N$. N is the number of training samples from the database.

$$y_i(b + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_d x_{id}) > 1 \tag{2.3}$$

In general, if the database is perfectly separated using a hyper-plane, there will be in theory an infinite number of such hyper-planes. The different hyper-planes can usually be shifted up or down, or rotated, without coming into contact with any of the feature vectors from the training set. The SVMs are focused on finding the optimal hyper-plane i.e., the one that is farthest from the training data. The process of finding such optimal hyper-plane starts by finding the perpendicular distance from the closest feature vectors to the separating hyper-plane. The distance to each feature vector $\mathbf{x_i}$ to the separating hyper-plane is given by $\frac{\langle \boldsymbol{\beta}, \mathbf{x_i} \rangle + b}{||\boldsymbol{\beta}||}$, where $||\cdot||$ denotes the Euclidean norm. This means that the closest feature vectors to the hyper-plane (those points where $y_i \cdot (\langle \boldsymbol{\beta}, \mathbf{x_i} \rangle + b) = 1$) will be separated a distance $\frac{1}{||\boldsymbol{\beta}||}$. Hence, the width of the margin from the closest feature vectors to the hyper-plane will be $\frac{2}{||\boldsymbol{\beta}||}$. These closest points to the hyper-plane are called support vectors. Figure 2.1 shows an example of the separating hyper-plane for 2-dimensional feature vectors.

The optimization problem to find the optimal hyper-plane that maximizes the margin in the training data is defined according to Equation 2.4.

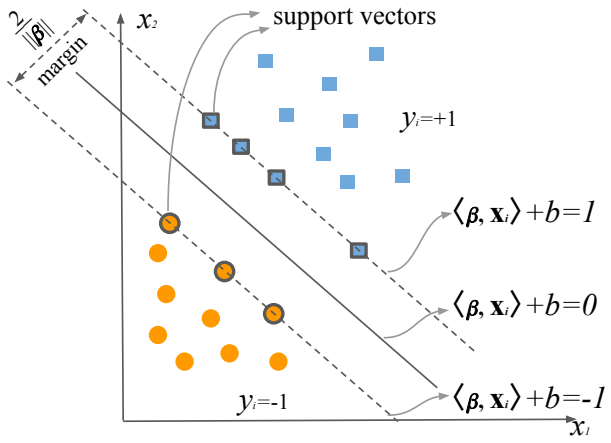

Figure 2.1: Hard-margin SVM

$$\begin{array}{ll} \underset{\boldsymbol{\beta},b}{\text{maximize}} & \frac{2}{||\boldsymbol{\beta}||^2} \\ \text{subject to} & y_i \cdot (\langle \boldsymbol{\beta}, \mathbf{x_i} \rangle + b) \geq 1,\ i = 1, \cdots, N \end{array} \tag{2.4}$$

The previous problem can be re-written as a constraint convex optimization problem, according to Equation 2.5.

$$\begin{array}{ll} \underset{\boldsymbol{\beta},b}{\text{minimize}} & \frac{1}{2} ||\boldsymbol{\beta}||^2 \\ \text{subject to} & y_i \cdot (\langle \boldsymbol{\beta}, \mathbf{x_i} \rangle + b) \geq 1,\ i = 1, \cdots, N \end{array} \tag{2.5}$$

This type of problems are typically solved by introducing the Lagrangian multipliers λ_i to include the restrictions as a linear combination into the objective function. Thus, the optimization problem can be re-formulated according to Equation 2.6. Our aim is now to minimize the Lagrangian of the primal problem $L_p(\boldsymbol{\beta}, b, \boldsymbol{\lambda})$.

$$L_p(\boldsymbol{\beta}, b, \boldsymbol{\lambda}) = \frac{1}{2} ||\boldsymbol{\beta}||^2 - \sum_{i=1}^{N} \lambda_i [y_i \cdot (\langle \boldsymbol{\beta}, \mathbf{x_i} \rangle + b) - 1] \tag{2.6}$$

The optimal conditions are found by setting the partial derivatives of $L_p(\boldsymbol{\beta}, b, \boldsymbol{\lambda})$ to zero, thus we obtain:

$$\frac{\partial L_p}{\partial \boldsymbol{\beta}} = \boldsymbol{\beta} + \sum_{i=1}^{N} \lambda_i y_i \mathbf{x_i} = 0 \quad \therefore \boldsymbol{\beta} = \sum_{i=1}^{N} \lambda_i y_i \mathbf{x_i} \tag{2.7}$$

$$\frac{\partial L_p}{\partial b} = -\sum_{i=1}^{N} \lambda_i y_i = 0 \quad \therefore \sum_{i=1}^{N} \lambda_i y_i = 0 \tag{2.8}$$

and substituting these in Equation 2.6 we obtained the so-called Wolfe dual optimization problem shown in 2.9. In addition, the solution must satisfy the Karush-Kuhn–Tucker conditions, which include Equations 2.7, 2.8, $\lambda_i \geq 0$, and $\lambda_i [y_i \cdot (\langle \boldsymbol{\beta}, \mathbf{x_i} \rangle + b) - 1] = 0$. The solution for L_D can be found using a standard optimization software.

$$L_D(\boldsymbol{\lambda}) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^\intercal \mathbf{x}_j \tag{2.9}$$

According to 2.7, the solution vector $\boldsymbol{\beta}$ is defined as a linear combination of the support vectors, i.e., those feature vectors $\mathbf{x_i}$ associated to those $\lambda_i > 0$, in order to fulfill the Karush-Kuhn–Tucker conditions. The bias term b is obtained by solving $\lambda_i[y_i \cdot (\langle \boldsymbol{\beta}, \mathbf{x_i} \rangle + b) - 1] = 0$ for any of the support vectors. Finally the optimal separating hyper-plane produces Equation 2.10 to classify the feature vectors from the test set.

$$f(\mathbf{x}) = \sum_{i=1}^{N} \lambda_i y_i \mathbf{x}^\intercal \mathbf{x}_i + b = 0 \tag{2.10}$$

Note that until now, the perfect separability of the training data is assumed. However, this is not the case in real-world scenarios. For those problems when the training data is not perfectly separable i.e., where there are errors that could be made by the machine, the soft-margin SVMs are considered, and explained as follows.

One way to deal with classes that are not perfectly separated is still to maximize the width of the margin, but allow for some points to be on the wrong side of the hyper-plane or inside the margin. This is done by the inclusion of positive slack variables $\xi \in \mathbb{R}^N$ to penalize those errors. Hence, our constraint from Equation 2.3, is re-formulated as follows.

$$y_i(b + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_d x_{id}) > 1 - \xi_i \tag{2.11}$$

The values of ξ_i measure the proportional amount by which the predictions from Equation 2.10 are on the wrong side of the margin. Miss-classification appear when $\xi_i > 1$, thus the sum $\sum_x i_i$ is an upper bound for the training errors. The cost associated to the miss-classification errors is included in the optimization problem from Equation 2.5, which is re-written as 2.12. The cost hyper-parameter C represents the chosen penalty to the errors. The case of the hard-margin SVM explained before corresponds to $C = \infty$.

$$\begin{array}{ll} \underset{\boldsymbol{\beta}, b, \xi}{\text{minimize}} & \frac{1}{2} ||\boldsymbol{\beta}||^2 + C \sum_{i=1}^{N} \xi_i \\ \text{subject to} & y_i \cdot (\langle \boldsymbol{\beta}, \mathbf{x_i} \rangle + b) \geq 1 - \xi_i \\ & \xi_i \geq 0 \end{array} \tag{2.12}$$

We can also define the Lagrange primal function re-formulating Equation 2.6 as Equation 2.13. ν_i corresponds to the Lagrange multipliers associated to the restriction $\xi_i \geq 0$.

$$L_p(\boldsymbol{\beta}, b, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \frac{1}{2} ||\boldsymbol{\beta}||^2 + C \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \lambda_i [y_i \cdot (\langle \boldsymbol{\beta}, \mathbf{x_i} \rangle + b) - 1 + \xi_i] - \sum_{i=1}^{N} \nu_i \xi_i \tag{2.13}$$

Similar to the case of the hard-margin SVM, we set the derivatives of $L_p(\boldsymbol{\beta}, b, \boldsymbol{\lambda}, \boldsymbol{\nu})$ to zero, thus we obtain:

$$\frac{\partial L_p}{\partial \boldsymbol{\beta}} = \boldsymbol{\beta} + \sum_{i=1}^{N} \lambda_i y_i \mathbf{x_i} = 0 \quad \therefore \boldsymbol{\beta} = \sum_{i=1}^{N} \lambda_i y_i \mathbf{x_i} \tag{2.14}$$

$$\frac{\partial L_p}{\partial b} = -\sum_{i=1}^{N} \lambda_i y_i = 0 \quad \therefore \sum_{i=1}^{N} \lambda_i y_i = 0 \tag{2.15}$$

$$\frac{\partial L_p}{\partial \xi} = C - \lambda_i - \nu_i = 0 \quad \therefore \lambda_i = C - \nu_i \tag{2.16}$$

By substituting these restrictions into the primal function, we also obtain the Wolfe Lagrangian dual objective function as 2.17. L_D is maximized subject to $0 \leq \lambda_i \leq C$ and $\sum \lambda_i y_i = 0$.

$$L_D(\boldsymbol{\lambda}) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^\intercal \mathbf{x}_j \tag{2.17}$$

The Karush-Kuhn–Tucker conditions for this problem include the conditions 2.14, 2.15, and 2.16, in addition to the constraints

$$\lambda_i [y_i \cdot (\langle \boldsymbol{\beta}, \mathbf{x_i} \rangle + b) - 1 + \xi] = 0 \tag{2.18}$$

$$\nu_i \xi_i = 0 \tag{2.19}$$

$$y_i \cdot (\langle \boldsymbol{\beta}, \mathbf{x_i} \rangle + b) - 1 + \xi \geq 0 \tag{2.20}$$

The support vectors for the case of the soft-margin SVM are those feature vectors that lie on the edge of the margin, and are associated by $\xi_i = 0$. Those feature vectors with $0 \leq \xi \leq 1$ are inside the margin but are on the right side of the hyper-plane i.e., they are well classified. Finally, those points associated to $\xi_i \geq 1$ are the miss-classification errors penalized by the hyper-parameter C. Given the solutions of the dual problem for b and $\boldsymbol{\beta}$ join with the tuning hyper-parameter C, the decision function of the soft-margin SVM is written according to Equation 2.21. Figure 2.2 shows an example of the soft-margin SVM for 2-dimensional feature vectors.

$$f(\mathbf{x}) = \sum_{i=1}^{N} \lambda_i y_i \mathbf{x}^\intercal \mathbf{x}_i + b = 0 \tag{2.21}$$

So far, we show that SVMs are powerful classifiers; however, in practice their computational and storage costs highly increase with the size of the training set. the computational cost of the SVM is associated to the quadratic programming solver used to find the support vectors, and usually ranges from $O(d \times N^2)$ to $O(d \times N^3)$ (d is the number of features and N is the size of the training set) depending on how efficient is the cache management of the solver (dataset dependent) [Chan 11].

Until now, we assume that the classes from the database are linearly separable with an hyper-plane. This restriction can be more flexible by increasing the feature space using expansion functions $\phi(\mathbf{x}_i)$ such as polynomials or splines. These functions transform the non-linear decision in the original space into a linear decision

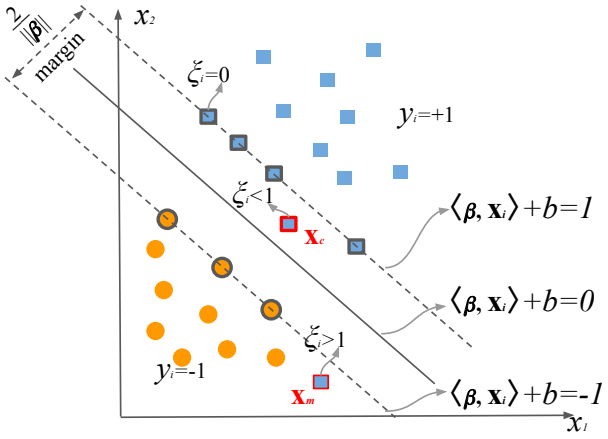

Figure 2.2: Soft-margin SVM. $\mathbf{x}_m$ corresponds to a miss-classified feature vector. $\mathbf{x}_C$ is a correctly classified feature vector which lies inside the margin.

in the expanded space, which will have a much higher dimension (infinite in some cases). The Lagrangian dual function L_D for the transformed feature vectors $\phi(\mathbf{x}_i)$ can be estimated in a similar way to the previous case, and has the form of Equation 2.22. The decision function for this dual problem is found and written according to Equation 2.23.

$$L_D(\boldsymbol{\lambda}) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle \tag{2.22}$$

$$f(\mathbf{x}) = \sum_{i=1}^{N} \lambda_i y_i \langle \phi(\mathbf{x}), \phi(\mathbf{x}_i) \rangle + b = 0 \tag{2.23}$$

Note that both Equations 2.22 and 2.23 involve $\phi(\mathbf{x})$ only through inner products, which indicates that we do not need to know the specific transformation function $\phi(\mathbf{x})$ but only the *Kernel* function $K(\mathbf{x}, \mathbf{x}_i) = \langle \phi(\mathbf{x}), \phi(\mathbf{x}_i) \rangle$. Such a kernel function computes the inner product in the transformed space. $K(\mathbf{x}, \mathbf{x}_i)$ should be a symmetric positive (semi-) definite function. The most common kernel functions used in the literature are the polynomial and Gaussian kernels, depicted in Equations 2.24 and 2.25, respectively. p and γ are hyper-parameters, and correspond to the order of the polynomial kernel and to the bandwidth of the Gaussian kernel, respectively.

$$K(\mathbf{x}, \mathbf{x}_i) = (1 + \langle \mathbf{x}, \mathbf{x}_i \rangle)^p \tag{2.24}$$

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\left\{ -\gamma \left|\left| \mathbf{x} - \mathbf{x}_i \right|\right|^2 \right\} \tag{2.25}$$

Support Vector Machines for Regression

The SVM paradigm can be adapted to solve regression problems were the target value is continuous ($y_i \in \mathbb{R}$). These new models are defined as SVRs. The core idea of SVRs is to minimize the prediction error $|y - f(\mathbf{x})|$ using the same principles of

SVMs designed for classification. For the case of the SVR we consider an hyper-tube to minimize a symmetric loss function with a ε-insensitive parameter to define a margin where some errors are tolerated. This is expressed according to Equation 2.26 and observed in Figure 2.3.

$$|y - f(\mathbf{x})| < \varepsilon \tag{2.26}$$

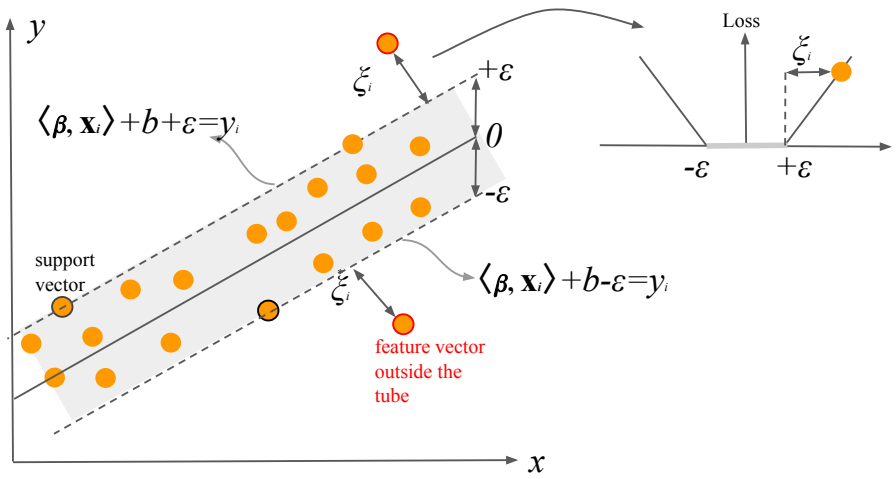

Figure 2.3: Support vector regressor

The optimization function for the SVR can be expressed according to Equation 2.27. C is the penalty hyper-parameter to determine the trade-off between $f(\mathbf{x})$ and the errors larger then ε that are allowed. Only those feature vectors for which $|y_i - f(\mathbf{x}_i)| > \varepsilon$ contribute to the loss function.

$$\underset{\boldsymbol{\beta}}{\text{minimize}} \ \frac{1}{2} ||\boldsymbol{\beta}||^2 + C \sum_{i=1}^{N} \{|y_i - f(\mathbf{x}_i)| > \varepsilon\} \tag{2.27}$$

The optimization problem from Equation 2.27 can be reformulated as a constraint optimization problem by the inclusion of slack variables ξ_i that count when $y_i - f(\mathbf{x}_i) > \varepsilon$, and ξ_i^* when $f(\mathbf{x}_i) - y_i > \varepsilon$. Hence the optimization problem is transformed to Equation 2.28.

$$\begin{aligned} \underset{\boldsymbol{\beta},,\xi,\xi^*}{\text{minimize}} \quad & \tfrac{1}{2} ||\boldsymbol{\beta}||^2 + C \textstyle\sum_{i=1}^{N} (\xi_i + \xi_i^*) \\ \text{subject to} \quad & \langle \boldsymbol{\beta}, \mathbf{x_i} \rangle + b - y_i \leq \varepsilon + \xi_i \\ & y_i - \langle \boldsymbol{\beta}, \mathbf{x_i} \rangle + b \leq \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0 \end{aligned} \tag{2.28}$$

The Lagrangian of the previous optimization problem is defined according to Equation 2.29. $\lambda_i, \lambda_i^*, \nu_i, \nu_i^*$ are the Lagrangian multipliers associated to the restrictions.

$$\begin{aligned} L_p = \quad & \tfrac{1}{2}\,||\boldsymbol{\beta}||^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) - \sum_{i=1}^{N}(\nu_i\xi_i + \nu_i^*\xi_i^*) \\ & - \sum_{i=1}^{N}\lambda_i(\varepsilon + \xi_i + y_i - \langle\boldsymbol{\beta}, \mathbf{x_i}\rangle - b) \\ & - \sum_{i=1}^{N}\lambda_i^*(\varepsilon + \xi_i - y_i + \langle\boldsymbol{\beta}, \mathbf{x_i}\rangle + b) \end{aligned} \tag{2.29}$$

Setting the partial derivatives of the primal problem to zero, we obtain the following expressions.

$$\frac{\partial L_p}{\partial \boldsymbol{\beta}} = \boldsymbol{\beta} + \sum_{i=1}^{N}(\lambda_i^* - \lambda_i)\mathbf{x_i} = 0 \quad \therefore \boldsymbol{\beta} = \sum_{i=1}^{N}(\lambda_i^* - \lambda_i)\mathbf{x_i} \tag{2.30}$$

$$\frac{\partial L_p}{\partial \xi_i} = C - \nu_i - \lambda_i = 0 \quad \therefore C = \nu_i + \lambda_i \tag{2.31}$$

$$\frac{\partial L_p}{\partial \xi_i^*} = C - \nu_i^* - \lambda_i^* = 0 \quad \therefore C = \nu_i^* + \lambda_i^* \tag{2.32}$$

$$\frac{\partial L_p}{\partial b} = \sum_{i=1}^{N}(\lambda_i^* - \lambda_i) = 0 \tag{2.33}$$

By substituting these partial derivatives in the primal problem, we obtain the dual problem according to Equation 2.34.

$$L_D = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\lambda_i^* - \lambda_i)(\lambda_j^* - \lambda_j)\langle\mathbf{x}_i, \mathbf{x}_j\rangle - \varepsilon\sum_{i=1}^{N}(\lambda_i^* - \lambda_i) + \sum_{i=1}^{N} y_i(\lambda_i^* - \lambda_i) \tag{2.34}$$

From the partial derivative with respect to the weights $\boldsymbol{\beta}$ in Equation 2.30 we found that the regression function for a sample from the test set can be written according to:

$$f(x) = \sum_{i=1}^{N}(\lambda_i^* - \lambda_i)\langle\mathbf{x}_i, \mathbf{x}\rangle + b \tag{2.35}$$

2.1.2 Gaussian Mixture Models - Universal Background Models

In previous studies [Aria 18a], we introduced the use of GMM-UBM systems to quantify the disease progression of PD patients. The main hypothesis is that if the speech of a patient is changing due to the disease progression, such changes can be modeled by comparing a model created for a patient in a specific session with respect to a reference model created with recordings of a group of speakers. The GMM-based systems are parametric probabilistic models represented as a linear combination of L Gaussian densities. For a feature vector $\mathbf{x}_i \in \mathbb{R}^d$ a GMM is defined according to Equation 2.36. The term ω_i represents the mixture weights. The probability densities $p_i(\mathbf{x}_i)$ are modeled as a multivariate Gaussian distribution with a mean vector

$\boldsymbol{\mu}_j \in \mathbb{R}^d$ and a covariance matrix $\boldsymbol{\Sigma}_j \in \mathbb{R}^{d\times d}$ [Reyn 00]. The parameters of the GMM for each Gaussian density are denoted as $\Lambda = \{\omega_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\}$, and they are estimated using the expectation maximization (EM) algorithm [Reyn 00].

$$p(\mathbf{x}_i|\Lambda) = \sum_{j=1}^{L} \omega_j p_j(\mathbf{x}_i) \tag{2.36}$$

$$\sum_{j=1}^{L} \omega_j = 1 \tag{2.37}$$

GMMs are used to represent the distribution of feature vectors extracted from a single speaker or a group of speakers. When the GMM is trained using features extracted from a large sample of speakers, the resulting model is a UBM. Ideally, the UBM is trained to represent the entire space of possible speakers. For a given set of speakers, the conditional probability $p(\mathbf{X}_{UBM}|\Lambda)$ is known as the maximum likelihood function that best represents the speaker's population. $\mathbf{X}_{UBM} \in \mathbb{R}^{N\times d}$ are the set of N feature vectors extracted from the group of speakers. The parameters Λ of the model are estimated using the expectation maximization (EM) algorithm, which increases the likelihood of the UBM $\mathrm{LL}(\mathbf{X}_{UBM}|\Lambda)$ in each iteration according to Equation 2.38.

$$\mathrm{LL}(\mathbf{X}_{UBM}|\Lambda) = \prod_{j=1}^{N} \sum_{i=1}^{L} \omega_i p_i(\mathbf{x}_j|\Lambda_i) \tag{2.38}$$

The model of the test speakers i.e., PD patients recorded in different sessions, is derived from the UBM by adapting its parameters Λ following a maximum a posteriori (MAP) process, which consists of a two step estimation process. The first step is the alignment of the feature vectors $\mathbf{x}_i$ to be adapted into each j-th component of the UBM, using Equation 2.39. Then, $\Pr(j|\mathbf{x}_t)$ and $\mathbf{x}_i$ are used to compute the sufficient statistics for the weights, the mean vectors, and the covariance matrices, using Equations 2.40, 2.41 and 2.42.

$$\Pr(j|\mathbf{x}_i) = \frac{\omega_j p_j(\mathbf{x}_i)}{\sum_{k=1}^{L} \omega_k p_k(\mathbf{x}_i)} \tag{2.39}$$

$$n_j = \sum_{i=1}^{N} \Pr(j|\mathbf{x}_i) \tag{2.40}$$

$$E_j(\mathbf{x}) = \frac{1}{n_j} \sum_{i=1}^{N} \Pr(j|\mathbf{x}_i)\mathbf{x}_i \tag{2.41}$$

$$E_j(\mathbf{x}^2) = \frac{1}{n_j} \sum_{i=1}^{N} \Pr(j|\mathbf{x}_i)\mathbf{x}_i^2 \tag{2.42}$$

The second step of the MAP process consists of using the sufficient statistics to update the parameters of the adapted model $\widehat{\Lambda} = \left\{\widehat{\omega}_j, \widehat{\boldsymbol{\mu}}_j, \widehat{\boldsymbol{\Sigma}}_j\right\}$ for the j-th mixture. The adapted parameters are computed using the following Equations for the weights,

means, and covariances, respectively. $\delta_{j\omega}, \delta_{j\mu}, \delta_{j\Sigma}$ are the adaptation coefficients to control the trade-off between the old and new estimates for the weights, means, and covariances, respectively. In addition, the scale factor r is considered to guarantee that $\sum \omega_j = 1$.

$$\widehat{\omega}_j = [\delta_{j\omega} n_j / N + (1 - \delta_{j\omega})\omega_j] r \tag{2.43}$$

$$\widehat{\boldsymbol{\mu}}_j = \delta_{j\mu} E_j(\mathbf{x}) + (1 - \delta_{j\mu})\mu_j \tag{2.44}$$

$$\widehat{\Sigma}_j^2 = E_j(\mathbf{x}^2) + (1 - \delta_{j\Sigma})(\Sigma_j^2 + \mu_j^2) - \mu_j^2 \tag{2.45}$$

Finally, the disease progression of the patients is estimated by comparing the adapted GMM model with respect to the UBM using a distance measure. We use the Bhattacharyya distance d_{Bha}, which considers differences in the mean vectors and covariance matrices between the UBM and the adapted models. d_{Bha} is defined by Equation 2.46, where $\widehat{\boldsymbol{\mu}}_j$ and $\widehat{\boldsymbol{\Sigma}}_j$ are the mean vector and the covariance matrix of the j-th component of the adapted model. $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_i$ are the parameters of the UBM [You 10]. The first term of Equation 2.46 measures the similarity between the mean vectors of the UBM and the adapted model. The second term measures the similarity between the covariance matrices.

$$d_{Bha} = \frac{1}{8} \sum_{i=1}^{L} \left\{ (\widehat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_i)^{\intercal} \left[\frac{\widehat{\Sigma}_i + \Sigma_i}{2} \right]^{-1} (\widehat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_i) \right\} + \frac{1}{2} \sum_{i=1}^{L} \left[\log \frac{\left| \frac{\widehat{\Sigma}_i + \Sigma_i}{2} \right|}{\sqrt{|\widehat{\Sigma}_i|\,|\Sigma_i|}} \right] \tag{2.46}$$

2.2 Deep Learning Methods

Deep learning is a subset of machine learning methods that has shown a lot success and has gained public attention in recent years, particularly because its ability to learn useful patterns from high dimensional non-structured data like images, video, speech, and natural language, among others. Deep learning methods usually are based on the concept of *Neural networks*, which are mathematical models inspired on how the brain processes information using multiple layers of abstraction to process the input information to make a decision.

Novel deep learning methods are considered to model the speech, handwriting, and gait data collected from PD patients and HC subjects. We propose different deep learning strategies based on convolutional and recurrent neural networks to process the speech, handwriting, and gait signals in an end-to-end fashion both to classify PD patients and HC subjects, and to evaluate the disease severity of the patients. We additionally propose two deep learning architectures to obtain robust speech features to model state of PD patients. The first architecture is designed to detect and extract phonological features related to the pronunciation of the patients of different groups of phonemes (see Section 4.3). The second one is designed two extract meaningful features based on a representation learning strategy using autoencoders to characterize the speech of PD patients (see Section 4.4). The following subsections

will explain the basic theory behind deep neural networks (DNN) and how they are considered to process the different biosignals collected from PD patients within the scope of this work.

2.2.1 Feed-Forward Neural Networks

The most common form of DNNs are the feed-forward neural networks, or *Multi-layer perceptrons* (MLPs). In this type of networks, the feature vector $\mathbf{x}_i \in \mathbb{R}^d$ is propagated forward via multiple processing layers until a set of output nodes. Figure 2.4 illustrates how the nodes are connected in the MLP to process the feature vectors. The nodes in the MLP are called *neurons*, and they are connected to process the information from the input vector. An example of a neuron of an MLP is shown in Figure 2.5.

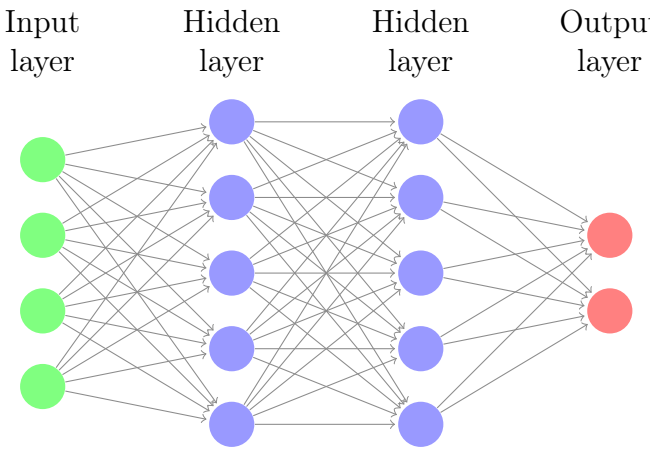

Figure 2.4: Feed-forward neural network with a depth of three layers. It consists of one *input layer*, two *hidden layers* and one *output layer*. All nodes are fully connected from the previous layer to the next one.

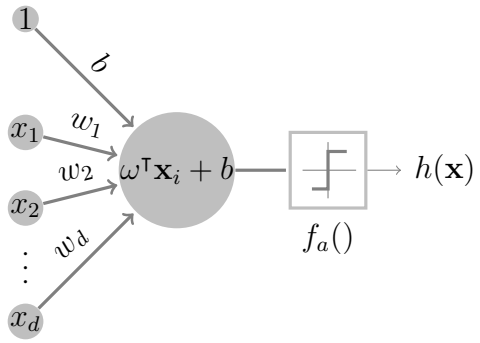

Figure 2.5: Single neuron. A linear combination of input values passes trough an activation function $f_a()$. b indicates the bias term. $h(\mathbf{x}_i)$ is the output value of the neuron.

A linear combination of the feature vector is computed with the weights $\boldsymbol{\omega} = \omega_1, \omega_2 \cdots, \omega_d$. b is the bias term. This linear combination then is processed by a non-

linear function $f_a()$ called *activation function* to create the output $h(\mathbf{x}_i)$. Formally, the feature vector is processed according to Equation 2.47 with a set of multiple neurons in one layer. $\mathbf{W} \in \mathbb{R}^{h \times d}$ is the weight matrix of the layer with h neurons, $b \in \mathbb{R}^h$ is the bias vector of the layer. For a neural network with more than one hidden layer, like the one shown in Figure 2.4 the output is expressed according to Equation 2.48. J is the number of layers and $\widehat{y}_i$ is the estimated output of the neural network.

$$h(\mathbf{x}_i) = f_a(\mathbf{W}^\intercal \mathbf{x}_i + \mathbf{b}) \tag{2.47}$$

$$\widehat{y}_i = h_J(h_{J-1}(\cdots h_2(h_1(\mathbf{x}_i)))) \tag{2.48}$$

There are different kinds of activation functions available. The sigmoid() or tanh() functions were the most popular when the topic was starting to gain attention [Good 16]. However, the introduction of the *rectified linear unit* (ReLU) [Nair 10] brought high improvements to deep learning. The ReLU function is piecewise linear, which makes it easy and fast to optimize (see Equation 2.49). In fact, Krizhevsky et al. [Kriz 12] reported a speed-up by a factor of six compared to a standard tanh() in their application. ReLU functions also help to prevent the vanishing gradient problem, which means that no further training is performed due to a neglectable gradient. However, if the input distribution tends to be more negative, the ReLU sets all activations to 0 which prevents further learning. This problem can be solved using the *Leaky-ReLU* [Maas 13] activation, where a slight slope factor α is applied (see Equation 2.50) in the negative value, which prevents the gradient from becoming 0.

$$f_a(x) = \max(0, x) \tag{2.49}$$

$$f_a(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha x & \text{otherwise} \end{cases} \tag{2.50}$$

When a classification problem needs to be solved, the network usually makes the decision according to the most likely output. Hence, it is beneficial to map the output values to a probability distribution to find the most probable output class. The activation function to perform this computation is the softmax(), which is expressed according to Equation 2.51.

$$\text{softmax}(\mathbf{x})_j = \frac{e^{\mathbf{x}_j}}{\sum_j e^{\mathbf{x}_j}} \tag{2.51}$$

Loss functions

The performance of the training process of a neural network is determined by the loss functions. They are a set of metrics designed to measure the difference between the output of the neural network and the expected value (label). The weights of the neural network are updated during the training process to minimize the loss function. One of the most common loss functions is the mean square error (MSE),

which is commonly used to solve regression problems. The MSE is defined according to Equation 2.52. The MSE loss penalizes large errors in the model and is insensitive to the small errors by the optimization of the $L2$-norm.

$$\mathcal{L}(y, \widehat{y}) = \frac{1}{N} \sum_{i=1}^{N} (y_i - \widehat{y}_i)^2 \tag{2.52}$$

An additional loss function used for regression problems is the Huber loss, described by Equation 2.53. It uses the scaled $L2$-norm if the error falls below 1 and the $L1$-norm in other cases. The Huber loss is less sensitive to outliers than the MSE and in some cases prevents exploding gradients [Girs 15].

$$\mathcal{L}(y, \widehat{y}) = \frac{1}{N} \sum_{i=1}^{N} \begin{cases} 0.5(y_i - \widehat{y}_i)^2 & \text{if } |y_i - \widehat{y}_i| < 1 \\ |y_i - \widehat{y}_i| - 0.5 & \text{otherwise} \end{cases} \tag{2.53}$$

Regarding classification problems, the most common loss function is the *cross-entropy*, which is designed to solve classification problems whose output is a probability value. Despite other loss functions like MSE penalize wrong predictions, cross-entropy gives a greater penalty when incorrect predictions have high confidence. The cross-entropy loss increases as the predicted probability diverges from the actual label. For binary classification problems the cross-entropy loss is defined according to Equation 2.54. For multi-class problems the loss function is extended to 2.55. N_c is the number of classes.

$$\mathcal{L}(y, \widehat{y}) = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(\widehat{y}_i) - (1 - y_i) \log(1 - \widehat{y}_i)] \tag{2.54}$$

$$\mathcal{L}(y, \widehat{y}) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{N_c} y_{i,c} \log(\widehat{y}_{i,c}) \tag{2.55}$$

Gradient-based optimization

The typical method to optimize the parameters of a neural network is the gradient descent, which minimizes the loss function in an iterative process by computing the gradient of the loss with respect to the parameters of the model. The optimization is performed following the negative gradient direction of the cost function $\mathbf{g}$ (see Equation 2.56). η is the learning rate, and denotes the step size taken at the k-th iteration during the optimization step.

$$\boldsymbol{\omega}_{k+1} = \boldsymbol{\omega}_k - \eta \mathbf{g} \tag{2.56}$$

In practice, the mini-batch gradient descent is used to approximate the gradient of the cost function in the training process. It replaces the actual gradient (computed from the entire dataset) with an estimation (computed from a randomly selected subset of the data). Such an estimation reduces the computational cost. The optimization using the mini-batch does not guarantee to reach the global minimum of the cost function, but it often finds a very low value fast enough to be useful. The core idea of the mini-batch gradient descent is that it is an expectation, which can

be approximated using a *batch* of samples, uniformly distributed from the training set. Hence, we can compute the gradient $\mathbf{g}$ from Equation 2.56 using Equation 2.57, where m is the batch size.

$$\mathbf{g} = \frac{1}{m}\nabla_\omega \sum_{i=1}^{m} \mathcal{L}(y_i, \widehat{y}_i) \tag{2.57}$$

In practice, it is necessary to gradually decrease the learning rate over time, so we now denote the learning rate at the k-th iteration as η_k. This is because the mini-batch gradient estimator introduces a source of noise due to the mini-batch sampling that does not vanish even when the minimum is reached. In practice, it is common to decay the learning rate linearly until the τ-th iteration, according to Equation 2.58, with $\alpha = k/\tau$ [Good 16].

$$\eta_k = (1-\alpha)\eta_0 + \alpha\eta_\tau \tag{2.58}$$

The most important property of the mini-batch gradient descent is that computation time per update does not grow with the size of the training data, which allows convergence even when the number of training examples is very large. For a large enough dataset, the mini-batch gradient descent may converge before it has processed the entire training set, within some fixed tolerance of the errors in the test set [Good 16].

Although the mini-batch gradient descent is mostly used for supervised learning because it is easy to compute the loss $\mathcal{L}(y, \widehat{y})$ between the actual and the predicted values, it can also be used for unsupervised learning, for instance when we train representation learning networks such as autoencoders [Vasq 20a]. In these scenarios, a common trick is to use the reconstruction loss $\mathcal{L}(\mathbf{x}, \widehat{\mathbf{x}})$ for the training process, as it is further explained in Section 4.4.

Adam optimizer

The Adam algorithm [King 14] is an adaptive learning rate optimization algorithm, which acts as an extension of the gradient descent. The name Adam is derived from *adaptive moment estimation*. The core idea of Adam is that the method computes individual adaptive learning rates for different parameters of the neural network according to the first and second moments of the gradients, shown in Equations 2.59 and 2.60, respectively. β_1 and β_2 are hyper-parameters known as the exponential decay rate for the first and second moment of the gradient, respectively, and they were originally defined as $\beta_1 = 0.9$ and $\beta_2 = 0.999$ [King 14]. These values for β_1 and β_2 have become standard for the community. For additional information about their inference and validation, the reader may refers to [King 14].

$$m_k = \beta_1 m_{k-1} + (1-\beta_1) g_k \tag{2.59}$$

$$v_k = \beta_2 v_{k-1} + (1-\beta_2) g_k^2 \tag{2.60}$$

Then, the update rule for the parameters of the neural network using Adam is shown in Equation 2.61. $\epsilon = 10^{-8}$ is chosen to avoid the division by 0. An overview

of other optimization methods based on gradient descent to train neural networks can be read on [Rude 16].

$$\boldsymbol{\omega}_{k+1} = \boldsymbol{\omega}_k - \frac{\eta}{\sqrt{v_k} + \epsilon} m_k \tag{2.61}$$

Backpropagation

The core idea of training a neural network is to update the weight matrix associated to each layer according to the current loss function. The backpropagation algorithm was proposed for such a purpose. It is an efficient way to compute the gradients with respect to the weights based on the chain rule and dynamic programming. The basic procedure is as follows:

1. **Forward pass:** Propagate the input mini-batch through the network in a layer-by-layer fashion to compute all activations and get the loss at the output layer.

2. **Backward pass:** Recursively apply the chain rule to propagate backwards through the network and compute all gradients. In this way each neuron weight is updated according to its contribution to the total loss.

3. Repeat the forward and backward steps iteratively for the different mini-batches from the training set until convergence of the loss function.

Figure 2.6 shows the basic concept of backpropagation. The green lines indicate the activations computed in the forward pass until the loss function $\mathcal{L}$ is computed. Consequently, the chain rule is applied to compute the gradient with respect to the output y and it is propagated backwards through the network (red lines).

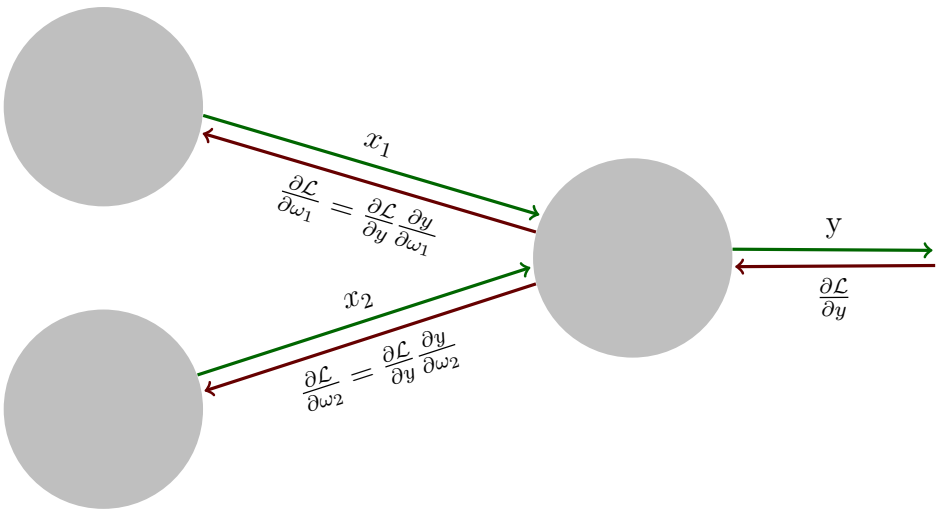

Figure 2.6: Illustration of the backpropagation algorithm. Green lines denote the activations computed in the forward pass. In the backward pass (red lines) the chain rule is used to compute the gradients. Adapted from the Stanford CS231n lecture notes [Li 18].

2.2.2 Convolutional Neural Networks

Convolutional neural networks (CNNs) are a specific case of neural networks that use convolutional operators instead of matrix multiplications in at least one of their layers [Good 16]. The fully connected layers in MLPs are replaced by convolutional layers. Hence, Equation 2.47 is replaced by 2.62 in CNNs.

$$h(\mathbf{x}_i) = f_a(\mathbf{W} \circledast \mathbf{x} + \mathbf{b}) \tag{2.62}$$

Convolutional layers can deal with 1, 2, or 3-dimensional signals. 1-dimensional signals include raw speech frames, online handwriting signals, or gait signals. 2-dimensional signals may include gray-scale images with only one channel, color images with three channels (red, green, blue), or time-frequency representations of speech signals. Finally, 3-dimensional signals may include video frames transmitted over time. The main advantage of CNNs is the local-connectivity i.e., each of the CNN filters captures a local context, which helps to get more accurate models and to reduce complexity because it requires less parameters. The discrete convolution operation used in CNNs is defined according to Equations 2.63 and 2.64 for the case of 1, and 2-dimensional signals, respectively.

$$\boldsymbol{\omega} \circledast \mathbf{x}[i] = \sum_{j=1}^{N} x[j] \cdot \omega[i-j] \tag{2.63}$$

$$\mathbf{W} \circledast \mathbf{X}[i,j] = \sum_{k=1}^{N_1} \sum_{l=1}^{N_2} x[k,l] \cdot \omega[i-k, j-l] \tag{2.64}$$

Given the nature of the convolution operation, which process the signals with CNN filters, make these layers particularly useful to learn time-invariant representations in 1D signals like speech waveforms, or time-frequency invariant in 2D signals such as spectrograms). Due to that reason, CNNs are commonly used in lower layers of the model because these are powerful to encode relevant local stationarities - like sinusoids in waveforms [Diel 14], or frequency traces in spectrograms.

After each convolutional layer, there is usually a pooling layer that down-samples the hidden representation in order to compress the feature space and use only the most relevant information. A pooling function replaces the output of the layer at a certain location with a summary statistic of the nearby outputs [Good 16]. For example, the *max pooling* operation reports the maximum output within a rectangular neighborhood. Pooling operation helps to make the representation approximately invariant to small translations of the input. Translation invariance means that if we shift or rotate the input by a small amount, the values of most of the pooled outputs do not change. Invariance to local translation can be a very useful property if we care more about whether some feature is present than exactly where it is [Good 16] e.g., we want to detect the presence of dysarthria in a speech frame but not the time-frame where it is present. Pooling layers also improve the computational efficiency of the network because the next layer has roughly k times fewer inputs to process (being k the size of the pooling neighborhood). Figure 2.7 shows a typical CNN architecture to process a spectrogram, consisting of 2 convolutional layers, 2 pooling layers and a classification layer formed with an MLP.

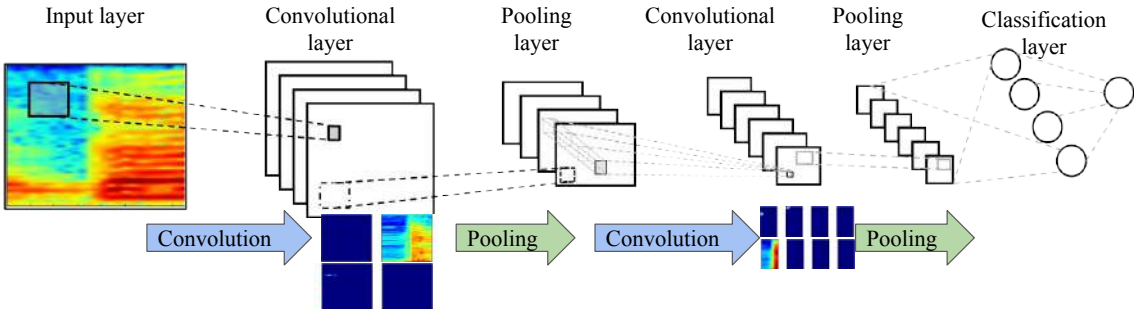

Figure 2.7: Typical structure of a CNN.

Residual layers

One of the architectures that has become the state-of-the-art for CNNs is ResNet [He 16]. The architecture uses residual layers to provide a solution to the vanishing gradient problem, which appears when deeper models are considered. Instead of training a set of stacked layers to directly fit an underlying distribution, the ResNet-based models fit a residual mapping. Formally, denoting the expected output of the layer as $h(\mathbf{x})$, the output of the residual block is designed to learn the function $\mathcal{F}(\mathbf{x}) = h(\mathbf{x}) - \mathbf{x}$. Hence, the desired function can be recovered as $\mathcal{F}(\mathbf{x}) + \mathbf{x}$. The residual function is easier to optimize than the originally expected. The formulation of $\mathcal{F}(\mathbf{x}) + \mathbf{x}$ can be implemented via *skip connections* (see Figure 2.8). These skip connections make it easier for the gradient to flow from output layers to layers nearer the input. The skip connections perform an identity mapping, whose output is added to the output of the convolutional layers.

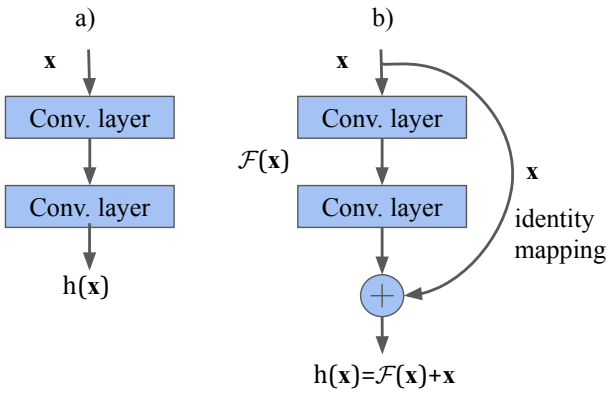

Figure 2.8: Comparison between **a)** a normal convolutional block and **b)** a residual block.

2.2.3 Recurrent Neural Networks

RNNs have been proposed to model a sequence of feature vectors $X = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_t, \cdots, \mathbf{x}_T\}$. These networks are designed in a way such that they have an output $h_t()$ that depends on both the feature vector in a time instant $\mathbf{x}_t$ and the output in the

previous time instant $h_{t-1}()$. Equation 2.65 represents the output of a hidden layer in an RNN in a time t.

$$\mathbf{h}_t(\mathbf{x}) = \tanh(\mathbf{W}[\mathbf{x}_t, \mathbf{h}_{t-1}] + \mathbf{b}) \tag{2.65}$$

The weight matrix $\mathbf{W}$ is learned to process the input and the previous output state of the RNN, respectively. These weights are shared across the different time steps of the network. If we had separate parameters for each time step, the network would not be able to generalize to sequence lengths not seen during training, nor to share statistical strength across different sequence lengths and across different positions in time [Good 16]. Such weight sharing is particularly important when a specific piece of information can occur at multiple positions within the sequence [Good 16]. The tanh() is usually considered because it does not vanish easily during the back-propagation trough time, as it occurs for other activation functions like the sigmoid().

Figure 2.9 shows the basic structure of an RNN. The sequence of feature vectors is fed to the recurrent layer, producing the output sequence $\mathbf{h}$, which is then used as input for the RNN in the next time steps. The hidden outputs are finally processed by an activation function to produce the output sequence $\mathbf{y}$.

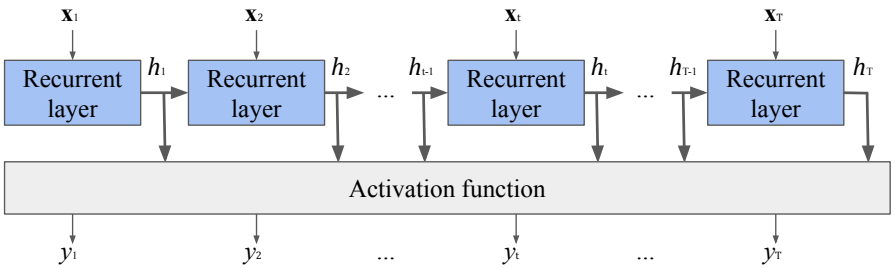

Figure 2.9: General scheme of a basic RNN.

The recurrent layer from Figure 2.9 has a *causal* structure, i.e., the state at time t only captures information from the past $\mathbf{x}_1, \cdots \mathbf{x}_{t-1}$ and the present input $\mathbf{x}_t$. However, in many applications it is important for the predictions to have information also from the future time steps. For instance, in speech recognition, the mapping of the current sound into a phoneme may depend on the next few phonemes because of co-articulation and potentially may even depend on the next few words because of the linguistic dependencies between nearby words [Good 16]. This is also true for handwriting modeling and many other sequence-to-sequence learning tasks, such as those addressed in this work. Bidirectional RNNs were created to address that need [Schu 97], and have been very successful in tasks such as speech recognition, handwriting modeling, machine translation, among others. Bidirectional RNNs combine an RNN that processes the input sequence forward through time with an additional RNN that moves backwards the input sequence. The output of the forward and backward RNNs is combined via addition, multiplication, or concatenation operation, as it is shown in Figure 2.10.

Traditional RNNs exhibit a vanishing gradient problem, which appears when modeling long temporal sequences. The problem occurs because the repeated application of the same operation at each time step of a long temporal sequence produces very

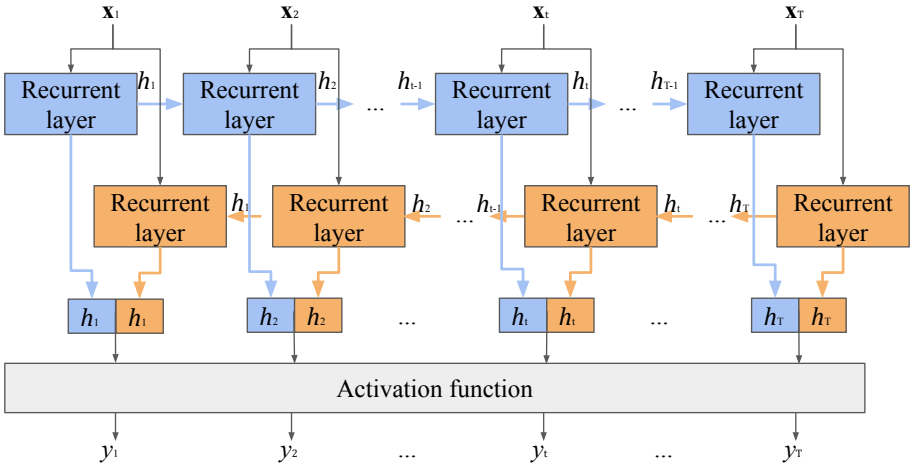

Figure 2.10: General scheme of a bidirectional RNN.

small gradients, which are vanish in the backward pass due to the chain rule, thus the weights of the network are not properly updated. In other words, during back-propagation through time, as gradients are calculated by the chain rule, the small numbers produced by the derivatives of the tanh() function are multiplied T times, which squeezes the final gradient to almost zero, thus there are no changes to update the weights of the RNN. Another problem that appears in traditional RNNs is the exploding gradients, which is associated to the consecutive multiplication by the weight matrix in each time step when the weights are too large. This causes a blowing up of the gradients, which makes the RNN training highly unstable. Several methods were proposed to address the problem of vanishing and exploding gradients in RNNs, such as adding skip-connections through time [Lin 98], or including *leaky units* to integrate signals with different time constants [Moze 92]. However, the most successful approach to solve the training problems in RNNs was the introduction of *gated RNNs*. These include the long short-term memory units (LSTMs) [Hoch 97], and the gated recurrent units (GRUs)[Cho 14].

Long Short-Term Memory Units

The core idea of LSTMs to solve the training problems of traditional RNNs is the inclusion of a *long term memory* using self-loops to produce paths where the gradient can flow for long term duration sequences [Hoch 97]. This self-loop is controlled by another hidden unit, which makes the memory of the LSTM to have a dynamical temporal context. The LSTM architecture consists of a set of four recurrently connected sub-networks, known as memory blocks. Each block contains one or more self-connected memory cells and three multiplicative units: the input, output and forget gates, which work analogously to write, read and reset operations for the cells [Grav 12].

The most important component in the LSTM block is the *state unit* s_t, which defines the self-loop for the long-term memory of the block. It is controlled by the *forget gate* f_g, which activates the state unit via a sigmoid function $\sigma()$ according

to Equation 2.66. $\mathbf{x}_t$ is the input vector in the present time, and $\mathbf{h}_{t-1}$ is the hidden output at the previous time instant. $\mathbf{W}_f$ and $\mathbf{b}_f$ are the weights and bias terms of the forget block.

$$f_g = \sigma\left(\mathbf{W}_f[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f\right) \tag{2.66}$$

The input gate i_g is used to control how much information is used to update the state unit of the LSTM block. The output of the input gate is obtained according to Equation 2.67, with the associated weights and bias terms $\mathbf{W}_i$ and $\mathbf{b}_i$, respectively.

$$i_g = \sigma\left(\mathbf{W}_i[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f\right) \tag{2.67}$$

The state unit of the LSTM block is updated using to Equation 2.68, according to the output values from the input gate i_g, and the self-loop controlled with the forget-gate.

$$s_t = f_g s_{t-1} + i_g \tanh\left(\mathbf{W}_c[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_c\right) \tag{2.68}$$

Finally, the output gate is used to control how much information is passed to the output of the LSTM block. The output gate is activated using Equation 2.69, in a similar way to the input and forget gates, but with its own parameters. The output of the LSTM block is then computed with Equation 2.70.

$$o_g = \sigma\left(\mathbf{W}_o[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o\right) \tag{2.69}$$

$$\mathbf{h}_t = o_g \tanh(s_t) \tag{2.70}$$

Figure 2.11a) shows an LSTM memory block with a single cell. The scheme is compared with the one obtained for a traditional RNN block in Figure 2.11b). An LSTM network is the same as a standard RNN, except that the summation units in the hidden layer are replaced by memory blocks [Grav 12]. Each cell has the same inputs and outputs as a traditional RNN, but has four times more parameters and a system of gating units that controls the flow of information [Good 16].

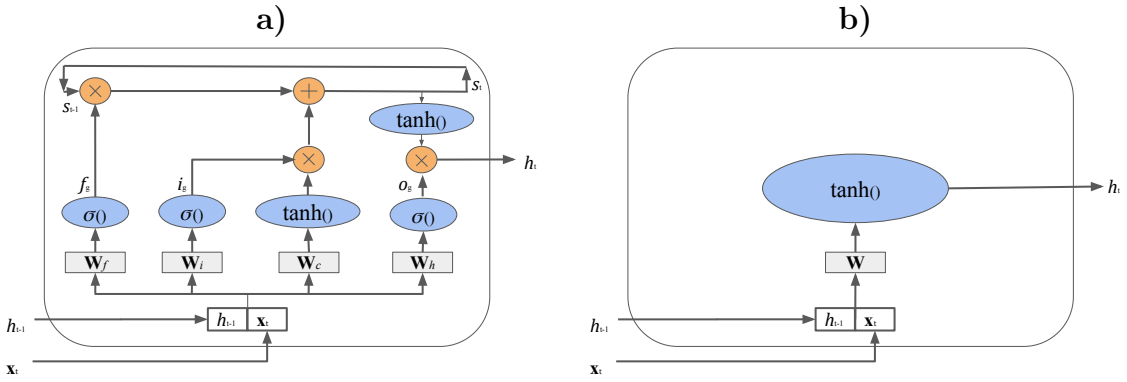

Figure 2.11: **a)** General scheme of an LSTM block. **b)** General scheme of an RNN block. $\sigma()$ represents sigmoid activation functions, and the symbol $\times$ denotes a matrix multiplication.

LSTMs have a much cleaner backpropagation compared to traditional RNNs. This structure avoids the vanishing and exploding gradient problems. The main reason is because there is no multiplication with the weight matrices during the backward pass, but only an element-wise multiplication with the forget gate, thus the complexity in the backward pass is reduced.

Gated Recurrent Units

The GRUs were introduced in [Cho 14] as an alternative to the LSTMs with the aim to reduce the number of parameters to learn, but keeping the core idea of gates to control the flow of information and to prevent vanishing and exploding gradients. The main difference between the GRU and the LSTM is that a single *update gate* unit u_g simultaneously controls the forgetting factor and the decision to update self-loop of the state unit. Conversely to the LSTM that has four gates, the GRU only has two: the update gate and the reset gate r_g, which acts similar to the forget gate of the LSTM. Figure 2.12 shows the main structure of the GRU block. Note that it includes only three weight matrices compared to four in the LSTM.

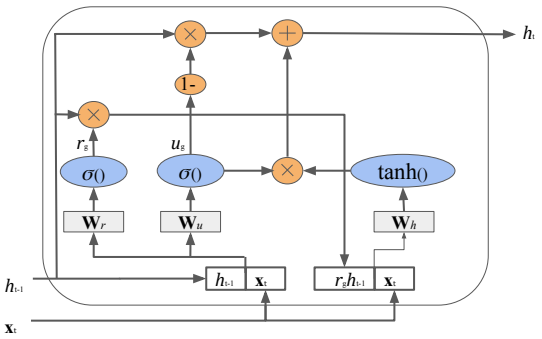

Figure 2.12: General scheme of a GRU block. $\sigma()$ represents sigmoid activation functions, and the symbol $\times$ denotes a matrix multiplication.

The update gate acts similar to the forget and input gate of an LSTM, and its output is computed by Equation 2.71. The update gate helps the GRU to determine how much of the past information needs to be transferred to the future.

$$u_g = \sigma\left(\mathrm{W}_u[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathrm{b}_u\right) \tag{2.71}$$

The reset gate is used in the GRU block to decide how much of the past information to forget, similar to the forget gate in the LSTM. The output of the reset gate is computed by Equation 2.72.

$$r_g = \sigma\left(\mathrm{W}_r[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathrm{b}_r\right) \tag{2.72}$$

Finally, the output of the GRU block is computed by Equation 2.73. The element-wise multiplication between the reset gate and $\mathbf{h}_{t-1}$ determines what information to forget from the previous time steps.

$$\mathbf{h}_t = (1 - u_g)\mathbf{h}_{t-1} + u_g \cdot \tanh(\mathrm{W}_h[r_g\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathrm{b}_h) \tag{2.73}$$

2.2.4 Regularization in Deep Learning

One of the biggest issues to train deep learning models is to guarantee generalization to new test data that appear in the final application. As we move towards more complex deep learning models, the method learns very well details and also noise from the training data, which ultimately results in poor performance on the unseen test data. Regularization comprises a set of techniques that make slight modifications to the learning algorithm to improve the generalization of the models, which is translated into an improvement in the performance on unseen test data. Different regularization methods have been proposed in the literature to improve generalization and to avoid over-fitting. Four different regularization strategies are considered within the scope of this work to obtain models that are more robust to process unseen data not included in the training set. Thus to create models suitable for the clinical practice.

Early Stopping

It is commonly observed in the training process of deep learning methods that the training error is reduced over the iterations, but in some cases the development error starts to increase. This behavior indicates that the model is starting to over-fit the training data. The core-idea of early-stopping is to monitor the loss in the development data, and when such loss starts to increase after a number of epochs, then we stop the training process (see Figure 2.13). The number of epochs with no further reduction of the development loss is usually known as *patience*, and it is an indicator of when to stop the training process.

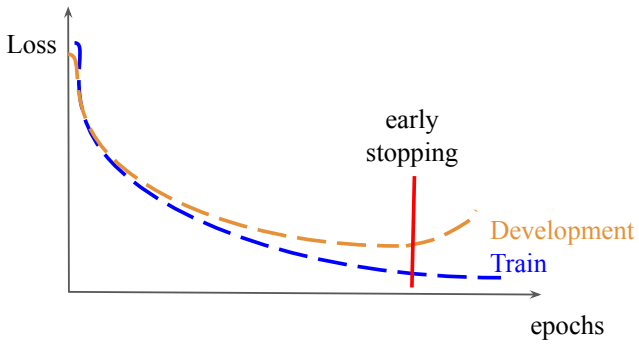

Figure 2.13: Early stopping strategy.

Dropout

This regularization method turns-off randomly a set of neurons in the model, according to a fixed probability value. On each training iteration different neurons are deactivated, therefore they no longer contribute neither to the forward nor the backward pass. Dropout helps to prevent that nearby neurons depend to each other, thus they have to learn the weights in a more independent way without considering the values of the neighbor neurons. The training process using dropout includes a hyper-parameter p_{drop}, which defines the probability of a neuron to be turned-off during each iteration. An additional consequence of dropout is that it roughly doubles the

number of iterations required to converge. However, training time for each epoch is also reduced, which compensates the training time. Similar to other regularization methods, dropout is more effective on those problems where there is a limited amount of training data, which is the typical scenario in clinical applications like the ones considered in this work.

Batch Normalization

Batch normalization [Ioff 15] is one of the most used methods to optimize the training process of deep learning models. It is a reparametrization technique to normalize the activation in intermediate layers of deep learning models [Bjor 18] using a z-score standardization function, according to Equation 2.74. The output of the hidden layer $\mathbf{h}(\mathbf{x})$ is standardized using statistics from the mini-batch μ_c and σ_c. The standardization is controlled by the parameters γ_c and β_c, which are learned during the training process. ϵ is a small value used for numerical stability.

$$\mathbf{h}(\mathbf{x}) = \gamma_c \frac{\mathbf{h}(\mathbf{x}) - \mu_c}{\sqrt{\sigma_c^2 + \epsilon}} + \beta_c \tag{2.74}$$

Batch normalization reduces the amount by which the hidden unit values shift around (internal-covariance shift). This has the effect of stabilizing the learning process and reducing the number of training epochs required to train the model [Ioff 15]. Batch normalization acts also as regularizer because the standardization of each mini-batch using its mean and standard deviation introduces some noise to each layer, providing a regularization effect that reduces over-fitting.

L2 Regularization

This is one of the most common methods to regularize the weights of deep learning models. It is also known as *Tikhonov regularization*. This method is used to regularize the cost function associated to the weights $\boldsymbol{\omega}$ by adding the term $\frac{1}{2}\left|\left|\boldsymbol{\omega}\right|\right|^2$ to penalize those weights with higher values. Due to the addition of the regularization term, the values of weight matrices decrease since it assumes that a neural network with smaller weight matrices leads to simpler models, thus reducing over-fitting. The regularization level is controlled by an additional hyper-parameter ζ. Hence, the cost function $\mathcal{L}$ is updated according to Equation 2.75. The update rule for the weights from Equation 2.56 is also re-formulated using Equation 2.76. L2 regularization is also known as *weight decay* as it forces the weights to decay towards zero (but not exactly zero).

$$\mathcal{L}(\boldsymbol{\omega})' = \mathcal{L}(\boldsymbol{\omega}) + \zeta \frac{1}{2}\left|\left|\omega\right|\right|^2 \tag{2.75}$$

$$\boldsymbol{\omega}_{k+1} = \boldsymbol{\omega}_k - \eta\mathbf{g} - \zeta\boldsymbol{\omega}_k \tag{2.76}$$

2.3 Experimental Evaluation

This section describes the strategies considered to validate and to optimize the proposed models based on classical pattern recognition and deep learning methods both to classify PD patients vs. HC subjects and to evaluate the severity of the disease of the patients.

All classification and regression models addressed in this work are validated following a nested 10-fold cross-validation strategy, which consists of dividing the feature sets extracted from all subjects into two partitions, one used to train and optimize the hyper-parameters of the algorithms, and the other one is used for test. At the same time, the training partition is divided into 9-folds, using 8 of them to train the models and the remaining one to optimize the hyper-parameters of the learning algorithm i.e., development set. A summary of this process is shown in Figure 2.14. The cross-validation is always performed subject-independent, i.e., it is guaranteed that all samples collected from the same subject are always in the same partition, and they are not mixed in the train and test set. The 10-folds are also randomly selected, but with the same seed in order to guarantee that all experiments are comparable to each other. In addition, the training sets for the classification problems are stratified i.e., we consider equal number of samples from both classes in order to have a balanced training data.

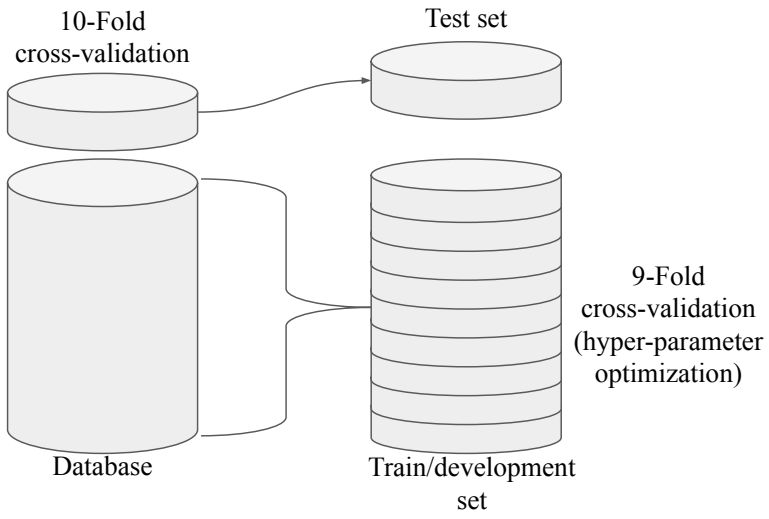

Figure 2.14: Distribution of the database into a nested 10-fold cross-validation.

The hyper-parameters of the learning algorithms are optimized based on the performance obtained on the development set. For the case of the SVM, the complexity hyper-parameter C, the bandwidth of the Gaussian kernel γ and the ε hyper-parameter for the loss function in the SVR are optimized in a randomized search strategy, which is a slight variation on grid search. Instead of searching over the entire grid of possible values of hyper-parameters, the randomized search only evaluates a random sample of points on the grid. This makes the optimization computationally cheaper than the full grid search optimization. Bergstra and Bengio show in [Berg 12] that in many scenarios, randomized search performs about as well as the full grid

search optimization. Even they show that using only 60 randomized samples from the hyper-parameters, there is a probability of 0.95 of reaching the global optimum with an error of 5%. To derive the previous conclusion, imagine the 5% interval around the global optimum. Now imagine that we generate sample points from the entire space and see if any of them lies within the selected optimal interval. Each random sample has a 5% chance of landing in the optimal space, thus if we select n points independently, the probability that all of them miss the desired interval is $(1 - 0.05)^n$. Hence, the probability that at least one of the n samples lies in the optimal interval is $1 - (1 - 0.05)^n$. If we want at least a 0.95 probability of success, we solve Equation 2.77 for n, and we get $n = 60$.

$$1 - (1 - 0.05)^n > 0.95 \tag{2.77}$$

The randomized search optimization is performed as follows: the values of the hyper-parameters C, γ, and ε are modeled with an exponential probability density function, which generates values for each hyper-parameter to be evaluated according to the performance in the development set. After several iterations with different generated values from the probability functions, the hyper-parameters that produced the highest accuracy are stored. After the 10-folds, the optimal hyper-parameters are found based on the median of the values of the hyper-parameters obtained for each fold. Finally, the 10-fold cross-validation is repeated but only with the train and test set in order to guarantee that all test samples are evaluated with the optimal hyper-parameters, which leads to more realistic and stable results.

For the case of the GMM algorithms, the number of Gaussian components is selected from the interval $M = \{2, 4, 8, 16, \cdots, 1024\}$, in the same way as the addressed in [Aria 18a], and the optimal value is found as the one that minimizes the Bayesian information criterion (BIC).

For the methods based on DNNs, the hyper-parameters include the kernel size of the convolutional layers, the number of hidden units in the recurrent cells, the number of neurons in the fully connected layers, the dropout probability, and the initial learning rate. The values for those hyper-parameters are manually optimized, based on prior knowledge about the problem, and aspects about training error, generalization error, and available computational resources.

Chapter 3

Clinical Assessment of Patients and Data Collection

This chapter describes the perceptual scales used to evaluate the disease severity of the patients. The motor state severity is evaluated with the MDS-UPDRS-III scale. In addition, the proposed m-FDA scale is used to evaluate the dysarthria severity of the participants. Both scales are described in Section 3.1. The chapter then describes in Section 3.2 existing databases used in the literature for motor examination of PD patients using information from speech, handwriting and gait. Finally the chapter in Section 3.3 includes a detailed description of the different corpora collected and used for the experiments of this thesis.

3.1 Clinical Assessment of the Participants

3.1.1 Movement Disorder Society - Unified Parkinson's Disease Rating Scale

There is no standard test to evaluate the severity of the symptoms associated to PD. Neurologists rely on clinical history and physical examination to assess the patients. Although there exist several scales to assess the neurological state of PD patients, the most widely used are the Movement Disorder Society - Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [Goet 08] and the Hoehn & Yahr scale. Particularly, the MDS-UPDRS is a perceptual scale used to evaluate motor and non-motor aspects of PD patients. The total MDS-UPDRS is divided into four parts: part I (13 items) concerns non-motor experiences of daily living such as cognitive impairment, depressed mood, sleep disorders, and fatigue. Part II (13 items) considers motor experiences of daily living such as eating, handwriting, and tremor. Part III (33 items) includes the motor examination in lower limbs, upper limbs, and speech production. Part IV (6 items) concerns motor complications such as time spend without medication. The ratings of each item range from 0 (normal) to 4 (severe) and the total score for each part is obtained from the sum of the corresponding items.

In this thesis only the third section (MDS-UPDRS-III) is considered because it evaluates the motor capabilities of the patients. The section has a total of 33 items to evaluate different motor capabilities. Thus, the ground truth to label the neurological

state of the patients is a score ranging between 0 and 132 (33 items×4=132). The MDS-UPDRS-III has only one out of 33 items to evaluate the speech of the patients. However, the speech production process involves different muscles and it is one of the most impaired symptoms of the patients [Horn 98]. Thus, it makes sense to consider a specific scale, in addition to the MDS-UPDRS-III to model only the speech impairments developed by PD patients.

3.1.2 Modified Frenchay Dysarthria Assessment Scale

Speech impairments developed by PD patients are described as *hypokinetic dysarthria*. Therefore, a scale to assess dysarthria is appropriate to assess speech of PD patients. Different scales to evaluate dysarthria include, for instance, the dysarthria profile [Robe 82], the dysarthria examination battery [Drum 93], and the Frenchay Dysarthria Assessment (FDA) [Ende 08]. Particularly, FDA was introduced in 1983 and later revised in 2008 [Ende 08]. The scale covers a wide range of aspects including reflexes, breathing, lips movement, palate movement, laryngeal capacity, tongue posture/movement, intelligibility, and swallowing. To evaluate swallowing the examiner requests the patient to drink different beverages like water and yogurt before speaking, therefore the evaluation requires the patient to be present during the assessment. In many cases the transportation to the clinic is not possible, especially for PD patients in intermediate or severe stages because their reduced mobility. Additionally, the scale is also not suitable for patients who live in remote rural areas where there is almost no clinical expert.

To overcome these issues we recently proposed a modified version of the FDA scale, namely m-FDA [Vasq 18b], which can be administered considering only speech recordings of the patients. Of course swallowing aspects are not covered in this version of the scale, however most of the speech aspects included in the original FDA scale are included in the modified version. The introduced m-FDA scale consists of 13 items and evaluates seven aspects of the speech including breathing, lips movement, palate/velum movement, laryngeal movement, intelligibility, and monotonicity. Each item ranges from 0 to 4 (integer values), thus the total score ranges from 0 (healthy speech) to 52 (completely dysarthric) [Vasq 18b]. Table 3.1 summarizes the speech aspects and items included in the scale. Different speech tasks are considered to evaluate each item of the m-FDA. Respiratory capability (Aspect: Breathing) is evaluated with sustained phonations of vowel /ah/ and diadochokinetic (DDK) tasks. Strength and control of lips closing (Aspect: Lips) are evaluated with DDK tasks and a read text, respectively. Nasal escape and velar movement (Aspect: Palate/Velum) are evaluated with the read text and a DDK task, respectively. Phonatory capability and effort to produce speech (Aspect: Laryngeal) are evaluated with the sustained vowel /a/ and the read text. Correctness and velocity in the tongue movement (Aspect: Tongue) are evaluated with DDK tasks. Finally, intelligibility and monotonicity are evaluated with the read text.

The labeling process for the m-FDA was performed by three phoniatricians who first agreed on the evaluations of ten speakers (five PD patients and five HC subjects, randomly chosen). These initial evaluations allowed the experts to standardize the evaluation criteria. Afterwards, the experts evaluated the speech of the additional

Table 3.1: Aspects and items included in the m-FDA scale

Aspect	m-FDA items	Speech task
Breathing	1) Duration of respiration	Vowel /ah/
	2) Respiratory capability	Vowel /ah/ and /pa-ta-ka/
Lips	3) Strength of closing the lips	/pa-ta-ka/
	4) General control the lips	Read text
Palate/Velum	5) Nasal escape	Read text
	6) Velar movement	/pa-ta-ka/
Laryngeal	7) Phonatory capacity in vowels	Vowel /ah/
	8) Phonatory capacity in continuous speech	Read text
	9) Effort to produce speech	Read text
Tongue	10) Velocity to move the tongue	/pa-ta-ka/
	11) Velocity to move the tongue	/ta/
Intelligibility	12) General intelligibility	Read text
Monotonicity	13) Monotonicity and intonation	Read text

speakers independently. The speakers from the multimodal, longitudinal, and At-home data, considered in this thesis were labeled according to the m-FDA scale. The inter-rater reliability among the phoniatricians was 0.75. It was computed by calculating the average Spearman's correlation between all possible pairs of raters [Vasq 18b]. The m-FDA scale has not been clinically validated yet, however, it can be used to get information about the progression of symptoms related with speech. Additionally, it can be administered based on speech recordings, i.e., patients can stay at home to do the exercises on their own or following the instructions given by the doctor who is in the clinic, or even given by a virtual agent. This scale represents a step towards the automatic administration of speech and language therapy for PD patients. The administration of this scale can be included in following releases of the Apkinson software (see Section 8.2)

3.2 Existing Data

The main aim of this section is to provide information about existing databases (some of them public) that could be used to start or deepen the study of motor impairments in PD patients using different bio-signals. Speech, gait, handwriting, and multimodal studies are reviewed.

3.2.1 Speech

One of the first studies for quantitative analysis of PD speech was the Parkinson's voice initiative[1]. Data collected during that project included utterances of sustained phonations of the vowel /ah/ pronounced by about 50 PD patients. Although the recordings are not publicly available, the main contribution of this initiative was to

[1]Parkinson's voice initiative, http://www.parkinsonsvoice.org/

capture the attention of the research community to address this problem. Few years later, Prof. Sabine Skodda in [Skod 11a] presented a study with data collected from 73 PD patients and 43 HC subjects, German native speakers. The participants were asked to do several speech tasks including the sustained phonation of the vowel /ah/, DDK exercises such as the rapid repetition of the syllables /pa-ta-ka/, the reading of a text with 81 words, and a monologue. This corpus was extended [Skod 11b, Oroz 16b] and its current version includes data from 88 PD patients and 88 HC subjects. In the same year, Prof. Jan Rusz introduced a corpus with 20 newly diagnosed PD patients and 16 HC subjects, Czech native speakers [Rusz 11]. The corpus included recordings of sustained vowels, DDK exercises, 12 isolated words, three sentences, a read text with 80 words and a monologue. The database has been updated since the first release, and now includes a total of 40 PD patients and the same number of HC speakers [Rusz 18b]. In 2013, the authores from [Saka 13] released a database with utterances of 20 PD and 20 HC subjects, Turkish native speakers. The speech tasks included sustained vowels, isolated words, digits, and sentences. This corpus is freely available to be used by the research community interested in the topic[2]. One year later, the PC-GITA corpus was released [Oroz 14]. This database contains utterances of 50 PD patients and 50 age and gender balanced HC subjects. All of the participants are Colombian Spanish native speakers. The participants were requested to perform several exercises, including: sustained phonation of the five Spanish vowels, six different DDK exercises, a set with 45 isolated words, 10 sentences, a read text with 36 words, and a monologue. This corpus is available upon request by contacting the first author of the paper where the data was released[3]. In 2015 another corpus was presented in [Baye 13]. The data include recordings of 168 PD patients, all of them English native speakers. No HC people participated in the study. The speech tasks included the sustained phonation of the vowel /ah/, one DDK task, and a reading passage. In 2016, the authors from [Nara 16] presented a corpus with data from 40 PD patients and 40 HC subjects, Spanish speakers from Extremadura (Spain). The subjects pronounced three repetitions of the vowel /ah/ (for at least five seconds, and on a single breath). Although the speech recordings are not available, a set of extracted features can be downloaded[4]. There is also a recent initiative that pushed the study of Parkinson's speech. It is lead by the mPower consortium and contains recordings of more than 2000 speakers, including PD patients and HC subjects [Bot 16b]. The corpus includes recordings of the sustained vowel /ah/ collected using smartphones. The data can be downloaded after registration in the Synapse portal[5]. The authors from [Beri 17] compiled and released a database with interviews with Muhammad Ali available on YouTube to track longitudinal changes in speech production due to the disease. The speech data contain utterances from 23 interviews performed between 1968 and 1981. These data are also available for the research community[6]. In 2018 the *in the wild speech medical corpus (WSM)* [Corr 18, Corr 19] was released. The

[2] https://archive.ics.uci.edu/ml/datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings

[3] email to Prof. Juan Rafael Orozco-Arroyave: rafael.orozco@udea.edu.co

[4] https://archive.ics.uci.edu/ml/datasets/Parkinson+Dataset+with+replicated+acoustic+features+

[5] https://www.synapse.org/mPower

[6] http://www.public.asu.edu/%7Evisar/software/AliSpeechData.zip

corpus contain data from PD, depression, and cold patients. The data was collected from video blogs (vlogs) from Youtube, when patients talk about different topics, including their disease, present experiences, or personal opinions. The PD data contain videos from 34 PD patients and 19 HC subjects. Recently, in [Moro 19a] the authors introduced the Neurovoz corpus, which contains speech utterances from 47 PD patients and 32 HC subjects, Castilian Spanish speakers. The speech tasks included DDK exercises, six sentences, and a description of a picture. Table 3.2 summarizes the main existing corpora for PD assessment from speech.

Table 3.2: Summary of existing data for speech assessment of PD patients

Database	Source	Participants	Speech tasks	Available [Y/N]
German PD	[Skod 11b]	88 PD, 88 HC	sustained vowels, DDKs, read text, monologue	N
Czech PD	[Rusz 11]	20 PD, 15 HC	sustained vowels, DDKs, read text, monologue	N
de-novo PD Czech	[Rusz 18b]	40 PD, 40 HC	sustained vowels, DDKs, read text, monologue	N
Turkish PD	[Saka 13]	20 PD, 20 HC	sustained vowels, words, read sentences	Y[2]
PC-GITA	[Oroz 14]	50 PD, 50 HC	sustained vowels, DDKs, read sentences, read text, monologue	Y[3]
English PD	[Baye 13]	168 PD	sustained vowels, DDK, read passage	N
Extremadura PD	[Nara 16]	40 PD, 40 HC	sustained vowels	N
mPower	[Bot 16b]	2000 speakers	sustained vowels	Y[5]
Ali speech data	[Beri 17]	23 interviews	from Muhammad Ali performed between 1968 and 1981	Y[6]
WSM	[Corr 18]	34 PD, 19 HC	monologues obtained from YouTube videos	N
Neurovoz	[Moro 19a]	47 PD, 32 HC	DDKs, read sentences, picture description	N

3.2.2 Handwriting

The automatic handwriting assessment of PD patients has increased in the recent years. One important aspect considered in the existing data is the handwriting exercises performed by the participants. Handwriting tasks can be divided into simple drawing exercises, writing tasks, and complex exercises, where the participants have to perform additional activities to the writing process. Handwriting data can be collected offline i.e., on paper and using a normal pen, or online i.e., collected with specialized tablets or smart-pens. For the first case, the obtained static images are analyzed with different computer vision methods. For the second case, it is possible to analyze the dynamics of the handwriting process, including information such as the pressure of the pen and kinematic aspects of the strokes.

There are several databases available to perform research on handwriting assessment of PD patients. For instance, the publicly available data from [Isen 14], which was released in 2014[7] and that contains drawings of Archimedean spirals performed by 62 PD patients and 15 HC subjects. These data contain three types of handwriting exercises: (1) the static spiral test, where three Archimedean spirals appeared on a tablet, and the patients have to retrace them. (2) The dynamic spiral test, where the spirals appear and disappear at certain time stamps, by forcing the patients to keep the pattern in mind while drawing. (3) The stability test, which consists of a

[7] https://www.kaggle.com/team-ai/parkinson-disease-spiral-drawings

red point in the screen where the participants were asked to hold the pen without touching the tablet's surface. In 2016, researchers from the Brno University of Technology released the PahaW database [Drot 16] to the public[8]. The corpus contains data from 37 PD patients and 38 HC subjects, who were requested to perform different handwriting tasks including drawings of Archimedean spirals, the repetition the graph '*l*', the bi-graph '*le*', a set of words, and the sentence in Czech language: *Tramvaj dnes uz nepojede* (the tram wont't go today). Several signals were collected including the on-surface movement, in-air movement, pressure, and position.

In addition to online handwriting data, in [Pere 16b] the authors presented the HandPD dataset, which is formed with Archimedean spirals drawn by 18 HC subjects and 74 PD patients, using normal pen and paper. The corpus was collected at Botucatu Medical School, São Paulo State University, in Brazil, and it is publicly available[9]. The HandPD dataset was updated and called the NewHandPD dataset [Pere 18a]. The new version contain information from 35 HC and 31 additional PD patients. Each subject was asked to perform 12 exercises, 4 of them related to spirals, 4 related to meanders, 2 circle movements (one in the air and another on the paper), and left and right-handed wrist movements. During the exercises, the handwriting dynamics was also captured by means of a smart-pen with several sensors: microphone, finger-grip, axial pressure, and tri-axial accelerometers. This database is also publicly available. In 2019, the authors from [Zham 19] introduced a database with handwriting data from 31 PD patients and 31 HC subjects, who were instructed to write a sentence in English, to repeat the scripts *b*, *d*, and *bd*, and to write and Archimedean spiral. Table 3.3 summarizes the main existing corpora for PD assessment from handwriting.

Table 3.3: Summary of existing data for handwriting assessment of PD patients

Database	**Source**	**Participants**	**Handwriting tasks**	**Available [Y/N]**
PD Spirals	[Isen 14]	62 PD, 15 HC	Spirals, hold the pen in a fixed point	Y[7]
PaHaW	[Drot 16]	37 PD, 38 HC	Spirals, graph repetition, sentence	Y[8]
Hand PD	[Pere 16b]	74 PD, 18 HC	Spirals, meanders	Y[9]
new Hand PD	[Pere 18a]	35 PD, 31 HC	Spirals, meanders, wrist movements	Y[9]
	[Zham 19]	47 PD, 32 HC	Spirals, script repetition, sentence, fluency test	N

Common drawing tasks in the existing data include Archimedan spirals, meanders, and circles. Drawing these type of figures is easy to perform and well tolerated by the patients; however, complexity may increase with the inclusion of more complex figures like houses, cubes, among others, when the patients usually apply different drawing strategies, i.e., the strokes of the figures can be drawn in different orders, which is then reflected in the extracted dynamic features [Impe 19c]. On the other hand, writing tasks included in the existing data include the repetition of scripts, which are easy to write and contain suitable information about the stability of the writing process. In addition, writing words or sentences is suitable to assess agraphia. Writing a sentence requires a higher neuromotor programming load than repeating scripts

[8] https://bdalab.utko.feec.vutbr.cz/
[9] http://wwwp.fc.unesp.br/~papa/pub/datasets/Handpd/

because it also involves linguistic skills, attention, and memory. Writing sentences also provides the possibility to evaluate the motor-planning activity within consecutive words. Finally, there are studies that consider more complex handwriting exercises such as the Rey Osterrieth complex figure [Shin 06], fluency tests based on writing a list of animals [Zham 19], or modified versions of the Fitt's task [Fitt 54], which evaluates the *act of pointing*, i.e., how to move a cursor to a target point [Smit 17b].

3.2.3 Gait

Several studies have performed gait analysis of PD patients. For such a purpose, some repositories with data available to work on the topic have been created. For instance the PhysioNet repository[10] was released in 2000. This database contains gait data from 93 PD patients and 73 HC subjects, where the participants were asked to walk during 2 minutes on level ground using force-sensitive sensors in the shoes. Later in [Bach 10], the authors introduced the Daphnet Freezing of Gait (FoG) dataset, which aimed to evaluate and detect FoG episodes. The recording system used accelerometers placed at the ankles, thighs, and the waist of the patients. About 8 hours of signals were captured from 10 PD patients and FoG episodes were observed in 8 of them. This corpus is available in the UCI-ML repository[11]. In [Bart 11] the authors presented a database with data from 92 PD patients and 81 HC subjects. Gait signals were collected using the eGaIT system[12], which consists of accelerometers and gyroscopes attached to the lateral heel of the shoes. The study was extended and now the database includes recordings of 190 PD patients and 101 HC subjects [Bart 17]. The tasks recorded include 20-meter and 40-meters walking with a pause every 10-meters, heel-toe tapping, and timed up and go (TUG) tests. Most of the recent studies of gait are based on wearable sensors attached to the body or to the shoes. However, there exists another way of collecting walking signals, by using a walkway. In [Hass 12a] the authors presented a database with signals recorded from 310 PD patients while walking on a walkway. Different gait parameters can be extracted from this recording device, however note that its use is restricted only to clinical environments. Table 3.4 summarizes the main existing corpora for PD assessment from handwriting.

3.3 Data Collected During this Thesis

3.3.1 Multimodal Corpus

The multimodal corpus is an extended version of the PC-GITA database [Oroz 14], described in Section 3.2.1. This extended version contains recordings of speech, hand-

[10] PhysioNet: the research resource for complex physiologic signals. https://physionet.org/content/gaitpdb/1.0.0/

[11] Daphnet Freezing of Gait dataset. https://archive.ics.uci.edu/ml/datasets/Daphnet+Freezing+of+Gait

[12] https://www.astrum-it.de/healthcare-medizintechnik/forschungsprojekte/sensorbasierte-bewegungsanalyse.html

Table 3.4: Summary of existing data for gait assessment of PD patients

Database	Source	Participants	Gait tasks	Available [Y/N]
Physionet	[Gold 00]	93 PD, 73 HC	Two minutes walking	Y[10]
Daphnet FoG	[Bach 10]	10 PD	Induced walking tests for FoG	Y[11]
	[Mazi 12]	10 PD	Free walking	N
	[Trip 13]	11 PD, 5 HC	Daily living activities	N
eGaIT	[Bart 17]	190 PD, 101 HC	Scripted walking, heel-toe tapping, TUG	N
	[Orne 17]	46 PD	Scripted walking	N
	[Cuzz 17]	156 PD, 24 HC	10 meters walking	N
	[Kuhn 17]	14 PD, 26 HC	10 meters walking, TUG	N
	[Cara 18]	25 PD, 25 HC	15 meters walking	N
	[Rehm 20]	81 PD, 61 HC	Two minutes walking	N
	[Agha 20]	19 PD	Heel tapping	N
	[Pfis 20]	30 PD	Daily living activities	N

writing, and gait collected from PD patients and HC subjects. Additional details are explained in the following subsections.

Recruitment Process

Data from 106 PD patients and 105 HC subjects were collected and included in the multimodal corpus. All of the subjects are Colombian Spanish native speakers. The inclusion criteria for the HC group guarantee that none of the participants has history of symptoms related to PD or any other kind of movement or speech disorder. In addition, we look for the age and gender distribution for the HC subjects to be non-significantly different than for the PD patients. Most of the recorded patients participate in the weekly meetings of *Fundalianza Parkinson Colombia*[13].

Speech, handwriting, and gait were captured in the same session during 1 hour, distributed as follows: 15 minutes for speech, 30 minutes for gait, and 15 minutes for handwriting. Unfortunately, not all bio-signals are available for all recorded sessions because the equipment to collect handwriting and gait signals was not available at the beginning, and we collect only speech signals. Information of the three bio-signals is available for 39 of the HC subjects and for 70 PD patients, which makes multimodal data available for 109 of the 211 (106 PD + 105 HC) recording samples.

Demographic and Clinical Information of the Participants

The database includes speech, handwriting, and gait signals collected from 106 PD patients and 105 HC subjects. 94 of the PD patients were evaluated by a neurologist expert and labeled according to the MDS-UPDRS-III scale. Additionally, the speech recordings from 93 of the PD patients and from 48 of the HC subjects were labeled by expert phoniatricians according to the m-FDA scale described in Section 3.1.2. The data were collected with the patients in ON state, i.e., under the influence of medication. The recording procedure is in compliance with the Helsinki Declaration

[13]Fundalianza Parkinson Colombia: https://sites.google.com/view/fundalianzaparkinsoncolombia

and it was approved by the Ethics Committee of the medical faculty of the University of Antioquia, and a written informed consent was signed by each participant. Table 3.5 summarizes clinical and demographic aspects of the participants included in the corpus. The neurological state of the patients was evaluated according to the MDS-UPDRS-III scale, which was administered by a neurologist in the *Pablo Tobón Uribe* Hospital in Medellín, Colombia. The average MDS-UPDRS-III of patients is 36.2, which indicates that the patients are in an intermediate state of the disease (maximum value of the scale is 132). The average m-FDA score for PD patients is 23.9, which indicates that their dysarthria severity is moderate to severe in some cases. For the HC subjects, the average m-FDA score is 7.8, which reflects a mild dysarthria severity, characteristic of the normal aging process. Statistical tests are included in the caption of Table 3.5 to validate the balance in gender and age, and the significant difference that exists between the m-FDA scores assigned to PD patients and HC subjects. The complete metadata of the participants from this corpus can be accessed online[14]. The distributions of age, MDS-UPDRS-III, m-FDA, and the time post diagnosis are shown in Figure 3.1. An interactive dashboard with information of the subjects from the corpus is also available online[15].

Table 3.5: Clinical and demographic information of the subjects from the multimodal corpus.

	PD patients	**HC subjects**	**Patients vs. controls**
Gender [F/M]	49/57	53/52	$^{*}p = 0.99$
Age [F/M]	60.9(11.2)/64.7(9.4)	59.9(8.7)/63.5(10.4)	$^{**}p = 0.08$
Time since diagnosis [F/M]	15.5(14.5)/8.1(5.9)	–	
MDS–UPDRS–III [F/M]	36.2(18.1)/36.3(18.9)	–	
m–FDA total [F/M]	23.7(8.9)/24.1(7.6)	6.6(7.0)/9.0(8.2)	$^{**}p \ll 0.005$

Time since diagnosis and age are given in years. [F/M]: Female/Male. Average(Standard deviation).
$^{*}p$–value calculated through Chi–square test. $^{**}p$–value calculated through Mann-Whitney U-test.

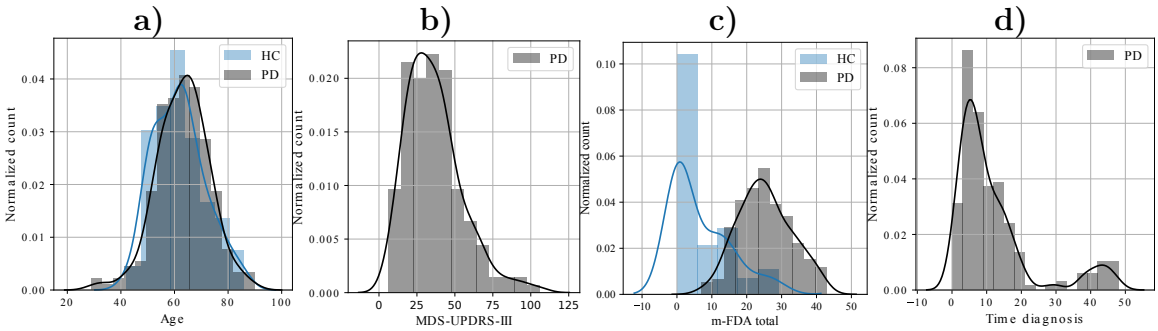

Figure 3.1: Distribution of: **a)** age, **b)** MDS-UPDRS-III, **c)** total score of the m-FDA scale, and **d)** time post PD diagnosis.

Speech Data

Speech data is available for the 106 PD patients and 87 of the HC subjects. The data include 48 PD patients and 48 HC subjects from the original PC-GITA corpus.

[14]Metadata Multimodal corpus: `https://bit.ly/2UwNUlG`
[15]Dashboard Multimodal corpus: `https://bit.ly/3cWUyYn`

Two of the original 50 HC subjects and 50 PD patients from PC-GITA were excluded to guarantee the age-balance among participants. The speech of the resulting 48 PD patients and 48 HC subjects from the PC-GITA corpus was recorded with a sampling frequency of 44.1 kHz and 16-bit resolution, in a soundproof booth from the Noel Clinic[16] at Medellin, Colombia. These recordings were re-sampled to 16 kHz. The remaining recordings were recorded with a noise-cancellation headset with a sampling frequency of 16 kHz, using the Neurospeech software [Oroz 18].

The speech protocol considers the same speech tasks recorded in the PC-GITA corpus [Oroz 14], except for the isolated words. The speech tasks include the sustained phonation of the vowel /ah/, six different DDK exercises (/pa-ta-ka/, /pa-ka-ta/, /pe-ta-ka/, /pa/, /ta/, /ka/), the reading of 10 different complex and simple sentences (from the syntactic point of view), a read text with 36 words phonetically balanced that contains all the Spanish phonemes (spoken in Colombia), and a spontaneous speech tasks where the participants were asked to speak about their daily routine. Detailed information about duration, number of words, number of phonemes, and phonetic distribution of the 10 sentences is included in Table 3.6 and Figure 3.2.

Table 3.6: Details of the sentences included in the corpus

	Sentence		**Duration**	# **words**	# **phonemes**	# **unique phonemes**
1	Mi casa tiene tres cuartos	Simple	1.9(0.4)	5	20	12
2	Omar, que vive cerca, trajo miel	Complex	2.6(0.7)	6	22	14
3	Laura sube al tren que pasa	Complex	2.2(0.5)	6	19	12
4	Los libros nuevos no caben en la mesa de la oficina	Simple	3.4(1.0)	11	39	15
5	Rosita Niño, que pinta bien, donó sus cuadros ayer	Complex	4.3(1.2)	9	37	17
6	Luisa Rey compra el colchón duro que tanto le gusta	Complex	4.0(1.2)	10	38	19
7	Viste las noticias? Yo vi ganar la medalla de plata en pesas, Ese muchacho tiene mucha fuerza!	Complex	7.9(2.2)	17	67	24
8	Juan se rompió una pierna cuando iba en la moto	Simple	3.2(0.9)	10	34	19
9	Estoy muy triste, ayer vi morir a un amigo	Simple	3.3(0.8)	9	32	16
10	Estoy muy preocupado, cada vez me es más difícil hablar!	Complex	4.3(1.1)	10	41	19

Duration is given in seconds. Average (Standard deviation)

Handwriting Data

Handwriting data is available for 76 PD patients and 57 HC subjects. The data consist of online drawings captured with a Wacom cintiq 13-HD[17] tablet with a sampling frequency of 180 Hz. The system to capture handwriting data is shown in Figure 3.3. The tablet captures six different signals: x-position, y-position, in-air movement, azimuth, altitude, and pressure of the pen. Figure 3.4 illustrate the difference in the azimuth and altitude angles.

The handwriting protocol includes a total of 17 exercises divided into drawing and writing tasks. On the one hand, drawing tasks consist of geometrical shapes like Archimedean spirals, circles, a house, two concentric rectangles, a rhombus, and a cube. Particularly, spirals and circles have been frequently used to evaluate motor impairments in patients with different neurodegenerative diseases [Vess 19], and they

[16] Clinica Noel: http://www.clinicanoel.org.co/en/home/

[17] Cintiq 13HD Graphic pen tablet for drawing https://www.wacom.com/es-ar/products/pen-displays/cintiq-13-hd

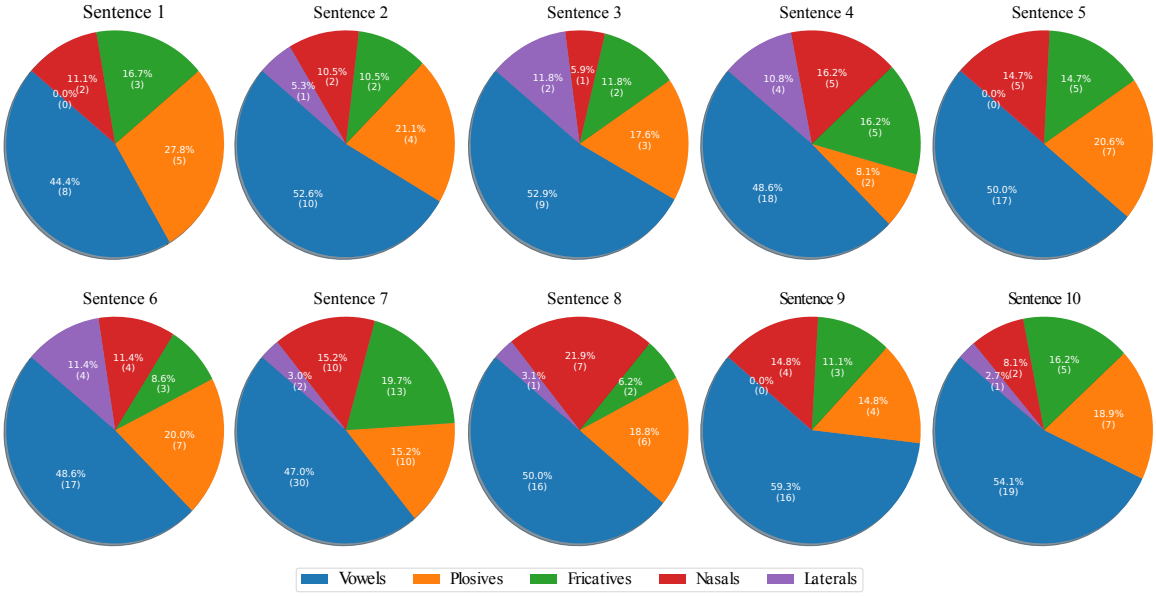

Figure 3.2: Phonetic details of the read sentences included in the corpus

Figure 3.3: System to capture handwriting data

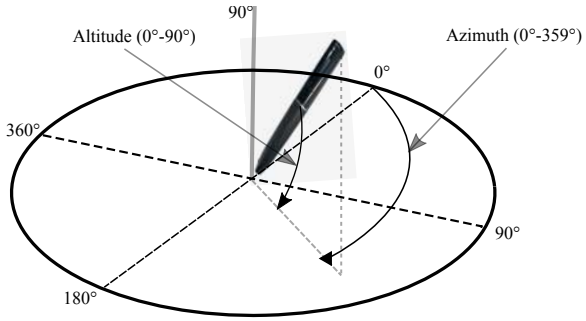

Figure 3.4: Difference between azimuth and altitude angles

are usually easy to perform by most patients. We additionally include the drawing of the Rey-Osterrieth complex figure [Shin 06] (see Figure 3.5), which is a neuropsychological task typically used to evaluate the spatial constructional ability and visual memory, and has been used to measure executive functions mediated by the pre-

frontal lobe [Shin 06]. Besides, writing tasks include simple writing exercises such as the writing the scripts l and m in a continuous and long trace. For these simple writing exercises, PD patients may produce slower and more irregular movements than the HC subjects. In addition, PD patients may exhibit micrographia over time when performing these exercises. The additional writing tasks include more complex exercises like writing the digits (0 to 9), the identification number of the patients, the name and signature of the participant, a free sentence, and the alphabet. These more complex exercises require a higher degree of simultaneous processing and cognitive load than the script repetition, since they also involve linguistic skills, attention, and memory [Vess 19]. Particularly, the sentences allow to capture a large number of in-air movements between the words. Table 3.7 summarizes the handwriting tasks included in the database. Particularly, the template used to draw Archimedean spiral is shown in Figure 3.6. The diameter of the spiral and the distance between the loops were set to 15.2 cm and 1.9 cm, respectively. Figure 3.7 includes some examples of the drawing tasks included in the corpus. The color of the drawings indicates the pressure of the pen.

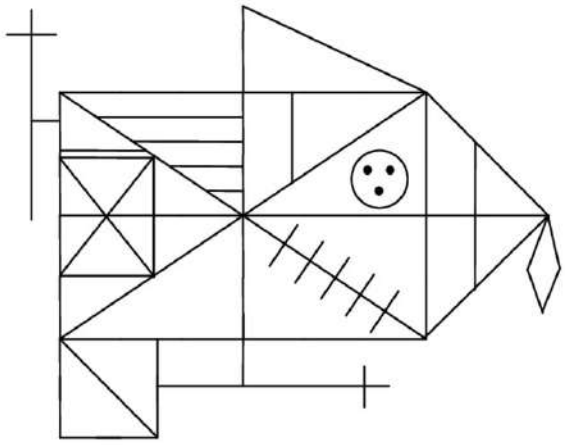

Figure 3.5: Rey-Osterrieth complex figure. Source: [Canh 00]

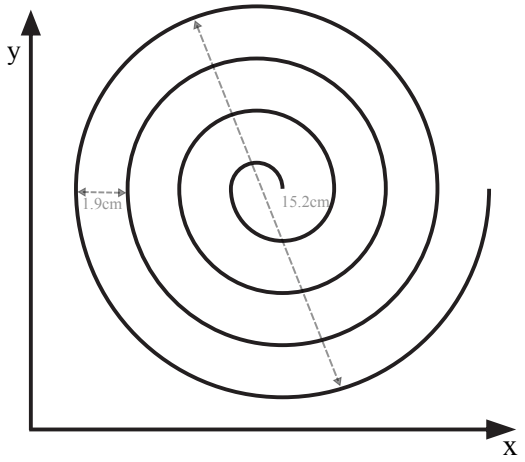

Figure 3.6: Template for the Archimedean spiral

Table 3.7: Handwriting tasks included in the corpus

Writing tasks	**Drawing tasks**
Alphabet	Circle
Free sentence	Guided circle
ID number	Cube
Name	House
Digits	Rectangles
Signature	Rhombus
Repetition of the graph l	The Rey-Osterrieth figure
Repetition of the graph m	Free Archimedean spiral
	Archimedean spiral with template

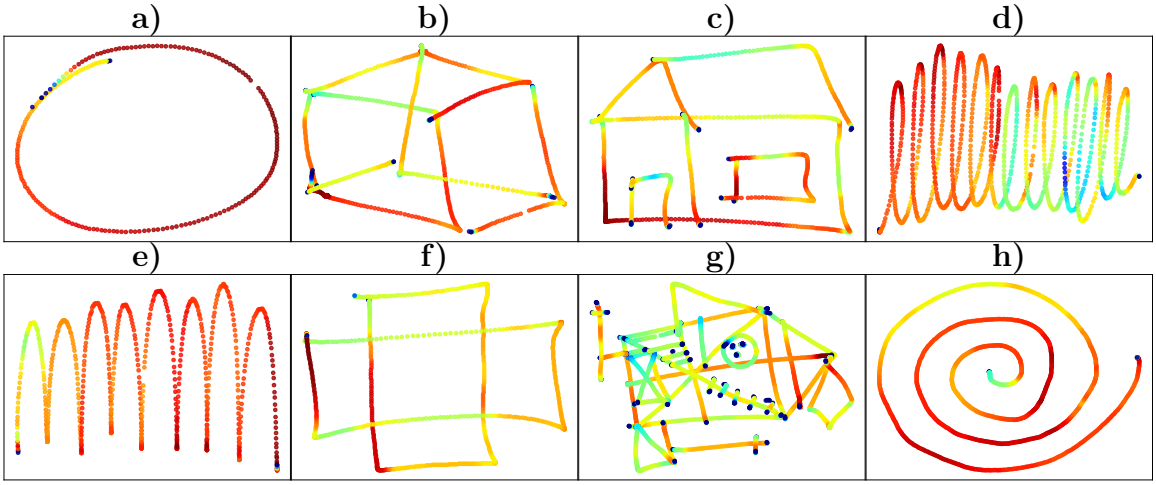

Figure 3.7: Example of drawings from the handwriting data: **a)** circle, **b)** cube, **c)** house **d)** graph *l*, **e)** graph *m*, **f)** rectangles, **g)** the Rey-Osterrieth figure, and **h)** Archimedean spiral.

Gait Data

Gait data is available for 76 PD patients and 57 HC subjects. Gait signals were captured with the eGaIT system[18], which consists of 3D-accelerometers (range ±6g) and 3D gyroscopes (range ±500°/s) attached to the external side (at the ankle level) of the shoes [Bart 17]. Data from both feet were captured at a sampling rate of 100 Hz and 12-bit resolution. Figure 3.8 shows the eGait system and the inertial sensor attached to the lateral heel of the shoe. The signals are transmitted by Bluetooth to a tablet where they are received and stored by an android app (see Figure 3.8a). Seven tasks are included in the protocol to capture gait data.

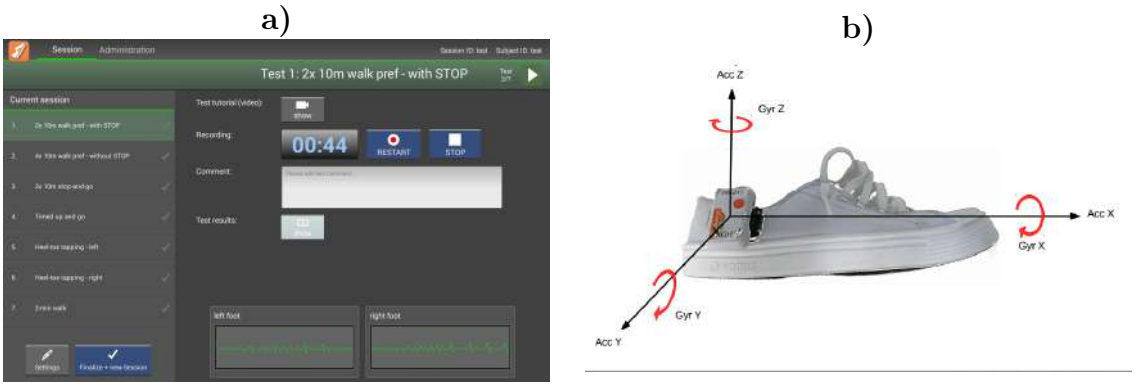

Figure 3.8: eGait system to capture gait data.

2–10m walk with stop (2x10): The test begins with the patient standing. The subject walks in a straight line for 10 meters at a comfortable pace, then stops for 2–3 seconds, then turns clockwise and walks back to the initial point.

[18] https://www.astrum-it.de/healthcare-medizintechnik/forschungsprojekte/sensorbasierte-bewegungsanalyse.html

4–10m walk without stop (4x10): This task begins also with the subject standing. The participant walks in a straight line for 10 meters at a comfortable pace. Then turns clockwise and walks back without pausing. In the starting point the subject turns counterclockwise and repeat the first 10 meters, then turns clockwise and walks back again.

2–10m stop-and-go: The participant starts walking a distance of 10 meters. For three times during the walk (every three meters), the subject is asked to stop and then to resume walking. The same instructions are executed again when the patient returns to the starting point.

TUG: The subject is sitting on a chair with his/her back leaned against the backrest. Then the participant stands up, walks a 3 meters distance, turns clockwise, returns, and sits down again.

Heel-toe tapping-left foot: The patient is on a chair, where (s)he alternate taps the ground with their left heel and their tiptoes for 20 seconds.

Heel-toe tapping-right foot: This tasks is the same as the previous one, but with the right foot instead of the left one.

2-min walk: The subject walks at his/her own pace for 2 minutes.

3.3.2 Longitudinal Corpus

The longitudinal corpus was built with a subset of 9 PD patients from the multimodal corpus, who were recorded in up to 7 different sessions. The aim of this corpus is to evaluate the impact of the motor deficits of the patients in long-term. The group of patients was recorded in seven sessions from 2012 to 2019. The data for the seven sessions were collected in 2012 (June), 2014 (June), 2015 (February), 2015 (August), 2016 (February), 2017 (December), and 2018 (December). Table 3.8 indicates the MDS-UPDRS-III and the m-FDA labels assigned to the patients of this corpus. Age and gender are also provided. Unfortunately, the MDS-UPDRS-III labels of the third recording session (S3), and the m-FDA labels for session S7 are not available. Patients PD08 and PD09 enrolled later in the study, thus data in the first session is not available for them.

Regarding the collected signals in the corpus, speech data is available in all sessions; however, handwriting and gait data is only available for sessions S2, S6, and S7. The protocols for collecting speech, handwriting, and gait data are the same as the ones described for the multimodal corpus. A professional audio setting was used for the first two sessions and the patients were asked to come to the clinic to perform the speech exercises; however, this represented a limitation for some of the patients due to their motor complications. The remaining five sessions were recorded with a conventional headset when the patients attend the weekly meetings at *Fundalianza Parkinson Colombia*.

Table 3.8: General information of patients included in the longitudinal corpus. $\mathbf{S_i}, i \in \{1, 2, \ldots, 7\}$:$i$th longitudinal session

Patient ID	Gender	Age	MDS-UPDRS-III							m-FDA						
			S1	S2	S3	S4	S5	S6	S7	S1	S2	S3	S4	S5	S6	S7
PD01	M	64	28	19	-	13	-	31	27	15	17	16	17	20	17	-
PD02	F	55	41	35	-	35	33	38	31	22	31	24	21	37	22	-
PD03	F	51	38	49	-	44	45	40	33	13	14	16	10	18	25	-
PD04	F	55	43	10	-	19	-	-	19	7	9	20	21	23	-	-
PD05	M	59	6	8	-	24	21	20	23	25	35	22	25	27	27	-
PD06	M	68	14	25	-	7	-	17	33	23	18	19	25	23	24	-
PD07	F	55	29	26	-	26	31	41	32	24	24	17	24	25	32	-
PD08	M	67	-	58	-	65	49	41	59	-	29	33	39	31	31	-
PD09	M	70	-	64	-	37	26	57	39	-	13	24	18	24	21	-

3.3.3 At-Home Corpus

The At-Home corpus is considered to monitor the progress of the speech deficits of PD patients in short-term periods of time and the impact of the medication. The data were recorded in 2016, within the scope of our participation in the *2016 Frederick Jelinek Memorial Summer Workshop (JSALT)*[19], and comprise a group of seven PD patients recorded four times per day (every two hours), once per month during four months. Thus, there is a total of 16 recording sessions per patient. Speech of the patients were recorded at their homes with a conventional headset. The neurological state of the patients was evaluated according to the MDS-UPDRS-III scale at the beginning of the first recording session. The speech recordings of the 16 sessions were evaluated following the m-FDA scale.

The speech data collected in the At-Home corpus include three DDK exercises (/pa-ta-ka/, /pa-ka-ta/, /pe-ta-ka/), the read text with 36 words, phonetically balanced from the PC-GITA corpus [Oroz 14], and continuous speech utterances from conversations between the patients and the interviewers. Table 3.9 shows demographic information about the patients from the At-Home corpus, and their m-FDA scores within the 16 sessions.

Table 3.9: General information of patients included in the At-Home corpus. $\mathbf{S_i}, i \in \{1, 2, \ldots, 16\}$:$i$th at-home sessions

Patient ID	Gender	Age	m-FDA															
			S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16
PD01	M	64	20	23	21	12	21	18	17	16	17	20	25	20	27	23	22	22
PD02	F	55	35	35	35	35	33	35	34	34	36	37	39	39	42	42	42	42
PD03	F	51	19	15	20	19	14	19	18	16	20	20	17	17	24	23	23	23
PD04	M	59	20	25	24	23	28	26	25	25	29	29	29	27	28	28	28	28
PD05	M	68	25	25	26	25	24	30	27	28	28	25	28	28	25	24	24	25
PD06	F	55	26	32	31	30	32	30	31	31	33	33	30	34	37	38	37	34
PD07	M	67	40	35	38	36	37	34	35	34	33	28	37	37	36	36	37	38

[19]Remote Monitoring of Neurodegeneration through Speech https://www.clsp.jhu.edu/workshops/16-workshop/remote-monitoring-of-neurodegeneration-through-speech/

3.3.4 Apkinson Corpus

This corpus corresponds to the data collected using the Apkinson android application [Vasq 19a, Oroz 20a]. Apkinson was designed to record several signals using sensors embedded on the smartphone (microphone, accelerometer, and gyroscope) and performs different analyses to model the neurological progression of PD patients. A detailed description of Apkinson is presented in Section 8.2.

General Description of the Participants

At the moment, data from up to 38 PD patients and 60 HC subjects were collected using Apkinson. The two groups are matched by age, gender, and scholarity. A total of 17 of the 38 patients were evaluated with the MDS-UPDRS-III. The reason for the other patients not to be evaluated is because the project that finances the cost of the neurological evaluation was not running for the time of those recordings. None of the participants in the HC group presented any neurological or movement disorder. Table 3.10 includes details of clinical and demographic information of the HC and the PD patients. The complete metadata of the participants from this corpus can be also accessed online[20]. The distributions of age, MDS-UPDRS-III, and the time post the diagnosis are shown in Figure 3.9.

Table 3.10: Clinical and demographic information of the subjects from the Apkinson corpus.

	PD patients	HC subjects	Patients vs. controls
Gender [F/M]	17/20	30/30	$^{*}p = 0.697$
Age [F/M]	66.5(12.5)/69.3(9.0)	63.6(6.3)/60.9(12.9)	$^{**}p = 0.008$
Sholarity [F/M]	12.1(4.0)/12.8(3.8)	9.6(3.5)/11.2(4.7)	$^{**}p = 0.007$
Time since diagnosis [F/M]	9.8(10.3)/8.3(5.5)	–	
MDS-UPDRS-III [F/M]	14.9(7.5)/19.9(6.9)	–	

Time since diagnosis, age, and scholarity are given in years. [F/M]: Female/Male. Average(Standard deviation). $^{*}p$–value calculated through Chi–square test. $^{**}p$–value calculated through Mann-Whitney U-test.

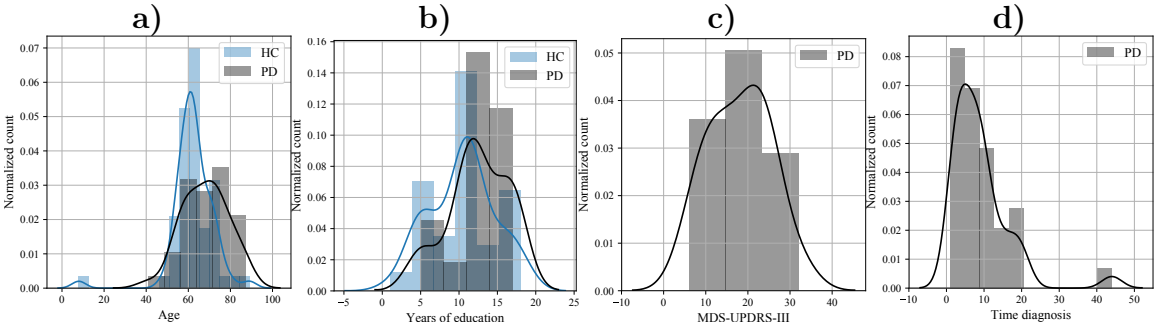

Figure 3.9: Distribution of metadata for the Apkinson corpus: **a)** age, **b)** scholarity, **c)** MDS-UPDRS-III, **d)** time post PD diagnosis.

[20]Metadata Apkinson corpus: `https://bit.ly/3490rPq`

Data collection

The data collected from Apkinson comprise 38 different exercises. The set includes tasks of different nature like speech production, hands movement, gait, and finger tapping. There are three groups of exercises, the first group has a total of 21 speech tasks including the sustained phonation of the vowels /ah/, /ih/, and /uh/, the same six DDK exercises from the multimodal corpus, the reading of the 10 sentences from Table 3.6, and the description of the cookie theft picture from the Boston Diagnostic Aphasia Examination [Boro 80], which is observed in Figure 3.10.

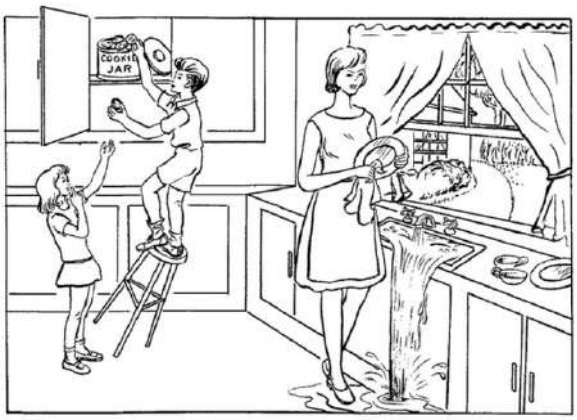

Figure 3.10: Cookie theft Picture from the Boston Diagnostic Aphasia Examination

The second group of exercises includes 11 tasks that are captured with the inertial sensors of the smartphone. The aim is to evaluate different abnormal aspects in movements including postural tremor, kinetic tremor, gait deficits, among others. These 11 tasks includes: (1) *Posture*, where the patient has to stands up straight during 30 seconds, (2-3) *circles*, where the patient has to make circles with the extended left and right arm, (4-5) *pronation/supination*, where the patient stretches out the left/right arm with the downward palm, and then turns the palm up & down, several times, (6-7) *finger to nose*, where the patient extends the left/right arm and then touches his/her nose and extends the arm again, several times, (8-9) *postural tremor*, where the patient extends the left/right arm and holds the smartphone in this position for at least 10 seconds, (10) *gait*, where the patient performs a short path walking four times, and (11) *2 minutes walk*, where the patient performs a normal walk exercises during two minutes.

Finally, the third group of exercises include three tasks to model fine-motor skills of the patients. The first one consists on tapping with the thumb of the dominant hand ladybugs that randomly appear on the screen. For the second task the finger tapping is repeated but now with both thumbs to hit two ladybugs that appear randomly on the screen (each ladybug is located in the right and left half of the screen thus each finger is close to a ladybug and a natural movement is guaranteed). The third task is to slide horizontally a bar until reaching a target point, which moves randomly every time it is reached. This third task is inspired in the Fitt's test to evaluate human computer interaction systems [Fitt 54]. Each fine-motor task requires rapid reaction, concentration, ability to associate, spatial location and repeated movements of extension and contraction of the fingers [Oroz 20a].

Chapter 4

Analysis of Parkinson's Disease from Speech

Speech is well known to be one of the most complex motor skills of humans, requiring precise/accurate control of about 100 different muscles, thus making our vocal system one of the more sensitive to the effects of PD [Chen 11]. Different speech impairments associated to PD are grouped as hypokinetic dysarthria, which appears in about 90% of the patients [Ho 99]. Hypokinetic dysarthria includes symptoms such as rigidity of the vocal folds, bradykinesia, reduced muscular control of the larynx and other organs related to the speech production. The effect of dysarthria in the speech of patients include increased acoustic noise [Horn 98], reduced intensity [Bake 98], harsh and breathy voice quality [Tsan 10], increased voice nasality [Spen 05], monopitch, monoludness, speech rate disturbances [Skod 11a], imprecise articulation of consonants [Tyka 17], and involuntary introduction of pauses [More 03]. In general, speech impairments appear in initial stages of the disease [Rusz 11], producing a negative impact in the communication skills and thus limiting social life of patients [Pell 06]. These symptoms can be highly dependent on the pharmacological therapy of the patients. According to [Rusz 16], speech impairments tend to improve or remain relatively stable after the initiation of dopaminergic treatment, especially for patients in early stages of the disease. Several studies have described the speech impairments developed by PD patients in terms of four different dimensions: phonation, articulation, prosody, and intelligibility [Rusz 11, Bock 13, Oroz 16b].

Phonation symptoms are related to the stability and periodicity of the vocal fold vibration, and with difficulties in the process of producing air in the lungs to make the vocal folds vibrate. For some cases, the speech of patients is not necessarily affected but the respiration does, which in the end affects phonation. Different phonation deficits are associated to PD patients, including differences in glottal noise compared to healthy speakers, incomplete vocal fold closure, and vocal folds bowing, which are typically characterized with measures such as noise to harmonics ratio (NHR), glottal to noise excitation ratio (GNE), harmonics to noise ratio (HNR), and voice turbulent index (VTI), among others [Tana 11]. On the other hand, phonation symptoms have been analyzed in terms of perturbation measures such as jitter, shimmer, amplitude perturbation quotient (APQ), pitch perturbation quotient (PPQ), and nonlinear dynamics (NLD) measures such as the correlation dimension (CD), the largest

Lyapunov exponent (LLE), the Hurst exponent (HE), or the Lempel-Ziv complexity (LZC) [Trav 17].

Articulation symptoms are related to the modification of the position, stress, and shape of several limbs and muscles to produce speech. These symptoms have been modeled by means of features such as vowel space area (VSA), vowel articulation index (VAI), formant centralization ratio (FCR), DDK regularity, and the onset energy. One of the first observed articulation impairments was the imprecise production of stop consonants such as /p/, /t/, /k/, /b/, /d/, and /g/ [Loge 78, Acke 91, Tyka 17]. Particularly, PD patients usually have incomplete vocal closure by maintaining a continuous level of vocal fold activity to avoid the difficulty of initiating the phonation [Blan 09]. This behavior causes that voiceless stops such as /p/, /t/, and /k/ are replaced by /b/, /d/, and /g/. Figure 4.1 shows an example of an HC subject and a PD patients uttering the syllables /pa-ta-ka/. The red line indicates the countour of the fundamental frequency (F_0). In Figure 4.1**b)**, for the case of the PD patient, the periodic signal before the consonant burst reveals an incomplete lip closure or a possible lack of control of the velum. When the velum is not well controlled, air continues to come through the mouth while producing the phoneme /k/. These effects are also visible in the spectrogram, especially for the plosive /k/, where the fundamental frequency from the previous vowel /ah/ is joined with the fundamental frequency from the following vowel /ah/, transforming the plosive /k/ into something similar to a fricative /g/, which in Spanish is the closest voiced phoneme to the /k/. This effect makes observable problems to control the velum and to stop the vibration of vocal folds. These effects are not observed for the case of the HC subject in Figure 4.1**a)**, where the closures of lips, palate, and velum are well defined both in time domain signal, the fundamental frequency, and the spectrogram.

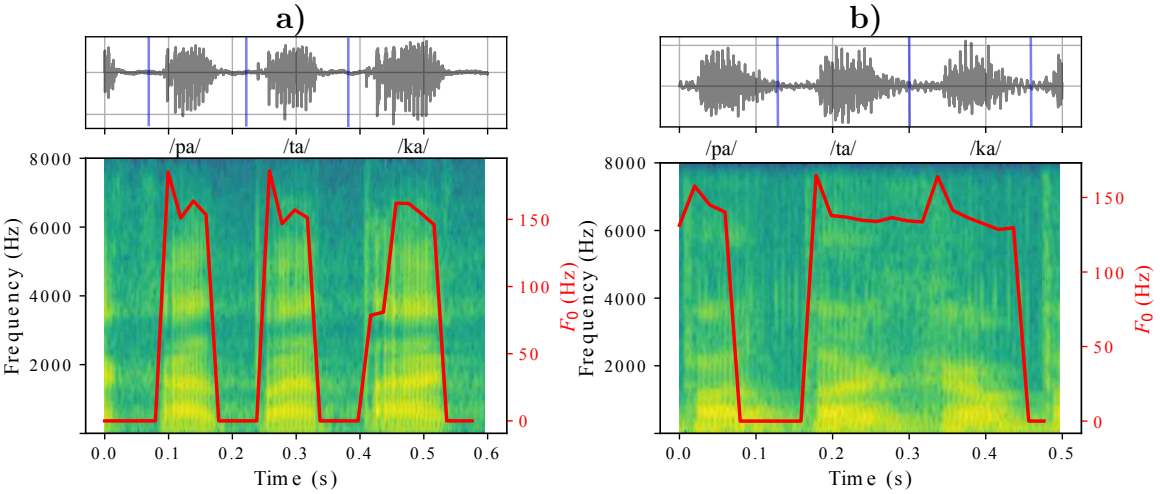

Figure 4.1: Speech signals, fundamental frequency and spectrograms of a PD patient and a HC speaker pronouncing the syllables /pa-ta-ka/. **a)** 49 year old male healthy speaker. **b)** 48 year old male PD patient with MDS-UPDRS-III: 9, and m-FDA:36

The literature also reported that the first and second formant frequencies exhibit shallower slopes for the case of PD patients with respect to HC subjects, which indicates decreased tongue/jaw movements [Kim 09, Wals 12]. Other articulation

symptoms include reduced duration of voiced segments and transitions, difficulties to start/stop vocal fold vibration while speaking [Oroz 16b], and increased voiced onset time (VOT), which increases with the disease severity [Forr 89].

Prosody deficits in PD are manifested as monotonicity, monoloudness, reduced stress, and changes in speech rate and pauses [Skod 11a]. Monrad-Krohn, in [Monr 57] denoted as *aprosody* all these symptoms related to pitch attenuation, rate, and loudness variation, often observed in PD patients. In addition to *aprosody*, the bradykinesia and freezing of movement sometimes cause difficulty in the initiation of voluntary speech and inappropriate long silences. Prosody is commonly evaluated with features related to pitch, intensity and duration. Prosody impairments may be caused also by non-motor symptoms like depression, which also affects PD patients [Star 90].

Finally, intelligibility is a measure of how comprehensible is the speech of a person. In other words, how much of the speech of a person can be understood. Several studies have reported a reduction of the perceived intelligibility in PD patients [Barn 16, Dima 17]. The intelligibility assessment have been performed using speech recognizers, where the word error rate (WER) can be computed for HC and PD patients [Oroz 16a].

The rest of the chapter is divided as follows: Section 4.1 shows a review of the literature about automatic speech evaluation of PD patients from a pattern recognition point of view, both to classify PD patients and HC subjects, to evaluate the neurological state of the patients, and to estimate the dysarthria severity of the patients. Then, Section 4.2 describes the methods classically addressed in the literature to model the speech of patients in terms of phonation, articulation, and prosody dimensions. These methods are used within the scope of this thesis as baselines for the proposed approaches. Then, Section 4.3 describes the proposed approach to model the speech of PD patients using phonological features. Section 4.4 described our proposed approach to model the speech of PD patients based on an unsupervised learning strategy using recurrent autoencoders. The chapter finishes in Section 4.5, where we describe methods to model the speech of PD patients in an end-to-end fashion using different configurations of CNNs.

4.1 A Review on Automatic Assessment of Speech in PD Patients

The clinical observations in the speech of PD patients can be objectively and automatically measured by computer aided methods supported in signal processing and pattern recognition with the aim to address three main aspects: (1) to support the diagnosis by classifying HC subjects and PD patients, (2) to evaluate the neurological state of patients according to a neurological scale such as the MDS-UPDRS, and (3) to evaluate the level of degradation of the speech of the patients according to a specific dysarthria scale such as the m-FDA. The review presented in the following subsections will cover studies focused on these three aspects. In addition, we will cover both classical pattern recognition approaches and novel studies based on deep learning strategies.

4.1.1 Automatic Classification of PD and HC Subjects

The studies addressed here aim to support the diagnosis of PD patients using several strategies based on speech processing and pattern recognition to classify PD patients vs. HC subjects. One of the first studies was proposed by Tsanas et. al. [Tsan 12], who extracted phonation features such as jitter, shimmer, noise measures, and NLD features. The authors considered several feature selection algorithms and classifiers based on SVMs and random forest (RF) to classify 263 utterances from the sustained vowel /ah/ from 43 subjects (10 HC and 33 PD), English speakers. The authors reported accuracies of up to 97.7% depending on the selected features. Although the high reported accuracies, the authors mixed the speakers from the train and test sets, which makes the results highly optimistic and biased. Phonation features were also considered by Sakar et al. [Saka 13], who computed perturbation features such as jitter, shimmer, APQ, and PPQ. The authors classified utterances from 20 PD patients and 20 HC subjects, Turkish speakers, who pronounced sustained vowels, isolated words, and short sentences. Classifiers based on K-nearest neighbors (KNN) and SVM were considered. The authors reported accuracies of up to 75% with utterances of the sustained vowels. The phonation analysis considered in [Oroz 15] included four different approaches to characterize sustained vowels: stability and periodicity, noise measures, spectral wealth, and NLD. The four approaches were used to classify utterances from the PC-GITA database [Oroz 14], using an SVM classifier. The authors reported accuracies of up to 84%, depending on the analyzed vowel and on the feature set. Similar phonation features were considered in [Nara 16] to characterize sustained phonations of vowel /ah/ from 40 PD patients and 40 HC subjects, Spanish speakers. The features included: (1) perturbation measures such as jitter, shimmer, PPQ, and APQ, (2) noise measures such as HNR and the GNE, (3) Mel frequency cesptral coefficients (MFCC), and (4) NLD measures such as the recurrence period density entropy, the detrended fluctuation analysis (DFA), and the pitch period entropy. The authors proposed a Bayesian classification strategy and reported accuracies of up to 75.2%. In [Vill 15] the authors proposed a phonation analysis based on time frequency representations to assess tremor in the speech of PD patients. The extracted features were based on energy and entropy computed from three time frequency representations: modulation spectra, the wavelet packet transform, and the Wigner-Ville distribution. The proposed features were extracted from short sentences from the PC-GITA corpus [Oroz 14], and classified using GMMs and SVMs algorithms. The proposed approach achieved an accuracy of up to 77%. In [Hemm 16] the authors proposed a novel phonation analysis based on the Hilbert-Huang transformation computed upon modulated (varying between low and high pitch) and sustained vowels. The authors analyzed the fundamental frequency (F_0) contour and its range to measure monotonicity in PD speakers. The classification was performed with an SVM, and accuracies of up to 86% were reported in the PC-GITA database [Oroz 14]. In [Meky 16] the authors characterized utterances of sustained vowels using perceptual features to classify 84 PD patients and 49 HC subjects, Czech speakers. The perceptual features were based on MFCC, linear frequency cepstral coefficients, linear prediction coefficients, perceptual linear predictive coefficients (PLP), among others. The authors considered a feature selection method based on sequential forward feature selection and a RF classifier, and reported accuracies

of up to 91.7%. The authors from [Saka 19], proposed a phonation analysis based on the Q-factor wavelet transform computed upon sustained phonations of vowel /ah/, pronounced by 188 PD patients and 64 HC subjects, Turkish speakers. This version of the wavelet transform has higher frequency resolution than the standard discrete wavelet transform (DWT). Wavelet decompositions were computed over the F_0 contour of sustained vowels. The authors extracted features based on the energy-content and entropy of the decompositions. The extracted features were classified with an ensemble of several classifiers. The authors reported an F1-score of up to 0.84 when the proposed features were combined with standard phonation features based on perturbation and noise measures. In [Aror 19], the authors classified utterances from sustained vowels from 14483 PD patients and 15321 HC subjects, obtained via telephone calls. The phonation features included classical perturbation, noise, and NLD features, as in [Tsan 12]. The authors considered also several feature selection strategies, and RF classifiers. The best reported results showed a sensitivity of up to 64.9%, and a specificity of up to 67.9%. However, the HC subjects were younger than the PD patients, which may induce a bias in the results [Aria 17]. NLD features were also considered in [Visw 20], where the authors computed the fractal dimension and the normalized mutual information in sustained phonations of /ah/, /uh/, and /m/ sounds, pronounced by 22 PD patients and 24 HC speakers. A linear SVM classifier achieved an accuracy of 81% with the extracted features. In [Ali 19b] the authors proposed a method based on linear discriminant analysis and neural networks optimized with a genetic algorithm. The proposed model was trained with phonation features extracted from sustained vowels from the Turkish corpus from [Saka 13]. The authors reported an accuracy of up to 95%. The authors from [Wodz 19] proposed a deep learning-based phonation analysis. The method aimed to classify sustained phonations of PD patients and HC subjects using a ResNet architecture trained with spectrograms. The network was pre-trained with utterances from the Saarbruecken voice database [Wold 07] and then fine-tuned to classify utterances from sustained vowels from the PC-GITA corpus [Oroz 14]. The proposed strategy achieved an accuracy of up to 92%.

There are some studies that have considered a reconstruction of the glottal signal to evaluate different phonation aspects of PD patients. For instance, in [Bela 16], the authors performed a reconstruction of the glottal source signal using an iterative adaptive inverse filtering algorithm, and extracted several temporal, spectral, and NLD features from the reconstructed glottal signal. The temporal features included the opening, closing, speed, amplitude, and normalized amplitude quotients. The authors classified utterances of the five Spanish vowels from the PC-GITA database [Oroz 14] with an SVM classifier, and reported accuracies of up to 78%, when the features computed from the five vowels are concatenated. In [Novo 20], the authors also performed a reconstruction of the glottal source and extracted different features such as the quasi open quotient, the normalized amplitude quotient, and the harmonic richness factor of the glottal signal. The features were extracted from 40 HC subjects and 40 newly diagnosed PD patients, Czech native speakers, who produced sustained phonations of the vowel /ah/. The classification of PD vs. HC subjects was performed with an SVM. The AUC reported by the authors was 0.78.

Regarding the articulation analysis, the authors in [Rusz 13] considered several features such as the VSA, the VAI, the formant frequencies (F_1 and F_2), and the ratio F_2i/F_2u. The extracted features were used to classify utterances of 20 early PD patients and 15 aged-matched HC Czech native speakers. The patients pronounced sustained phonations of the Czech vowel /ih/, the repetition of short sentences, the reading of a text with 80 words, and a monologue. The monologue was the most accurate task to differentiate speech of early PD patients and HC speakers, with an accuracy of up to 80%. In [Novo 14], the authors modeled six articulatory deficits in PD: vowel quality, coordination of laryngeal and supra-laryngeal activity, precision of consonant articulation, tongue movement, occlusion weakening, and speech timing. The authors considered DDK exercises from the same data from [Rusz 13], and reported an accuracy of 88% discriminating between PD patients and HC speakers, using an SVM classifier. An additional articulation model was proposed in [Oroz 16b], to model the difficulty of PD patients to start/stop the vocal fold vibration in continuous speech. The model was based on the energy content in the transitions from unvoiced to voiced and from voiced to unvoiced segments. The authors classified PD patients and HC speakers with speech recordings in three different languages (Spanish, German, and Czech), and reported accuracies ranging from 80% to 94% depending on the language. A combined phonation and vowel articulation model was considered in [Aria 17]. Phonation features included perturbation measures such as jitter, shimmer, APQ, and PPQ. Articulation features included formant frequencies, the VSA, the FCR, and MFCCs. The authors classified PD patients and two groups of HC subjects: (1) young speakers with ages ranging from 22 to 50 years and (2) HC subjects with age matched compared to the PD patients. The aim was to analyze the impact of aging when discriminating PD vs. HC subjects. Accuracies of up to 79% were reported, which also concluded that phonation and articulation capabilities are impaired not only due to the presence of PD but also due to the aging process. In [Mont 18] the authors proposed articulation features based on VOT segments to classify 27 PD patients and 27 HC subjects, Spanish speakers. Different temporal and spectral features were extracted from the VOT from DDK exercises. Temporal features included the VOT duration, the VOT ratio, the vowel variability quotient, and the articulation rate, while the spectral features considered 13 MFCC extracted from the VOT segments. The authors considered an SVM classifier, and reported an accuracy of up to 92.2%. A different articulation analysis were introduced in [Godi 17] to model the dynamics of the amplitude envelope of DDK exercises. The proposed features were based on the permutation entropy and other NLD features computed over the derivatives of the amplitude envelope of DDK exercises. The features were used to classify the speakers from the PC-GITA database [Oroz 14]. The authors reported accuracies over 85% using an SVM classifier. A different articulation model was proposed in [Moro 19a]. The authors considered a forced alignment strategy based on GMM-HMM systems to segment the different phonemes in the utterances, with the aim to train an independent GMM-UBM system for each phoneme. The classification was performed with a threshold of the difference between the posterior probabilities from models created for HC subjects and PD patients. The model was tested with utterances from the PC-GITA database [Oroz 14], the Czech data from [Rusz 13], and an additional data in Spanish language from Madrid (Neurovoz). The authors re-

ported accuracies of up to 81% for the PC-GITA and the Neurovoz data, and of up to 94% for the Czech data. The study from [Moro 19a] was extended in [Moro 19b]. The authors trained individual GMM-UBM systems for specific phonological classes such as fricatives, liquids, nasals, plosives, and vowels. The same classification strategy was considered. The proposed methods were tested with the same data from [Moro 19a], using the same classification strategy. The authors reported accuracies of up to 85% for the PC-GITA corpus, 89% for the Neurovoz data, and 94% for the Czech data. In [Kara 19] the authors proposed a combined phonation and articulation analysis based on the empirical mode decomposition to compute new features called intrinsic mode function cepstral coefficients. The authors claimed that the first four intrinsic mode functions give information about vocal tract whereas higher order functions give information about the vocal fold vibrations. The features were used to classify PD and HC subjects using the data from [Saka 13], and a subset of 20 PD patients and 25 HC subjects from the PC-GITA corpus [Oroz 14]. The authors considered an SVM classifier. The reported accuracy was 95% for the data from [Saka 13], and 93% for the subset of the PC-GITA corpus [Oroz 14]. The results are not conclusive due to the small amount of data considered. In addition, the authors performed an independent cross-language test, i.e., training with the recordings from one corpus and test with the utterances from the other one, obtaining an accuracy of only 58.3%. These independent cross-language experiments have also been performed in other studies [Vasq 17a, Oroz 16b]; however, the reported accuracy is lower than 60% for all cases.

Deep learning models have also been proposed to model articulation impairments in PD patients. In [Zhan 17] the authors combined perturbation and articulation features with a deep learning model based on autoencoders to classify PD patients and HC subjects. Different acoustic features were used as input for the autoencoder. The bottleneck features from the autoencoder were used to feed a KNN classifier. The authors considered the Turkish PD data from [Saka 13]. The authors reported accuracies of up to 94%; however, the results were slightly optimistic because the hyperparameters of the autoencoder were optimized according to the accuracy obtained in the test set. A deep learning based articulation model was proposed in [Vasq 17a] to model the difficulties of the patients to stop/start the vibration of the vocal folds based on the transition between voiced and unvoiced segments. The segmented transitions were modeled with time-frequency representations based on the short time Fourier transform and the continuous wavelet transform. The time-frequency representations were used as input for a CNN to classify PD patients and HC speakers in three languages: Spanish, German, and Czech, and reported accuracies ranging from 70% to 89%, depending on the language. An additional deep learning model was proposed in [Korz 19] to classify patients with dysarthria and HC speakers. The proposed model consisted of a combination of convolutional and recurrent layers trained with a multitask learning strategy to address two tasks: (1) to detect the presence of dysarthria, and (2) to reconstruct the Mel spectrogram from the input. The authors reported a recall of up to 0.93 in the dysarthria detection. The authors in [Mall 20] classify PD patients and HC subjects using a CNN-LSTM network with a transfer learning strategy. A trained model to classify between ALS and HC speakers was

used to initialize a model to classify a set of 60 PD patients and 60 HC subjects. The authors reported an accuracy of up to 90%.

Regarding prosody analysis, the authors from [Bock 13] considered voiced segments to compute prosody features based on the F_0 contour, energy contour, duration, and pitch periods to classify 88 PD patients and 88 HC speakers, German speakers. The prosody features were combined with acoustic and glottal features. Accuracies of up to 82% were reported. In [Zhao 14] the authors considered prosody features such as the voiced ratio, the average and variance of F_0, and average duration of pauses to classify speech of 5 PD and 7 HC subjects, and to detect emotions in the speech of the patients, since for PD patients, the capacity to produce emotional speech is also reduced [Pint 04]. The utterances were labeled with five different emotions. The authors reported accuracies of up to 73.3% discriminating between PD and HC subjects, and of up to 65.5% by classifying the emotions of the patients, using an SVM classifier. In [Gala 16] the authors considered several prosody features based on the F_0 and intensity variation, and on the speech rate to classify utterances from 98 PD patients and 51 HC subjects, Czech speakers. The authors considered a RF classifier, and reported accuracies of up to 67%. In [Beru 19] the authors computed several phonation and prosody features to classify the 20 PD patients and 20 HC subjects, Turkish speakers from [Saka 13]. The extracted features were classified with a neural network. Accuracies of up to 86.5% were reported by the authors, by combining different speech tasks using a voting strategy. In [Benb 19] the authors considered the articulation and prosody features introduced in [Hlav 17] to evaluate whether patients with rapid eye behavioral movement disorder (RBD) are classified as PD patients or HC subjects. The extracted features included the entropy, rate, and acceleration of speech timing, the duration of unvoiced stops, the decay of unvoiced fricatives, the relative loudness of respiration, among others. The authors trained an SVM classifier with data from 30 PD patients and 50 HC subjects, Czech native speakers, which achieved and accuracy of 85%. Then, a set of 50 RBD patients were tested by the classifier, of which 66% of them were assigned to the PD group.

Classical feature extraction approaches have been also used to train speaker models to represent the presence of PD in a group of speakers. For instance, in [Garc 17] the authors considered phonation, articulation, and prosody features to train speaker models based on i-vectors [Deha 11]. Reference i-vectors were computed for HC and PD subjects. The cosine distance between the reference i-vectors and a test speaker i-vector was used to classify PD patients and HC subjects. Accuracies of up to 78% were reported. In [Moro 18] the authors combined phonation and articulation features with state-of-art speaker recognition techniques. The authors considered GMM-UBMs and i-vectors trained with different phonation and articulation features. The authors reported accuracies of up to 87% in the PC-GITA database [Oroz 14]. In [Wu 18] the authors proposed a strategy based on Mel-scale spectrograms and K-means clustering to create reference models for PD and HC speakers. The features of the speaker to be evaluated were encoded by reference models based on a multiplication between the features and the centroids of the clusters. The encoded features were used to classify speech utterances from 27 PD patients and 446 HC subjects using an SVM classifier. The authors considered a data augmentation strategy called adaptive synthetic sampling to balance the data in the training set. The authors

reported an accuracy of up to 90.2%. In [Lope 19] the authors considered phonation, articulation, and prosody features to train speaker models based on a Fisher vector approach. The extracted Fisher vectors fed a linear SVM classifier to discriminate between PD and HC subjects. The authors reported an accuracy of up to 84% for the PC-GITA corpus. Recently, in [Moro 20] the authors considered novel speaker models based on the state-of-the-art speaker recognition technique called X-vectors [Snyd 18], obtained as embeddings from a time delay neural network trained with MFCC for a speaker recognition application. The authors classify the 43 PD patients and 46 HC subjects from the Neurovoz corpus [Moro 19b]. The extracted X-vectors were classified with probabilistic linear discriminant analysis (PLDA) classifier, which achieved an accuracy of up to 90%.

In addition to the previous studies, an important aspect considered in the classification of PD vs. HC subjects is the acoustic conditions of the speech signals. For instance, in [Vaic 17] the authors evaluated the effect of non-controlled acoustic conditions on several algorithms to detect PD from speech. The authors considered speech signals from 99 PD patients and 98 HC Lithuanian speakers, which were recorded with a high quality cardioid microphone and a smartphone. The speakers pronounced the sustained vowel /ah/ and short sentences. The authors compared the performance of 18 different acoustic feature sets. Most of the feature sets were computed using the OpenSMILE toolkit [Eybe 15]. The authors reported accuracies of up to 80.7% when the high quality microphone was considered. The performance was reduced around 5% when the speech signals captured from the smartphone were considered; however, the experiments were performed in a matched scenario, i.e., the train and test sets were formed with recordings captured in the same acoustic conditions. A more realistic approach should consider the mismatched conditions. In [Vasq 17b] the authors analyzed the impact of several acoustic conditions on the performance of several methods to classify PD and HC subjects both in matched and mismatched scenarios. The extracted features considered the phonation model from [Oroz 15], the articulation model based on the transitions between voiced and unvoiced segments from [Oroz 16b], a speaker model based on super-vectors [Bock 13], and features extracted using OpenSMILE [Eybe 15]. The acoustic conditions considered include saturation, dynamic compression, additive white Gaussian noise, different environmental background noises, audio codecs, and real telephone channels. Accuracies of up to 82% were reported with the noise-free speech signals. The authors concluded that background noise produced the highest impact in the accuracy of the models, especially in mismatched conditions. In addition, the impact of telephone channels, dynamic compression, and saturation was not as critical as in the case produced by background noise. In [Corr 19] the authors evaluated the difference to classify PD and HC subjects from data in controlled and non-controlled acoustic conditions. The PC-GITA corpus [Oroz 14] was considered as the controlled database. The WSM corpus [Corr 18], which contains data from vlogs from 34 PD patients and 19 HC subjects was used as the non-controlled acoustic conditions corpus. Both corpora were classified with a deep learning model based on a CNN-LSTM network with self attention mechanism. The authors reported an UAR of 0.94 for PC-GITA and of 0.83 for WSM, the later one obtained in a cross domain experiment, using the PC-GITA corpus for training. In [Rusz 18a] the authors evaluated the impact of speech

captured with a smartphone to detect speech deficits in patients with RBD, and early untreated PD patients. Speech data were recorded from 50 RBD, 30 PD patients, and 30 HC subjects, all of them Czech speakers. The authors computed phonation, articulation, and prosody features, and classified the pairs of the three different groups with a binary logistic regression algorithm. The results indicated that a combination of three features representing monopitch, inappropriate silences, and decreased rate was able to discriminate between PD and HC subjects with an AUC of 0.85. In addition, the same features were able to classify the patients with the sleep disorder and HC subjects with an AUC of 0.69. The results also suggested that there are strong correlations between the features computed from the smartphone signals and those computed from professional microphone recordings. A different approach using data captured from smarthphones was proposed in [Zhan 19], where the authors considered non-speech body sounds such as breathing, clearing throat, and swallowing to classify PD vs. HC subjects. The non-speech body sounds were modeled using a deep learning strategy based on ResNet architectures. The proposed method achieved an accuracy of up to 83.3% in a dataset formed with 321 PD patients and 569 HC subjects. The results obtained were comparable to the ones obtained with normal speech sounds. However, the speaker independence was not guaranteed in the training process, which leads to biased and optimistic results.

4.1.2 Automatic Evaluation of the Neurological State of Patients

The evaluation of the neurological state of the patients is focused on the automatic estimation of clinical scales assigned to the patients by neurologist experts. The most common scale that has been automatically predicted is the UPDRS scale, especially the part III, which is related to the motor symptoms of the patients [Tsan 10, Eski 12, Baye 13, Schu 15, Gros 15, Oroz 16a]. The assessment of the neurological state has been performed in two ways: (1) a regression approach, where the clinical score is predicted [Tsan 10, Eski 12, Baye 13, Schu 15, Gros 15, Oroz 16a], and (2) a multiclass classification strategy, where the patients are divided into three or four groups according to their neurological state [Bock 13, Oung 18, Vasq 18a].

One of the first approaches to assess the neurological state of the patients was proposed in [Tsan 10]. The authors considered recordings of sustained vowels, which were modeled using phonation features. The UPDRS-III scores were estimated using different regression techniques. The speech of 42 PD patients was recorded once per week during six months. Neurologist experts evaluated the patients three times along the study. The weekly UPDRS scores were obtained by the authors using a piecewise linear interpolation. The best result reported corresponds to a MAE of 7.5 points in the prediction of the total the UPDRS scale. The scores of the motor section in the UPDRS (UPDRS-III) were estimated with a MAE of 6 points. The same data and features from [Tsan 10] were considered in [Eski 12] to predict the neurological state of the patients based on the total UPDRS score and its third part. The authors used regression algorithms based on SVRs and neural networks. The performance of the regression methods was evaluated according to the MAE and the Pearson correlation coefficient (r). The authors reported correlations of up to 0.65 (MAE=6.32) to pre-

dict the total UPDRS score, and of up to 0.63 (MAE=4.96) to predict only the motor part of the scale. In [Cast 14], the authors also considered the same data and features from [Tsan 10] to predict both the total UPDRS scale and its third part using a combination of genetic programming and semantic genetic operators. Phonation features computed from sustained vowels were used to train the genetic algorithms. The authors reported a MAE of 7.5 by predicting the total UPDRS score, and a MAE of 5.5 predicting only the motor part of the scale. In [Nila 18] the authors considered the same data and features from [Tsan 10] to predict both the total UPDRS score and the motor section of the scale (part III). The authors proposed a novel regression strategy called incremental SVR, combined with a clustering method based on self-organizing maps, and a dimensionality reduction strategy based on non-linear iterative partial least squares. The authors reported a Pearson correlation coefficient of up to 0.885 (MAE=0.466) for the UPDRS-III score and of up to 0.868 (MAE=0.497) for the total scale. The same authors in [Nila 19] predicted the total UPDRS scale and its third part in the same data from [Tsan 10]. The authors proposed a technique based on an ensemble of adaptive neuro-fuzzy inference system networks with singular value decomposition for dimensionality reduction and a clustering strategy based on self-organizing maps. The authors reported a Pearson correlation coefficient of up to 0.956 (MAE=0.491) for the UPDRS-III score and of up to 0.967 (MAE=0.480) for the total scale. Despite the results from [Tsan 10, Eski 12, Cast 14, Nila 18] and [Nila 19] look promising, the results are optimistic and biased for two main reasons: (1) the authors do not guarantee the speaker independence because they include data from the same speaker both in the train and test sets, and (2) most of the labels of the considered data were interpolated, which makes the data not reliable for the addressed problem. In addition, the MAE and the Pearson correlation coefficient are not the most reliable performance metrics for the addressed problem, especially the MAE, which only makes sense when there is a baseline to compare the performance of the models.

A different approach was introduced in [Baye 13]. The authors considered phonation features such as jitter, shimmer, and the HNR to predict the UPDRS-III score of 168 PD patients, English speakers, who performed different tasks such as the sustained phonation of the vowel /a/, DDK exercises, and a read text. The authors reported a MAE of 5.5 points using a ridge regression algorithm; however, as we mention previously, the MAE is not a reliable metric to evaluate the performance of the addressed problem. Later on, the *2015 computational paralinguistic challenge (ComParE)* [Schu 15] had one of the sub-challenges about the automatic estimation of the neurological state of PD patients according to the MDS-UPDRS-III score. The baseline of the challenge was computed using features extracted with the openSMILE toolkit [Eybe 15] and the regression was performed with an SVR. The reported baseline for the challenge corresponded to a Spearman's correlation (ρ) of 0.39. The winners of the challenge [Gros 15] reported a Spearman's correlation of 0.65 when grouping automatically the speech tasks per speaker and using Gaussian processes and deep neural networks to perform the prediction of the clinical score. In [Oroz 16a] the authors predicted the neurological state of PD patients according to the MDS-UPDRS-III scale combining articulation and intelligibility features. The articulation was modeled by computing the energy content in the transitions from unvoiced to voiced (onset) and from voiced to unvoiced (offset) segments. The authors reported

Spearman's correlations of up to 0.69 between the original MDS-UPDRS-III scores and the predicted ones using an SVR. In [Tu 17b] the authors aimed to predict the MDS-UPDRS-III score of 61 PD patients using several features based on spectral and glottal analysis. The prediction was performed using a non-parametric regression strategy based on the Hausdorff distance between a speaker from the test set and the speakers in the training set. The neurological state of the patients was predicted with a Pearson's correlation of up to 0.58. The authors also predicted the dysarthria level of 56 patients, English speakers, and reported a Pearson's correlation of up to 0.78. In [Rame 17] the authors considered acoustic features computed with the openSMILE toolkit [Eybe 15] to predict the neurological state of the patients according to the MDS-UPDRS-III score. The authors proposed a feature selection algorithm based on the maximal relevance and minimal redundancy based on correlations (mRMRc) criterion. Gaussian mixture regression and SVRs were considered to predict the MDS-UPDRS-III scale. The authors reported a Spearman's correlation of up to 0.52 and concluded that the most informative features were those related to spectral flatness and the energy content distributed in the spectrum. In [Smit 17a] the authors aimed to predict motor, cognitive, and depressive symptoms of 35 PD patients, English speakers. The motor state was predicted based on the UPDRS score, the cognitive state was predicted based on the Montreal cognitive assessment (MoCA) scale, and the depressive state was evaluated with the geriatric depression scale. The clinical scales were predicted with articulation features such as formant frequencies and the derivatives of MFCCs; and prosody features based on the phoneme rate. The extracted features were used to train a Gaussian staircase regression algorithm. The authors reported moderate correlations predicting motor severity (ρ=0.42) and global cognition (ρ=0.52) but not depression (ρ=-0.21).

An additional approach was proposed in [Aria 16] to track the disease progression per speaker. The authors created individual speaker models for seven PD patients recorded in five sessions during three years. The speaker models followed a GMM-UBM approach trained with articulation features. The Bhattacharyya distance was computed between the speaker model and a UBM trained with utterances of 62 PD patients and 50 HC, Spanish speakers. The distance measure was correlated with the MDS-UPDRS-III scale assigned to each patient to assess the progress of the disease. The authors reported an average Pearson's correlation of 0.60 for all patients. Another study to evaluate the progress of the neurological state of the patients was presented in [Gala 18]. The authors predicted the changes in the UPDRS score for 51 PD patients, Czech native speakers, recorded in two sessions within two years. The authors compute several phonation features for sustained vowels. The difference of the UPDRS score between the two sessions for all patients was predicted with a Gradient boosted trees regressor. The authors reported an estimated error rate (EER) of 11% (MAE=1.7) for the prediction of the part IV of the UPDRS, and an EER of 26% (MAE=7.3) for the prediction of the changes of the part III of the UPDRS.

Regarding the evaluation of the neurological state based on grouping patients in several stages of the disease (initial, intermediate, and severe), in [Bock 13] the authors considered phonation, articulation and prosody features to classify PD patients in three levels of the disease, and reported accuracies of up to 46.6%.

4.1.3 Automatic Evaluation of the Dysarthria Severity of Patients

Although the MDS-UPDRS-III evaluates motor skills including the movement of hands and arms, gait, and posture, among others, it is not suitable nor fair to assume that the scale can accurately be predicted only based on speech recordings. To evaluate the impact of PD solely on the speech, a scale to evaluate speech would be a valuable tool. Several studies have considered the application of scales to assess only the speech deficits of PD patients [Skod 11a, Pate 16]. There are several studies to predict the speech deficits of the patients based on clinical scores. In [Tsan 14] the authors used phonation features to evaluate the response of 14 PD patients to the Lee Silverman voice treatment as *acceptable* or *unacceptable*. The authors considered only information from the sustained vowel /ah/, and reported accuracies close to 90% discriminating between acceptable vs. unacceptable utterances. In [Tu 17a] the authors proposed a deep learning model to assess dysarthric speech. The model aimed to predict the dysarthria severity adding an intermediate interpretable hidden layer that contains four perceptual dimensions: nasality, vocal quality, articulatory precision, and prosody. The authors evaluated the performance of the model on a dysarthric speech corpus and showed that their approach provided an interpretable output highly correlated (ρ=0.82) with a subjective evaluation performed by speech and language pathologists. In [Vasq 18b] the authors computed phonation, articulation, prosody, and intelligibility features to predict the dysarthria severity of 68 PD patients and 50 HC subjects according to the introduced m-FDA scale, explained in Section 3.1.2. The dysarthria level was estimated using several regression models. In addition, speaker models based on i-vectors were also explored. The results indicated that articulation features were the most accurate to predict the m-FDA score, obtaining Spearman's correlations of up to 0.69 between the predicted scores and those assigned by the phoniatricians. In [Cern 17] the authors modeled the composition of non-modal phonations, i.e., voice quality spectrum, in PD using a deep learning approach to compute phonological posteriors from the speech signal. Those posteriors were used to assess the dysarthria level of the speakers from the PC-GITA corpus [Oroz 14]. The authors correlated (ρ=0.56) the phonological posteriors and the subjective evaluation performed by speech therapists following the m-FDA scale [Vasq 18b]. In [Beri 17] the authors computed prosody and articulation features from interviews performed by Muhammad Ali to assess the progress of the disease. The results reveal that Mr. Ali's speaking rate sharply declined over time (Pearson correlation of -0.574) as did his ability to clearly articulate vowels. In [Laar 17] the authors considered an i-vector approach to predict the dysarthria level of 129 patients with several diseases, including 31 PD patients. The labeling of the dysarthria level was based on the DEB. All patients were native French speakers, and they were asked to pronounce a read text with 550 phonemes. The prediction of the dysarthria level considered three dimensions of the speech impairments: articulation, intelligibility, and the total severity. The proposed method consisted of training an i-vector extractor with 19 linear frequency cepstral coefficients, and then use those extracted i-vectors to train an SVR. A Spearman's correlation coefficient of 0.88 was reported for the prediction of the total dysarthria level. The correlations reported for the

specific levels of intelligibility and articulation were of 0.84 and 0.87, respectively. In [Vasq 18a] the authors improved the articulation model introduced in [Vasq 17a] to predict the neurological state and the dysarthria level of the patients from the PC-GITA corpus [Oroz 14]. The model based on CNNs was trained using a multitask learning strategy to jointly classify PD patients and HC subjects, PD patients in different stages of the disease according to the MDS-UPDRS-III score, and to classify the participants into several groups based on their level of dysarthria according to the m-FDA scale. The proposed model was able to classify the PD patients and HC subjects with accuracies of up to 89%, to classify the patients according to their neurological state with accuracies of up to 55%, and to classify the subjects according to their level of dysarthria with accuracies over 43.3%.

An additional approach to monitor the disease progression per speaker was introduced in [Aria 18a]. The authors considered the speaker models based on GMM-UBMs [Aria 16] and i-vectors to monitor the progression of the level of dysarthria in a longitudinal study. The speaker models based on i-vectors and GMM-UBMs were trained with phonation, articulation, and prosody features. The speaker models were tested in two different scenarios: (1) a test set with utterances from seven patients captured in five sessions distributed from 2012 to 2016, and (2) the At-home database, described in Section 3.3.3. The authors considered the effect of different communication channels (Skype, Hangouts, mobile phone, and land-line) to test the suitability of the speaker models to perform a remote monitoring of the speech impairments of the patients. The speaker models were able to monitor the progression of the m-FDA score in the at-home data with a Spearman's correlation of up to 0.55. The results for the long-term progression indicated a correlation of up to 0.77.

4.1.4 Main Outcomes from the Literature

Different applications have been considered for the speech assessment of PD patient. A summary of the literature about the considered applications within the last years is shown in Figure 4.2 and Table 4.1. Most of the papers are focused on the classification of PD vs. HC subjects, the assessment of the neurological state of the patients, or the evaluation of the severity of the speech impairments of the patients, following a dysarthria scale. The number of papers about these applications has increased within the years. Particularly, regarding the assessment of the neurological state of the patients and the prediction of their dysarthria severity, there are a couple of studies focused on longitudinal evaluation of the patients for a continuous and individual monitoring of the disease progression [Aria 18a, Gala 18]. Other studies are focused on evaluating the effect of medication in the speech production of patients [Rusz 16, Pomp 20, Nore 20, Hemm 20], the assessment of the intelligibility of patients [Dima 17], or the prediction of whether patients with RBD are classified as PD patients or HC subjects [Benb 19].

There are other applications that are not well not enough studied yet and that can have a direct impact for the medical community, patients, or caregivers. Those applications include the automatic assessment of non-motor impairments exhibited by patients such as depression or cognitive decline, which highly affect the quality of life of the patients, and their communication capabilities. A couple of studies to

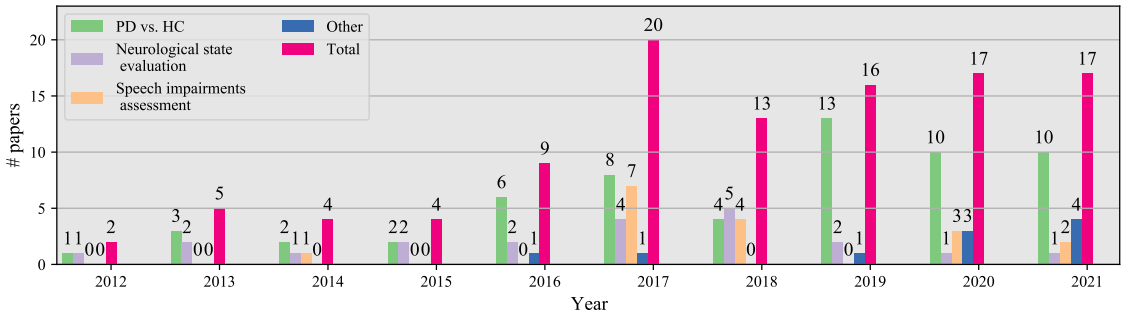

Figure 4.2: Different applications addressed in the literature for speech assessment of PD.

Table 4.1: Different applications addressed in the literature for speech assessment of PD.

References	Application
[Tsan 12], [Rusz 13], [Saka 13], [Bock 13], [Novo 14], [Zhao 14], [Oroz 15], [Vill 15], [Bela 16], [Gala 16], [Hemm 16], [Meky 16], [Nara 16], [Oroz 16b], [Aria 17], [Garc 17], [Godi 17], [Rusz 17], [Vaic 17], [Vasq 17b], [Vasq 17a], [Zhan 17], [Mont 18], [Moro 18], [Vasq 18a], [Wu 18], [Aror 19], [Benb 19], [Corr 19], [Kara 19], [Korz 19], [Lope 19], [Moro 19a], [Moro 19b], [Rued 19], [Saka 19], [Vasq 19c], [Zhan 19], [Wodz 19], [Moro 20], [Novo 20], [Visw 20], [Mall 20], [Kodr 20], [Oroz 20b], [Jean 20], [Kara 20], [Rios 20], [Mill 20], [Vasq 21c], [Rios 21], [Garc 21], [Quan 21], [Np 21], [Kara 21b], [Kara 21a], [Amat 21], [Vasq 21a], [Jean 21]	Classification PD vs. HC
[Tsan 10], [Eski 12], [Baye 13], [Bock 13], [Cast 14], [Gros 15], [Schu 15], [Aria 16], [Oroz 16a], [Garc 17], [Rame 17], [Smit 17b], [Tu 17b], [Aria 18a], [Nila 18], [Oung 18], [Gala 18], [Vasq 18a], [Nila 19], [Vasq 19c], [Hemm 20], [Sech 21]	Assessment of the neurological state
[Tsan 10], [Beri 17], [Cern 17], [Garc 17], [Laar 17], [Rusz 17], [Tu 17a], [Tu 17b], [Aria 18a], [Vasq 18b], [Vasq 18a], [Gala 18], [Oroz 20b], [Kara 20], [Mill 20], [Vasq 21b], [Kara 21b]	Assessment of speech impairments
[Rusz 16], [Dima 17], [Benb 19], [Pomp 20], [Nore 20], [Hemm 20], [Pere 21b], [Pere 21c], [Garc 21], [Roma 21]	Others

address such applications have been recently performed. [Garc 21, Pere 21b]. Other application include the detection of patients in prodromal stages of the disease, which would benefit the development of neuroprotective therapies [Post 15]. Patients in prodromal stages may include subjects with genetic mutations responsible for producing PD but with no clinical signs of the disease. Another potential application is the discrimination between PD and other neurological disorders with similar symptoms such as Huntington's disease or essential tremor [Rusz 15].

The studies addressed in the literature considered different speech tasks for the assessment of the patients. A summary of the most common addressed tasks is shown in Figure 4.3 and Table 4.2. The most common speech tasks include the phonation of sustained vowels, reading isolated sentences, or read texts, DDK tasks like the rapid repetition of /pa-ta-ka/, or spontaneous speech. The use of continuous speech tasks like monologues has increased within the last years, mainly because the motivation of developing technology for continuous monitoring of the patients. There are additional tasks that have not been considered enough, but that contain suitable information about the disease including phonation of modulated vowels [Hemm 16], non-speech

body sounds like swallowing or coughing [Zhan 19], the sustained phonation of nasal consonants like *m* [Visw 20].

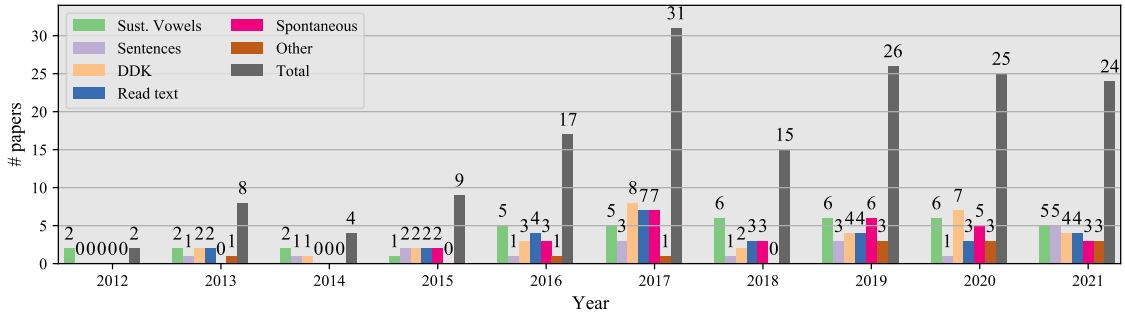

Figure 4.3: Different speech tasks considered in the literature for speech assessment of PD.

Table 4.2: Different speech tasks considered in the literature for speech assessment of PD.

References	Speech task
[Tsan 10], [Eski 12], [Tsan 12], [Baye 13], [Rusz 13], [Cast 14], [Tsan 14], [Oroz 15], [Bela 16], [Hemm 16], [Meky 16], [Nara 16], [Rusz 16], [Aria 17], [Rusz 17], [Vaic 17], [Vasq 17b], [Zhan 17], [Gala 18], [Moro 18], [Nila 18], [Oung 18], [Vasq 18b], [Wu 18], [Aror 19], [Beru 19], [Kara 19], [Nila 19], [Saka 19], [Wodz 19], [Novo 20], [Mall 20], [Pomp 20], [Kara 20], [Hemm 20], [Vasq 21b], [Quan 21], [Np 21], [Kara 21a], [Jean 21]	Sustained vowels
[Saka 13], [Zhan 17], [Beru 19], [Korz 19], [Kodr 20], [Kara 20], [Kara 21b], [Amat 21]	Isolated words
[Saka 13], [Zhao 14], [Schu 15], [Vill 15], [Oroz 16b], [Vaic 17], [Tu 17a], [Zhan 17], [Vasq 18b], [Beru 19], [Moro 19a], [Vasq 19c], [Moro 20], [Pomp 20], [Oroz 20b], [Sech 21], [Quan 21] [Kara 21b], [Vasq 21a], [Jean 21]	Isolated sentences
[Bock 13], [Baye 13], [Gros 15], [Schu 15], [Aria 16], [Gala 16], [Oroz 16b], [Oroz 16a], [Cern 17], [Dima 17], [Garc 17], [Laar 17], [Rame 17], [Rusz 17], [Tu 17a], [Aria 18a], [Moro 18], [Vasq 18b], [Benb 19], [Lope 19], [Moro 19b], [Vasq 19c], [Pomp 20], [Oroz 20b], [Rios 20], [Garc 21], [Sech 21], [Roma 21], [Jean 21]	Read texts
[Bock 13], [Baye 13], [Novo 14], [Gros 15], [Schu 15], [Oroz 16b], [Oroz 16a], [Rusz 16], [Cern 17], [Garc 17], [Godi 17], [Laar 17], [Rame 17], [Rusz 17], [Vasq 17a], [Tu 17a], [Moro 18], [Vasq 18b], [Lope 19], [Moro 19a], [Moro 19b], [Rued 19], [Vasq 19c], [Moro 20], [Mall 20], [Pomp 20], [Nore 20], [Oroz 20b], [Rios 20], [Mill 20], [Vasq 21c], [Rios 21], [Vasq 21b], [Vasq 21a],	DDK
[Gros 15], [Schu 15], [Oroz 16b], [Oroz 16a], [Rusz 16], [Beri 17], [Garc 17], [Laar 17], [Rame 17], [Rusz 17], [Vasq 17b], [Tu 17a], [Aria 18a], [Moro 18], [Vasq 18b], [Benb 19], [Corr 19], [Lope 19], [Moro 19b], [Moro 19a], [Vasq 19c], [Mall 20], [Pomp 20], [Oroz 20b], [Rios 20], [Jean 20], [Pere 21b], [Pere 21c], [Jean 21]	Spontaneous speech
[Hemm 16], [Zhan 19], [Visw 20], [Nore 20], [Garc 21]	Others

There are additional speech tasks that can be explored to find markers about other aspects present in the speech of patients, especially those focused on the evaluation of non-motor symptoms that affects the speech and language production. These speech tasks include retelling, i.e., the examiner tells a story to the patient, who has to tell later what (s)he remembers about the story. This task is suitable to evaluate aspects

about memory load and hesitations of the patients [Garc 18a, Garc 21]. Other task that can be explored is image description [Nore 20], where a picture with different actions in it is shown to the patients, who have to describe it with as much details as possible. This type of task is useful to evaluate aspects of cognitive decline in the patients.

The studies addressed in the literature have also considered or proposed different methods to evaluate the speech of PD patients. The methods were divided in different categories such as phonation, articulation, prosody, intelligibility, those based on deep learning, and others. A summary of the use of those methods is observed in Figure 4.4 and Table 4.3. In the initial years, most of the studies were only focused on phonation or articulation analyses. In the last five years, works on deep learning started to appear, showing interesting and accurate results in modeling different aspects of the speech of PD patients. Additional methods that appear in the literature that have not been enough explored include the use of phonological features [Cern 17, Oroz 20b], which is deeply explored in the scope of this thesis and explained with details in Section 4.3. Additional methods explored in the literature include speaker models based on speaker recognition methods such as i-vectors [Aria 18a, Garc 17, Laar 17], or the X-vectors [Moro 20, Jean 20].

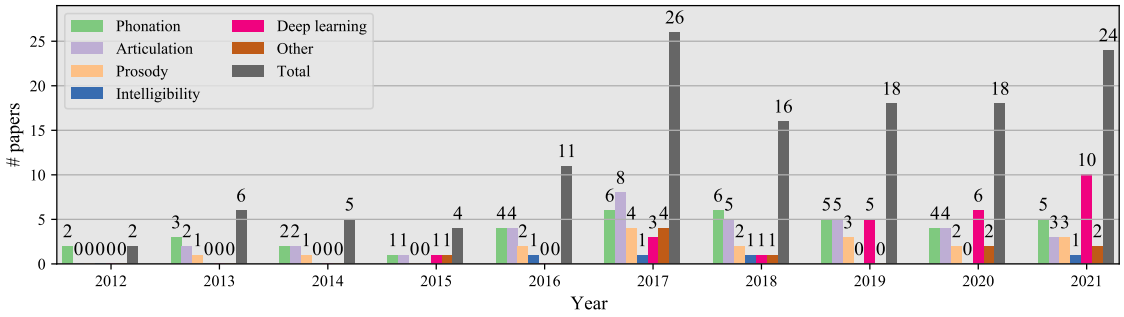

Figure 4.4: Different methods considered in the literature for speech assessment of PD.

4.2 Acoustic Analysis of Speech

This section aims to cover classical acoustic analyses methods designed to characterize the speech of PD patients. We include a detailed description of phonation, articulation, and prosody features that are used to model different phenomena exhibited in the speech of the patients. For instance, phonation analysis give information about the capability to control respiration and expel air from the lungs to make the vocal folds vibrate. Articulation analysis is considered to model the capability of patients to control the movement of several articulators to produce speech e.g., tongue, lips, jaw, and velum. Finally, prosody analysis provides information about the intonation, loudness, and timing in continuous speech. The analysis of these three dimensions of speech was included in the Neurospeech software [Oroz 18], released to model the speech of PD patients. In addition, the implementation of these features is avail-

Table 4.3: Different methods considered in the literature for speech assessment of PD.

References	Methods
[Tsan 10], [Tsan 12], [Eski 12], [Baye 13], [Bock 13], [Saka 13], [Cast 14], [Tsan 14], [Oroz 15], [Bela 16], [Hemm 16], [Meky 16], [Nara 16], [Aria 17], [Garc 17], [Rusz 17], [Tu 17a], [Vasq 17b], [Zhan 17], [Aria 18a], [Gala 18], [Moro 18], [Nila 18], [Oung 18], [Vasq 18b], [Aror 19], [Kara 19], [Nila 19], [Beru 19], [Lope 19], [Saka 19], [Novo 20], [Visw 20], [Oroz 20b], [Hemm 20], [Vasq 21b], [Moro 21], [Quan 21], [Np 21], [Kara 21b]	Phonation
[Bock 13], [Rusz 13], [Novo 14], [Zhao 14], [Vill 15], [Aria 16], [Oroz 16b], [Oroz 16a], [Rusz 16], [Aria 17], [Beri 17], [Garc 17], [Godi 17], [Rusz 17], [Smit 17b], [Vasq 17a], [Vasq 17b], [Aria 18a], [Mont 18], [Moro 18], [Vasq 18b], [Kara 19], [Benb 19], [Lope 19], [Moro 19b], [Rued 19], [Oroz 20b], [Kara 20], [Hemm 20], [Mill 20], [Garc 21], [Quan 21], [Amat 21]	Articulation
[Bock 13], [Zhao 14], [Gala 16], [Beri 17], [Garc 17], [Rusz 16], [Rusz 17], [Smit 17b], [Aria 18a], [Vasq 18b], [Benb 19], [Beru 19], [Lope 19], [Nore 20], [Oroz 20b], [Garc 21], [Quan 21], [Roma 21]	Prosody
[Oroz 16a], [Dima 17], [Parr 18], [Roma 21]	Intelligibility
[Gros 15], [Vasq 17a], [Tu 17b], [Zhan 17], [Vasq 18b], [Corr 19], [Zhan 19], [Korz 19], [Vasq 19c], [Wodz 19], [Moro 20], [Mall 20], [Pomp 20], [Jean 20], [Rios 20], [Vasq 21c], [Pere 21b], [Rios 21], [Pere 21c], [Vasq 21b], [Moro 21], [Quan 21], [Kara 21a], [Vasq 21a] [Jean 21]	Deep learning
[Schu 15], [Cern 17], [Laar 17], [Rame 17], [Vaic 17], [Wu 18], [Kodr 20], [Oroz 20b], [Mill 20], [Garc 21], [Sech 21]	Others

able as an open source toolkit called *Disvoice*, available online[1]. We additionally described in this section a set of features extracted with the OpenSMILE [Eybe 15] toolkit, which was designed to extract features to recognize paralinguistic aspects from speech.

The features described in this section are considered as a baseline and complement for the proposed methods in this thesis for speech assessment of PD patients: the phonological analysis described in Section 4.3 and the representation learning analysis described in Section 4.4. The reliability of all these methods is evaluated in two different applications: the automatic classification of PD patients and HC subjects, and the prediction of the dysarthria severity of the speakers, according to the introduced m-FDA scale, described in Section 3.1.2.

4.2.1 Phonation Features

The phonation features are used to model abnormal patterns in the vocal fold vibration and are extracted from the voiced segments, where there is vibration of the vocal folds. They can be computed in sustained vowels, in DDK exercises, and in continuous speech signals. The phonation analysis comprises seven features computed for short-time frames of the speech signal.

[1]Disvoice: python framework to extract features from speech https://github.com/jcvasquezc/DisVoice

Jitter and shimmer: these two features describe temporal perturbations in the fundamental frequency and amplitude of the speech signal, respectively. Jitter is computed according to Equation 4.1. N is the number of frames in the speech utterance, M_f is the maximum of the fundamental frequency, and F_0 corresponds to the fundamental frequency computed on the k-th frame. Shimmer is computed using Equation 4.2, where M_a is the maximum amplitude of the signal, and $A(k)$ corresponds to the maximum amplitude on the k-th frame.

$$\text{Jitter}(\%) = \frac{100}{N \cdot M_f} \sum_{k=1}^{N} |F_0(k) - M_f| \tag{4.1}$$

$$\text{Shimmer}(\%) = \frac{100}{N \cdot M_a} \sum_{k=1}^{N} |A(k) - M_a| \tag{4.2}$$

Amplitude perturbation quotient (APQ): this feature measures the long-term variability of the peak-to-peak amplitude of the speech signal. The computation includes a smoothing factor of 11 voiced periods.

Pitch perturbation quotient (PPQ): similar to APQ, PPQ measures the long-term variability of the fundamental frequency, with a smoothing factor of five periods.

Both APQ and PPQ are computed as the absolute average difference between the amplitude or period values (for APQ or PPQ, respectively) of each frame and the average of its neighbors, divided by the average values for the signal. The perturbation quotients are computed using Equation 4.3, where $L = N - (k-1)$, $D(i)$ is the pitch period sequence when computing the PPQ or the pitch amplitude sequence when computing the APQ. N is the number of frames, k is the length of the moving average (11 for APQ or 5 for PPQ), and $m = (k-1)/2$.

$$\text{PQ} = \frac{1}{L} \sum_{i=1}^{L} \frac{\frac{1}{k} \sum_{j=1}^{k} D(i+j-1) - D(i+m)}{\frac{1}{N} \sum_{n=1}^{N} D(i)} \tag{4.3}$$

Additionally, the first and second derivatives of F_0 are included in the feature set, along with the energy content of the signal. Additional details of the methods can be found in [Oroz 18]. Four statistical functionals are calculated per feature (mean, standard deviation, skewness, and kurtosis), forming a 28–dimensional feature vector per utterance. Figure 4.5 shows an example with two speech signals and their F_0 contour produced by two male speakers with similar age: an HC speaker (left) and a PD patient (right). Note the stability of the F_0 contour for the HC subject compared to the one obtained for the PD patient, who exhibits clear signs of hypokinetic dysarthria because of the uncontrolled movement of his vocal folds.

4.2.2 Articulation Features

Articulation reflects the ability of a speaker to move and put the muscles of the vocal tract in the correct position, on the correct time, and with the appropriate energy and duration while producing speech. The evaluation of articulation in PD patients is performed typically with measurements of the vocal space and with spectral and cepstral analyses. This thesis considers the articulation analysis introduced in [Oroz 16b]

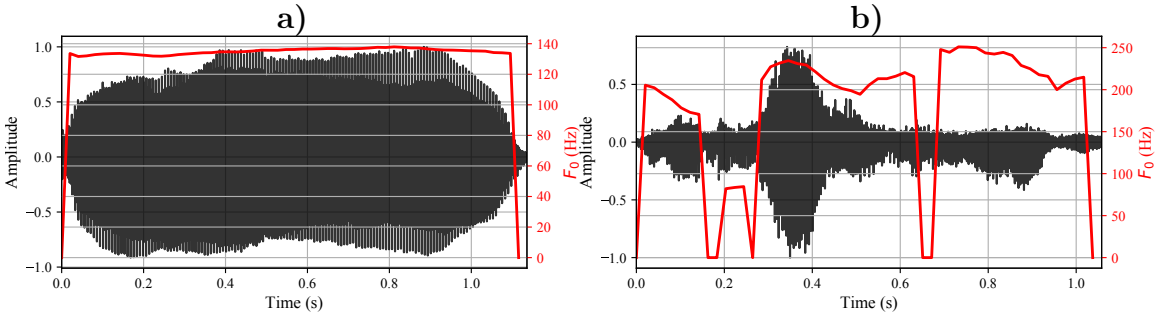

Figure 4.5: Sustained phonations of vowel /a/ and their corresponding F_0 contour for: **a)** a 71 years old healthy speaker with m-FDA = 0, and **b)** a 77 years old PD patient with m-FDA = 41 and MDS-UPDRS-III = 92

to model the difficulties of the patients to start or stop the vocal fold vibration based on an analysis of the energy content in the transitions between voiced and unvoiced segments.

Articulation capabilities of the patients are evaluated considering the energy content in the transition from unvoiced to voiced segments (onset) and from voiced to unvoiced segments (offset). The detection of voiced/unvoiced segments is based on the computation of F_0. Those segments where there are F_0 values are considered as voiced, conversely segments where no F_0 values are found, are labeled as unvoiced. Then, the border between voiced and unvoiced sounds are detected, and 40 ms of the signal are taken to the left and to the right, forming a segment with 80 ms length. Figure 4.6 shows an example of the resulting spectrograms of an onset transition for a HC speaker and a PD patient. Note that the transition for the HC subject is well defined, while the transition for the PD patient is not clearly identified. In addition, note for the case of the PD patient there is a small vibration in the unvoiced part that produces a F_0 around 100 Hz in the left part. This is a characteristic of the lack of control of the vocal fold vibration since the patient was unable to completely stop the vibration for the unvoiced part in the left.

Once the transition is detected, the spectrum of the segment is distributed into 22 critical bands according to the Bark scale according to [Zwic 80]. For frequencies below 500 Hz the bandwidths of the critical bands are constant at 100 Hz, while for medium and high frequencies the increment is proportional to the logarithm of frequency. The distribution of the frequency according to the Bark scale is shown in Equation 4.4 . $\arctan(\cdot)$ is measured in [radians] and f in [Hz].

$$\text{Bark}(f) = 13 \cdot \arctan(0.00076 f) + 3.5 \arctan\left(\frac{f}{7500}\right)^2 \tag{4.4}$$

Finally, after dividing the spectrum into the critical bands, the energy content is computed for each frequency band, forming the Bark band energies (BBE). In addition to the BBE computed upon the transitions, 12 MFCCs and their first two derivatives are also calculated per segment to obtain a smooth representation of the voice spectrum in the transitions. The aim of introducing the MFCCs is to take into account the human auditory perception, which makes it more suitable to model speech signals. A summary of this process is shown in Figure 4.7. At the end, a set

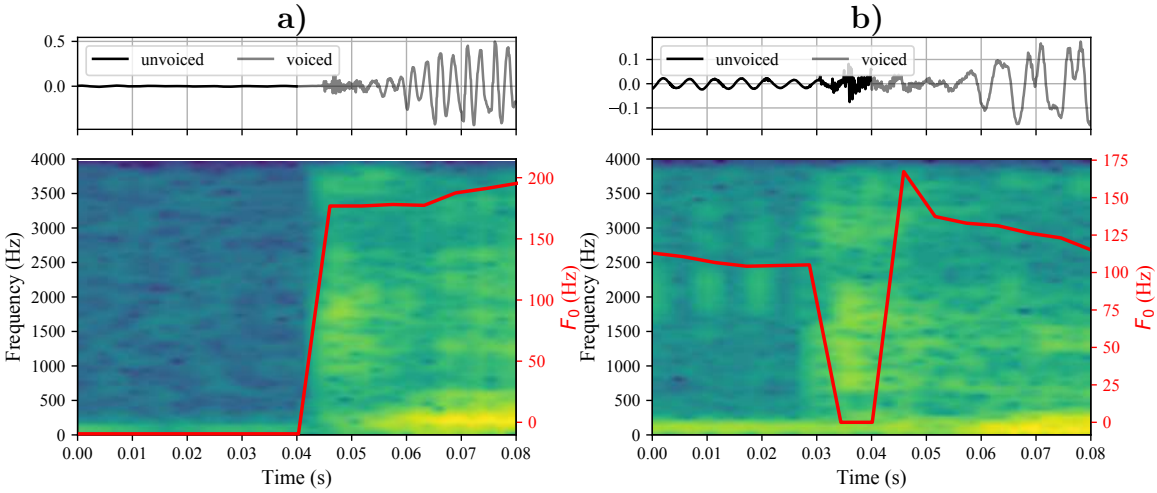

Figure 4.6: Speech signals, fundamental frequency, and spectrograms of an onset transition for **a)** a 71 years old healthy speaker with m-FDA = 0, and **b)** a 77 years old PD patient with m-FDA = 41 and MDS-UPDRS-III=92

of 58 features (22 BBEs+12×3 MFCCs with Δs) are computed for each transition. The features extracted for onset and offset segments are also concatenated forming a vector with 116 components.

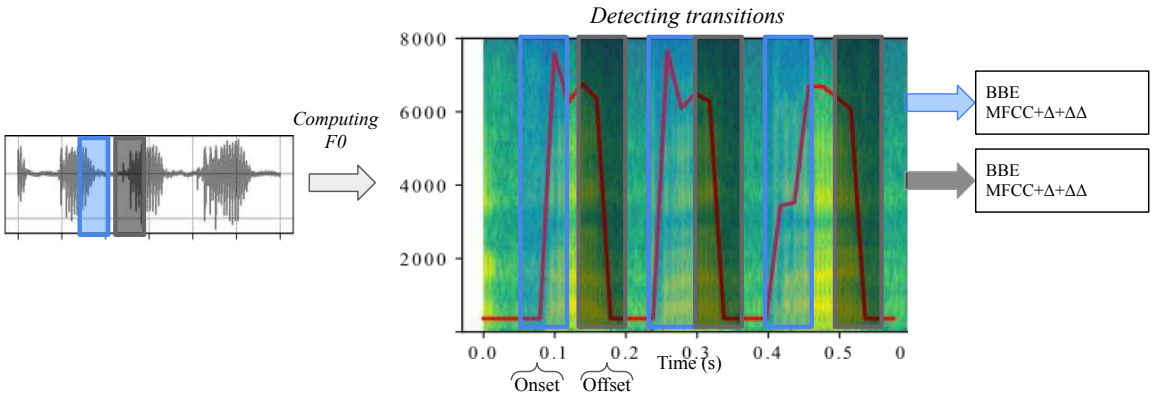

Figure 4.7: Model of articulation features extracted from onset and offset segments.

The articulation features are complemented with the computation of the first and second formant frequencies, and their first two derivatives (six features in total). These additional features allow to represent resonances in the vocal tract and the capacity of the speaker to keep the tongue in a certain position while producing voiced sounds. For instance, the literature has determined that the first derivative of the second formant is highly correlated with the speed of the tongue, representing a precise measure of articulatory speech [West 94]. Finally, similar to the phonation analysis, four statistical functionals are calculated per feature (mean, standard deviation, skewness, and kurtosis), forming a 488-dimensional ((6+116)×4) feature vector per utterance to model the articulation of each speaker.

4.2.3 Prosody Features

This group of features is designed to model timing, intonation, and loudness during the production of natural speech. Prosody has been typically modeled with measures related with duration and the contours of F_0 and energy of the speech signal. Prosody features allow to model the monotonicity, monoloudness, and speech rate disturbances in PD patients. Figure 4.8 shows an example of the prosody differences that exist between an HC subject (left) and a PD patient (right). Note that the variability of the F_0 contour (measured according to the standard deviation of F_0) is lower for the case of the patient, which is translated into monotonicity. In addition, the tilt, i.e., slope of a linear reconstruction, is more negative for the case of the PD patient, i.e., the patient tends to lower the pitch at the end of the words more frequently than the HC subject.

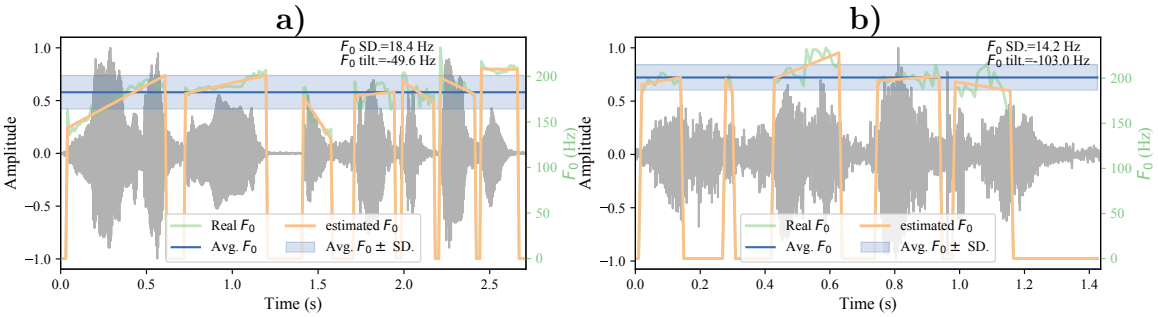

Figure 4.8: Speech signals and fundamental frequency contours for **a)** a 71 years old healthy speaker with m-FDA = 0, and **b)** a 77 years old PD patient with m-FDA = 41 and MDS-UPDRS-III=92. F_0**SD:** standard deviation of F_0.

Different features are considered in this thesis to model the prosody impairments exhibited by PD patients. A complete list of the considered features is shown in Table 4.4. The total feature set is divided into three groups to model the pitch contour (30 features), the energy contour (48 features), and the duration and speed (25 features). The F_0 contour is computed using the robust algorithm for pitch tracking (RAPT) [Talk 95], which shown to be more accurate than other methods, like the ones based on autocorrelation function from Praat [Garc 14].

4.2.4 OpenSMILE Features

This feature set comprises a set of 6373 features extracted with the OpenSMILE toolkit [Eybe 15], which is highly used for research in recognition of paralinguistic aspects from speech since several years now. This feature set is classically used as a baseline in the annual *Computational paralinguistic challenge (ComParE)* since 2009. The extracted features comprise several acoustic measures based on spectral, cepstral, linear predictive coding, and perceptual linear prediction analysis, among others. It includes also articulation-based features based on formant frequencies, and prosody features based on pitch and loudness. The toolkit is available online to be used for the community [2]. The extracted features from OpenSMILE are used in

[2]OpenSMILE: `https://www.audeering.com/opensmile/`

Table 4.4: Description of prosody features. **Avg:** Average, **SD:** standard deviation, **Max:** maximum value, **Min:** minimum value.

Num.	Feature	Description
	Features based on F_0	
1-6	F_0 contour	Avg, SD, Max, Min, Skewness, Kurtosis
7-12	Tilt of a linear estimation of F_0 for voiced segments	Avg, SD, Max, Min, Skewness, Kurtosis
13-18	MSE of a linear estimation of F_0 for voiced segments	Avg, SD, Max, Min, Skewness, Kurtosis
19-24	F_0 on the first voiced segment	Avg, SD, Max, Min, Skewness, Kurtosis
25-30	F_0 on the last voiced segment	Avg, SD, Max, Min, Skewness, Kurtosis
	Features based on energy	
31-34	Energy-contour for voiced segments	Avg, SD, Skewness, Kurtosis
35-38	Tilt of a linear estimation of energy contour for voiced segments	Avg, SD, Skewness, Kurtosis
39-42	MSE of a linear estimation of energy contour for voiced segments	Avg, SD, Skewness, Kurtosis
43-48	Energy on the first voiced segment	Avg, SD, Max, Min, Skewness, Kurtosis
49-54	Energy on the last voiced segment	Avg, SD, Max, Min, Skewness, Kurtosis
55-58	Energy-contour for unvoiced segments	Avg, SD, Skewness, Kurtosis
59-62	Tilt of a linear estimation of energy contour for unvoiced segments	Avg, SD, Skewness, Kurtosis
63-66	MSE of a linear estimation of energy contour for unvoiced segments	Avg, SD, Skewness, Kurtosis
67-72	Energy on the first unvoiced segment	Avg, SD, Max, Min, Skewness, Kurtosis
73-78	Energy on the last unvoiced segment	Avg, SD, Max, Min, Skewness, Kurtosis
	Features based on duration	
79	Voiced rate	Number of voiced segments per second
80-85	Duration of Voiced	Avg, SD, Max, Min, Skewness, Kurtosis
86-91	Duration of Unvoiced	Avg, SD, Max, Min, Skewness, Kurtosis
92-97	Duration of Pauses	Avg, SD, Max, Min, Skewness, Kurtosis
98-103	Duration ratios	Pause/(Voiced+Unvoiced), Voiced/Pause, Pause/Unvoiced, Unvoiced/(Voiced+Unvoiced), Unvoiced/Pause , Voiced/(Voiced+Unvoiced),

this thesis as a baseline and complement to the classical features based on phonation, articulation, and prosody, described previously; and to the proposed features based on phonological analysis and representation learning, described in the following sections.

4.3 Phonological Analysis of Speech

Different groups of features such as MFCCs, PLPs, or embedding from neural networks are commonly used for several speech processing applications. However, for pathological speech processing such as PD classification or dysarthria modeling, only a small subset of features are used. The most important ones were those described in the previous sections to model phonation, articulation, or prosody impairments of the patients. More complex feature sets are rarely used by the medical community to model pathological speech, mainly because their lack of interpretability. Nevertheless, those high-dimensional feature vectors contain a great amount of information about the state of the patients that can be exploited by clinicians. Thus, it is important to provide clinically meaningful features that at the same time carry meaningful information about the health state of the patients.

Phonological features are more understandable for clinicians than the traditional high-dimensional features used in speech processing. Phonological features are represented by a vector with information about the mode and manner of articulation of the speaker, which are specifically related with the movements of the articulators in the vocal tract. Phonological features have been considered for different pathological speech processing applications [Midd 09], including assessment of dysarthric

speech [Jiao 17, Cern 17], the evaluation of progressive apraxia of speech [Asae 16], or the assessment of speech of cochlear implant users [Aria 19].

Different models have been proposed to extract phonological features from speech. In [Zhao 15] the authors detected phonological categories such as consonant, nasal, and bilabial, among others. The model was trained with utterances from isolated phonemes, syllables, and English words using a deep-belief network that achieves accuracies over 90%. In [Cern 16] the authors presented a toolkit called *Phonvoc*[3] to estimate 15 phonological posteriors based on the sound patterns of English. The authors considered a parallel bank of fully connected networks to recognize each phonological class. Accuracies over 96% were reported to detect the phonological classes. In [Jiao 17] the authors detected 15 phonological classes using a model based on RNNs with LSTM units trained with the TIMIT corpus. The phonological classes were detected with accuracies over 90%. In spite of the success on the use of phonological features to characterize pathological speech, there is a lack of models available for the research community that can be used and adapted for different pathological speech applications. The availability of models is even more scarce for languages different to English. That is the reason why we proposed a new model to extract phonological features from speech, in the context of PD evaluation. The model was designed for Spanish native speakers, and is available as a toolkit for the research community. The next subsections describe the *Phonet* toolkit to extract phonological posterior probabilities from speech, and how we consider those phonological posteriors to extract meaningful features to model the speech impairments of PD patients.

4.3.1 Phonet

Phonet is a toolkit designed to estimate phonological posteriors based on bidirectional RNNs with GRU units. The models are available online[4] to be used by the research community interested in pathological speech assessment. The toolkit was originally presented in [Vasq 19b]. However, by the time of writing this thesis, the architecture of the model has been updated from its original version to get more accurate phonological features. The model is trained with Spanish language utterances to test the reliability of the phonological analysis in a language different to English.

The phonetic alphabet for Spanish includes 24 different phonemes, represented by 5 vowels and 19 consonants [Hier 93]. These phonemes are grouped into phonological classes based on the mode and manner of articulation of the sounds. Tables 4.5 and 4.6 show the distribution of the Spanish phonemes into the phonological classes for vowels and consonants, respectively. The notation of the phonemes is based on the international phonetic alphabet.

Some conventions were considered in the design of Phonet to extract the phonological features based on Spanish language: (1) the phoneme /θ/ was not considered because it is only used in Spanish from central Spain. (2) The phoneme /ʝ/ from the word *cayado* and the phoneme /ʎ/ from the word *callado* were grouped together since they are pronounced similarly in many Latin American countries [Bagu 17]. (3) The

[3]Phonvoc: Phonetic and phonological vocoding platform `https://github.com/idiap/phonvoc`

[4]Phonet: Keras-based python framework to compute phonological posterior probabilities from audio files `https://github.com/jcvasquezc/phonet`

Table 4.5: Distribution of Spanish vowels into phonological classes. Source: [Vasq 19b].

	Front	**Central**	**Back**
High	/i/		/u/
Mid	/e/		/o/
Low		/a/	

Table 4.6: Distribution of the Spanish consonant into phonological classes based on the mode and manner of articulation. Source: [Vasq 19b].

	Labial		**Dental**		**Alveolar**	**Palatal**	**Velar**	
Nasal	/m/				/n/	/ɲ/		
Stop	/p/	/b/	/t/	/d/		/tʃ/	/k/	/g/
Fricative	/f/		/θ/		/s/	/ʝ/	/x/	
Lateral					/l/	/ʎ/		
Flap					/ɾ/			
Trill					/r/			

phoneme /n/ from the word *cana* and the phoneme /ɲ/ from the word *caña* were also grouped together because they belong to the same phonological categories. Based on the aforementioned conventions, we have a phonetic alphabet with 21 phonemes to train the phonological feature extractor in Phonet. Those phonemes are distributed into 18 phonological classes defined according to Table 4.7. The silence is considered as an additional phonological class.

Table 4.7: Distribution of the different Spanish phonemes into phonological classes.

#	**Phonological class**	**List of phonemes**
	Manner of articulation	
1	Nasal	/m/, /n/
2	Stop	/p/, /b/, /t/, /k/, /g/, /tʃ/, /d/
3	Continuant	/f/, /b/, /tʃ/, /d/, /s/, /g/, /ʎ/, /x/
4	Lateral	/l/
5	Flap	/ɾ/
6	Trill	/r/
7	Strident	/f/, /s/, /tʃ/
8	High	/i/, /u/
9	Low	/a/, /e/, /o/
	Place of articulation	
10	Dental	/t/, /d/
11	Labial	/m/, /p/, /b/, /f/
12	Velar	/k/, /g/, /x/
13	Back	/a/, /o/, /u/
14	Front	/e/, /i/
	Other categories	
15	Vocalic	/a/, /e/, /i/, /o/, /u/
16	Consonantal	/b/, /tʃ/, /d/, /f/, /g/, /x/, /k/, /l/, /ʎ/, /m/, /n/, /p/, /ɾ/, /r/, /s/, /t/
17	Voiced	/a/, /e/, /i/, /o/, /u/, /b/, /d/, /l/, /m/, /n/, /r/, /g/, /ʎ/
18	Silence	/sil/

The phonological posteriors will be the conditional posterior probability of a speech frame to belong to one or more phonological classes. The phonological posteriors are computed with a deep learning model based on RNNs with GRU units, trained to detect the occurrence of the phonological classes. The deep learning model is trained following a multitask learning strategy to detect the different phonological classes. The proposed model to extract the phonological posteriors is shown in Figure 4.9. Speech segments of 400 ms length are windowed into frames of 25 ms with a time-shift of 10 ms to compute the feature sequence for the input layer of the neural network. The input features correspond to the log-energy of the signal distributed into 33 triangular filters separated according to the Mel scale. The feature sequences from the input are processed by two bidirectional GRU layers with 128 cells to model information from the past (backward) and future (forward) states of the sequence, simultaneously. The output of the second recurrent layer is connected to 18 time distributed dense layers (one per phonological class) with 128 neurons and ReLU activations. Then, the output for each time distributed dense layer is connected to its respective output layer with a Softmax activation function to get posterior probabilities for the output sequence. The model is trained with a multitask learning strategy because it has shown to improve generalization in the training process of a deep learning model [Caru 94]. When part of the network is shared across different tasks, i.e., different phonological classes, the feature maps are more constrained, yielding better generalization.

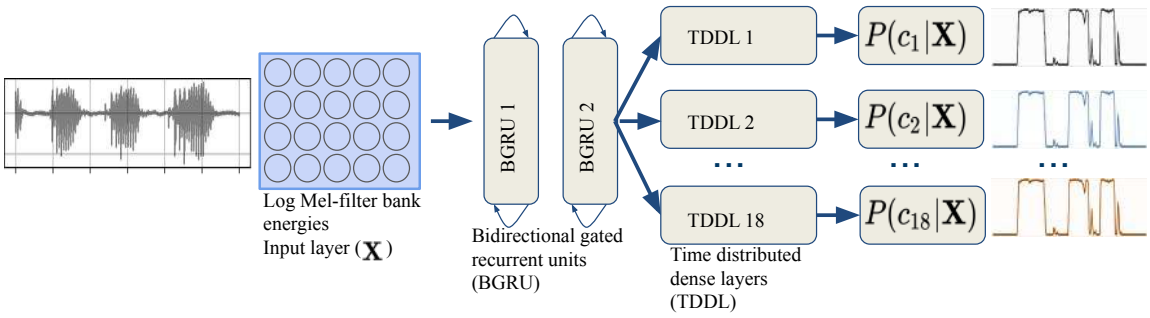

Figure 4.9: Architecture of the proposed neural network to estimate the phonological posteriors from the speech. $P(c_i|\mathbf{X})$: conditional probability for each phonological class $c_i, i : \{1, 2, \cdots, 18\}$.

The loss function in a multitask learning strategy is a linear combination of the individual loss functions for each task, following Equation 4.5 when two tasks are considered. The term γ is a weight hyper-parameter, $L_1(\Theta)$ is the loss for the first task, and $L_2(\Theta)$ is the loss function for the second task. When $\gamma = 0$, the network only learns the first task, and when $\gamma = 1$, the network is trained to predict only the second task. The loss function can be generalized using Equation 4.6 when more than two tasks are considered, subject to the condition $\sum_i \gamma_i = 1$. The values of γ_i were set to $1/18$ to give the same weight to each phonological class in the loss function. The loss function for each task $L_i(\Theta)$ corresponds to the weighted categorical cross-entropy, defined according to Equation 4.7 to avoid the unbalance of the classes in the training process. The weight factors w_i for each class are defined based on the percentage of samples from the training set that belong to each class. The network was trained

using an Adam optimizer [King 14]. In addition, dropout layers with probability of 0.2, and batch normalization were considered to improve the generalization of the proposed network. The model was trained with an early stopping strategy with a patience of 15 epochs.

$$L(\Theta) = \gamma L_1(\Theta) + (1-\gamma)L_2(\Theta) \tag{4.5}$$

$$L(\Theta) = \sum_i \gamma_i L_i(\Theta) \tag{4.6}$$

$$L_i(\Theta) = -w_i \log\left(p\left(c_i|\mathbf{X}\right)\right) + (1-w_i)\log\left(1-p\left(c_i|\mathbf{X}\right)\right) \tag{4.7}$$

The training of Phonet was performed with the CIEMPIESS corpus [Hern 14], which consists of 17 hours of FM podcasts in Mexican Spanish. The database was designed to be used in speech recognition systems, and it was annotated at word level, considering all phonemes of the Spanish language. The data contain 16717 audio files with a sampling frequency of 16 kHz and 16-bit resolution. 700 utterances from the entire corpus (from a set of different speakers) were separated to be used as the test set for the detection of phonological classes. The CIEMPIESS corpus was forced-aligned using the *BAS CLARIN* web service[5] [Kisl 17] based on the phonetic segmentation introduced in [Schi 99] for Spanish. The audio files and their corresponding transcriptions were uploaded to the server, which provides *Textgrid* files with phonetic alignment for each utterance. The aligned phonemes were used as labels to train Phonet. The accuracy of Phonet to recognize the different phonological classes are shown in Table 4.8. The proposed model shows to be highly accurate to detect the different phonological classes. The F-score ranges from 0.827 to 0.956, depending on the phonological class.

Table 4.8: Accuracy of Phonet to detect the different phonological classes.

	Phonological class	**F-score**	**Precision**	**Recall**		**Phonological class**	**F-score**	**Precision**	**Recall**
1	Nasal	0.892	0.944	0.869	10	Dental	0.894	0.947	0.867
2	Stop	0.862	0.902	0.847	11	Labial	0.846	0.931	0.806
3	Continuant	0.877	0.905	0.865	12	Velar	0.900	0.962	0.864
4	Lateral	0.873	0.960	0.822	13	Back	0.884	0.894	0.881
5	Flap	0.827	0.955	0.754	14	Front	0.889	0.913	0.882
6	Trill	0.945	0.995	0.903	15	Vocalic	0.853	0.854	0.853
7	Strident	0.933	0.954	0.924	16	Consonantal	0.837	0.842	0.836
8	High	0.886	0.927	0.868	17	Voiced	0.891	0.894	0.890
9	Low	0.843	0.855	0.842	18	Silence	0.956	0.965	0.952

Figures 4.10a) and 4.10b) show the difference between the phonological posteriors estimated for a PD patient (a) and an HC speaker (b) when they pronounce the Spanish sentence *mi casa tiene (my house has)*. Phonological posteriors for vocalic, stop, nasal, and strident are included. Note that the toolkit is accurate to estimate all phonological classes for the HC subject, but it is not able to accurately differentiate several classes for the PD patient, for instance nasal, stop, and vocalic posteriors

[5] https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic

are overlapped. Similar, Figures 4.10c) and 4.10d) show the difference in the labial, dental, vocalic, and velar phonological posteriors between a PD patient and an HC subject when they perform a DDK task like the repetition of /pa-ta-ka/. Note that for the case of the PD patient in Figure 4.10c) the dental posterior appears to be active when it should not, e.g., in the first and second /p/ phonemes. Same behavior is observed for velar, which is active for the /t/ phoneme between 0.9 and 1.0 seconds. For the case of the HC subject in Figure 4.10d) note that the posteriors are generally higher, more stable , and less overlapped than the observed for the PD patient.

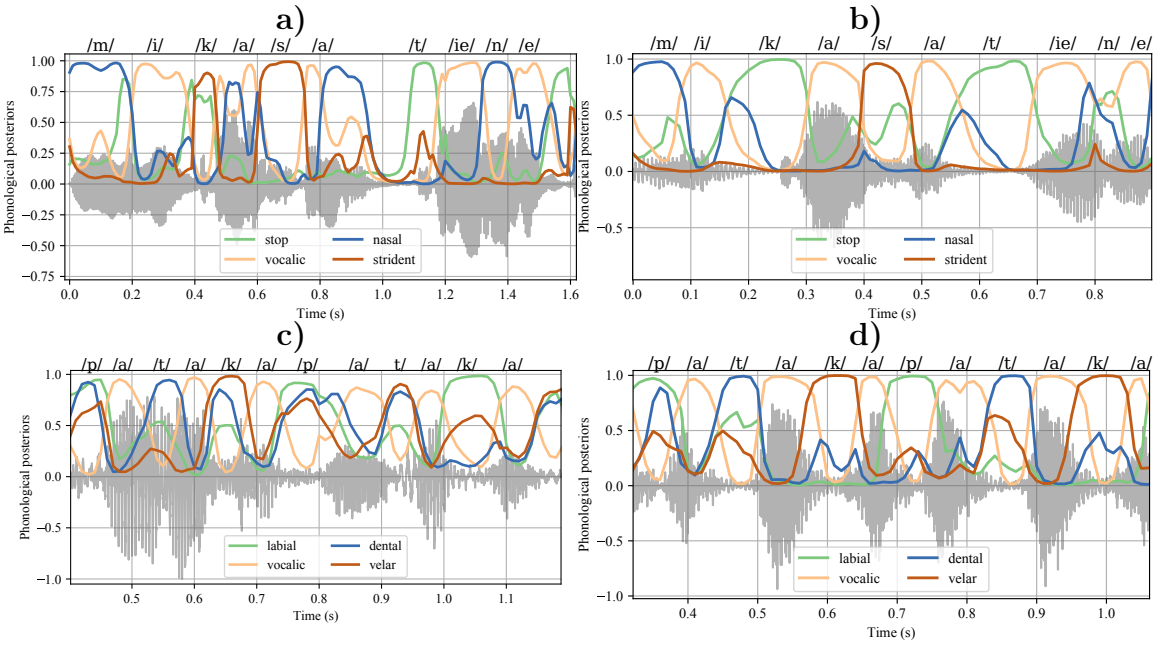

Figure 4.10: Phonological posterior probabilities for for **a)** a PD patient, pronouncing a sentence. **b)** an HC speaker, pronouncing a sentence. **c)** a PD patient performing a DDK task, and **d)** an HC subject performing a DDK task. The PD patient corresponds to a 77 years old subject with m-FDA = 41 and MDS-UPDRS-III=92. The HC subject is a 71 years old speaker with m-FDA = 0. The sentence produced by the subjects is the Spanish sentence /mi casa tiene/ (my house has), and the DDK task corresponds to the repetition of the syllables /pa-ta-ka/.

A different visualization of the difference in the phonological posteriors between PD patients and HC subjects is observed in Figure 4.11. The radar plot in Figure 4.11a) shows the phonological posteriors in the Spanish sentence *Mi casa tiene tres cuartos* (my house has three rooms), for a PD patient and an HC subject, who acts as a reference speaker. The posteriors are generally higher for the HC subject compared with the PD patient, especially for labial, stop, and nasal, which indicates a better pronunciation of those groups of phonemes for the case of the HC subject. Figure 4.11b) shows a similar analysis, but for a DDK task, where labial, dental, velar, vocalic, and stop classes appear. For this case note also that the posteriors are higher for the HC subject patient compared with the PD patient. Figure 4.11c) shows the difference in the phonological posteriors that are not active during the DDK task i.e., the ideal probability should be 0. For this case, the area of the radar is smaller for the HC subject than for the PD patient, which reflects problems in the

pronunciation of the phonemes present in the DDK task for the PD patient, e.g., the vowels appear to sound more low and frontal than back, for the case of vowel /a/, or the original stops appear to sound more continuant for the case of PD patient.

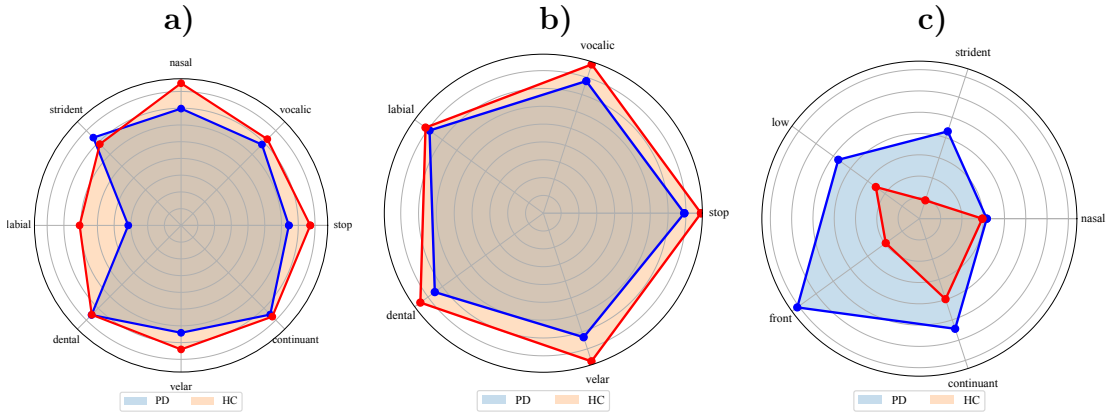

Figure 4.11: Difference in phonological posteriors between PD patients and HC subjects. **a)** Active phonological classes in the sentence *Mi casa tiene tres cuartos* (my house has three rooms). **b)** Active phonological classes during the repetition of /pa-ta-ka/. **c)** Not active phonological classes during the repetition of /pa-ta-ka/.

4.3.2 Phonological Features

Different features can be computed from the contour of the phonological posteriors. We propose the use of phonological log-likelihood ratio (PLLRs) features, as an effective set to characterize acoustic and phonetic aspects from PD patients. The features are designed to model the capabilities of the speakers to pronounce different groups of phonemes. These type of features have shown to be accurate to model different aspects of speech, such as spoken language [Diez 14a, Diez 14b], or non-nativeness [Abad 16] due to their ability to model the pronunciation of different groups of phonemes. PLLRs are computed from the set of N phonological posteriors p_i following Equation 4.8. $\sum_{i=1}^{N} p_i = 1$ and $p_i \in [0, 1]$.

$$r_i = \text{logit}(p_i) = \log \frac{p_i}{1 - p_i} \qquad i = 1, \cdots, N \tag{4.8}$$

As explained in [Diez 14b], PLLRs overcome the non-Gaussian nature of phonological posteriors, which is better to exploit different classification methods, or speaker models based on GMM-UBM or i-vectors. However, according to [Diez 14b], when the distribution of two (or more) PLLRs is observed, these Gaussian distributed features show to be a strongly bounded, which limits the distribution of the feature space. In order to avoid the bounding effect, and with the aim to obtain a smoother representation of the features, PLLRs are projected as described in [Diez 14b, Diez 14a] using the projection matrix P from Equation 4.9. $\widehat{\mathbf{1}} = \frac{1}{\sqrt{N}} [1_1, 1_2, \cdots, 1 - N]$ and $\mathbb{I}$ stands for the Identity matrix.

$$P = \mathbb{I} - \widehat{\mathbf{1}}^{\intercal} \widehat{\mathbf{1}} \tag{4.9}$$

The projected PLLRs are computed for the speech signals from the available corpora. Then six statistical functionals are computed: average, standard deviation, skewness, kurtosis, maximum, and minimum, in a similar way as in the phonation, articulation, and prosody features. The final phonological feature vector per utterance is formed with 18 PLLRs $\times$ 6 functionals= 108 features.

4.4 Unsupervised Representation Learning for Speech Analysis

As it was explained previously, traditional methods to model pathological speech are based on the computation of single hand-crafted features such as jitter, shimmer, or formant frequencies that may not completely model all of the phenomena that appear due to the presence of the disease and the dysarthria level of PD patients. Methods based on feature representation learning have the potential to extract more abstract and robust features than those manually computed. These features could help to improve the accuracy of different models to characterize pathological speech [Cumm 18]. There are recent studies focused on extracting features based on deep learning strategies to characterize speech signals for different applications. One of the most well known methods to characterize speech signals using deep learning corresponds to *X-vectors* [Snyd 18], which were designed for speaker recognition applications. X-vectors are extracted from time-delay neural networks to model a fixed temporal context size from the speech signal. *X-vectors* have even been considered to classify PD vs. HC subjects from speech [Moro 20]. Additional representation learning methods include the one presented in [Chor 19], where the authors aimed to learn a feature representation with information about the phonetic distribution of the utterance, and at the same time being invariant to the identity of the speaker or the acoustic conditions. The proposed model consisted of a vector quantized variational autoencoder based on Wavenet [Oord 16] to decode the raw waveform from the bottleneck space. The authors applied their proposed model for acoustic unit discovery and phoneme recognition problems. The model achieved an accuracy up to 64.5% recognizing 41 phonemes from the librispeech corpus. In [Pasc 19] the authors consider a self-supervised learning scheme to encode a feature representation from raw speech signals. The proposed method consisted of an end-to-end encoder-decoder neural network with multiple decoders to learn different speech features. The encoder processed the raw speech waveform with convolutional layers with Sincent filters [Rava 18] to get the feature embedding. The decoders aimed to learn different speech aspects such as the log-power spectrum, MFCC, and prosody features. The proposed model was evaluated in a speech recognition problem using the TIMIT corpus, achieving an accuracy up to 85.3%. The model from [Pasc 19] was recently updated in [Rava 20b], by including a quasi RNN and skip connections in the encoder, and more decoders to learn also Gammatone and Mel filterbanks. The model was applied in a phoneme recognition problem using the TIMIT corpus, and achieved a phoneme error rate of 32.7%. Although there are different strategies to learn feature representations from speech, to the best of our knowledge, non of them have been applied to model pathological speech utterances, with the exception of X-vectors. We

recently proposed a model based on unsupervised representation learning to extract suitable features for pathological speech classification considering convolutional and recurrent autoencoders [Vasq 20a]. The aim is not only to have a robust feature space in the bottleneck layer but also to obtain meaningful and interpretable features based on the reconstruction error of the autoencoders. The features proposed in [Vasq 20a] are applied and extended in this thesis to detect the presence of PD and to evaluate the dysarthria severity of the speakers. Only the features based on the recurrent autoencoder (RAE) are considered, since they showed to be more robust to model the presence of pathological speech [Vasq 20a].

4.4.1 Recurrent Autoencoders

The RAE is considered to characterize the temporal structures of the input spectrogram. The architecture of the implemented RAE is shown in Figure 4.12. The input is a spectrogram with 128 frequency bins distributed according to the Mel-scale and 126 time steps. The speech signal is segmented into chunks of 500 ms with a time-shift of 250 ms. The spectrogram is computed for each chunk with a window length of 32 ms and a step-size of 4 ms, forming the 126 time-steps. The STFT is computed with 512 frequency points, which are transformed into the Mel-scale using 128 filters, forming the input spectrogram observed in the left part of Figure 4.12. In these spectrograms we did not lose information about the fundamental frequency contour because the high number of Mel filters and the large frame size. Each column of the $n = 126$ time steps of the spectrogram serves as input for a sequence to one RNN in the encoder, which is formed with a bidirectional LSTM (BLSTM) with 128 cells to model information from the past (backward) and future (forward) states of the sequence, simultaneously. The output sequence of the BLSTM layer at the last time step $\mathbf{x}_n$ is stacked with the hidden state of the layer at the last time step $\mathbf{s}_n$ because they have observed and carry information about the whole input sequence. This stacked vector then passes trough a fully connected layer to get the bottleneck representation $\mathbf{h}$. The decoder is formed with a sequence of 2 LSTM layers to retrieve the original spectrogram from the bottleneck representation. The bottleneck features were replicated 126 times for the decoder because every LSTM cell in the decoder requires an input vector. The complete architecture of the RAE is shown in Figure 4.12. The Pytorch [Pasz 17] implementation of the trained models are available online[6] for the research community. The repository also contains scripts to train the autoencoders with different datasets, and methods to use the trained autoencoders to extract the proposed features.

The RAE was trained with the CIEMPIESS corpus [Hern 14], using the same strategy explained previously for the phonological features. The RAE was trained using the MSE loss function between the input and output spectrograms. We consider also an Adam optimizer [King 14].

[6] AEspeech: feature extraction from speech signals based on representation learning strategies using pre-trained autoencoders: `https://jcvasquezc.github.io/AEspeech/`

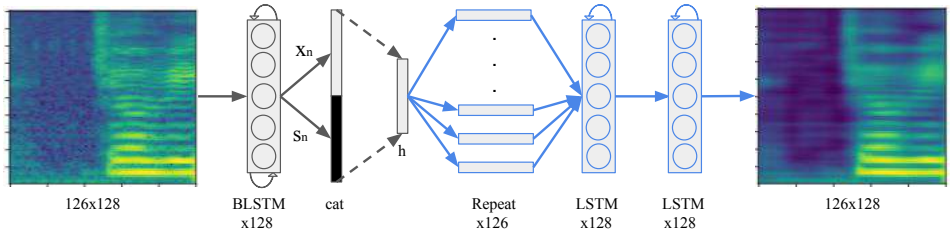

Figure 4.12: Scheme of the RAE. Source: [Vasq 20a]. **h**: bottleneck representation. $\mathbf{x}_n$ output of the BLSTM layer at the last time step, $\mathbf{s}_n$ hidden state of the BLSTM at the last time step.

4.4.2 Representation Learning Features

Two different feature sets are extracted and stacked from the trained RAE, according to Figure 4.13. The first set consists of the bottleneck features $\mathbf{h} \in \mathbb{R}^{128}$ obtained from the autoencoder, computed from the 500 ms length segments. The bottleneck features derived from Mel-scale spectrograms have shown to be more effective than the MFCC to characterize speech signals [Deng 10]. We consider 128 hidden units in the bottleneck space because it shown to be slightly more accurate to model the presence of PD, according to [Vasq 20a]. The second feature set is based on the MSE between the input and the decoded spectrograms, computed for each frequency band (MSE(f), $f \in \mathbb{R}^{128}$). We hypothesize that not all frequency regions of the spectrogram are reconstructed with the same error, and that such an error is related to the presence of paralinguistic aspects such as the presence of PD or the dysarthria level of the speakers. This approach is inspired from models designed for anomaly detection in time-series [Pere 18b, Fan 18], where for this case, the anomalies are referred to speech signals affected by the presence of PD, and which cannot be reconstructed properly by the RAE autoencoder. In [Vasq 20a] we showed that there are significant differences between the reconstruction error from the autoncoder between PD patients and HC subjects.

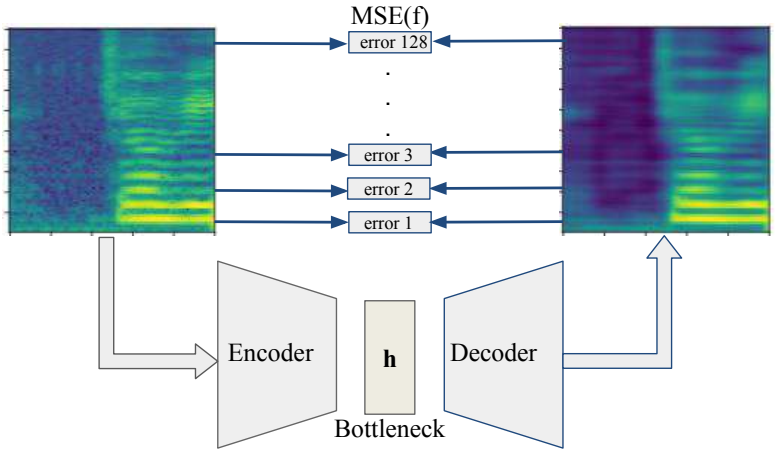

Figure 4.13: Features extracted from the autoencoders. Source: [Vasq 20a]

For each utterance, we obtain a feature matrix $X \in \mathbb{R}^{256\times N}$ formed with the concatenation of the bottleneck features and error features. N is the number of 500 ms length spectrograms extracted from the utterance. We then compute four statistical functions, similar to the other feature sets described previously (mean, standard deviation, skewness, and kurtosis), forming a vector $x_u \in \mathbb{R}^{1024}$ to represent the complete utterance.

4.5 Deep Learning Models for Speech Analysis

Besides the previous models, which are based on different feature extraction strategies for a latter classification, we also consider end-to-end deep learning models both to classify PD patients and HC subjects and to evaluate the dysarthria severity of the subjects. The proposed models are based on the CNNs described in Section 2.2.2 using time-frequency representations from the short-time Fourier transform. Two different spectral representations are considered for the input of the CNNs.

The first input comprises spectrograms of onset and offset transitions. This representation is an improved version of the proposed in [Vasq 17a, Vasq 19c]. The aim in this case is to model the difficulties of PD patients to start-stop the vocal fold vibration based on the transitions between voiced and unvoiced segments e.g., offset and between unvoiced to voiced segments e.g., onset. The offset and onset segments are detected according to the presence of the fundamental frequency using the RAPT algorithm, in a similar process than the addressed for the articulation features in Section 4.2.2 (see Figure 4.6). The borders are detected, and 80 ms of the signal are taken to the left and to the right of each border, forming chunks of signals with 160 ms length. Each one of those chunks is modeled using a Mel-spectrogram. The Mel-spectrograms of the transitions are computed with a frequency resolution of 512 points, 64 Mel filters, a window size of 32 ms and a time-shift of 4 ms. These parameters lead us to a time frequency representation of 41 time steps and 64 frequency bins, used as input for the CNNs. Figure 4.14 shows the difference in the onsets between a HC subject and three patients in different stages of the disease (mild, intermediate, and severe), according to their assigned m-FDA scores. Note that the HC speaker clearly defines the transition, conversely the patients are not able to produce clean transitions, especially for the patients in intermediate and severe levels of the disease.

The second input considered for the CNNs includes continuous speech segments. For this case, we consider similar Mel-spectrograms than the ones used to train the recurrent autoencoders from Section 4.4. For the second input, we consider Mel-spectrograms computed for speech segments with 500 ms length with a time-shift of 250 ms. The Mel-spectrograms of the speech segments are computed with a frequency resolution of 512 points, 64 Mel filters, a window size of 32 ms and a time-shift of 4 ms, similar to the previous case. These parameters lead us to a time frequency representation of 126 time steps and 64 frequency bins, used as input for the CNNs.

The two input spectrograms considered are processed by two different deep CNNs. The first model is based on a traditional CNN based on LeNet [LeCu 98]. It consists of three convolutional layers, with max-pooling and dropout, followed by a fully connected layer and the output layer. Leaky-ReLu activations are considered in the hidden layers, and a Softmax activation function is considered in the output to make

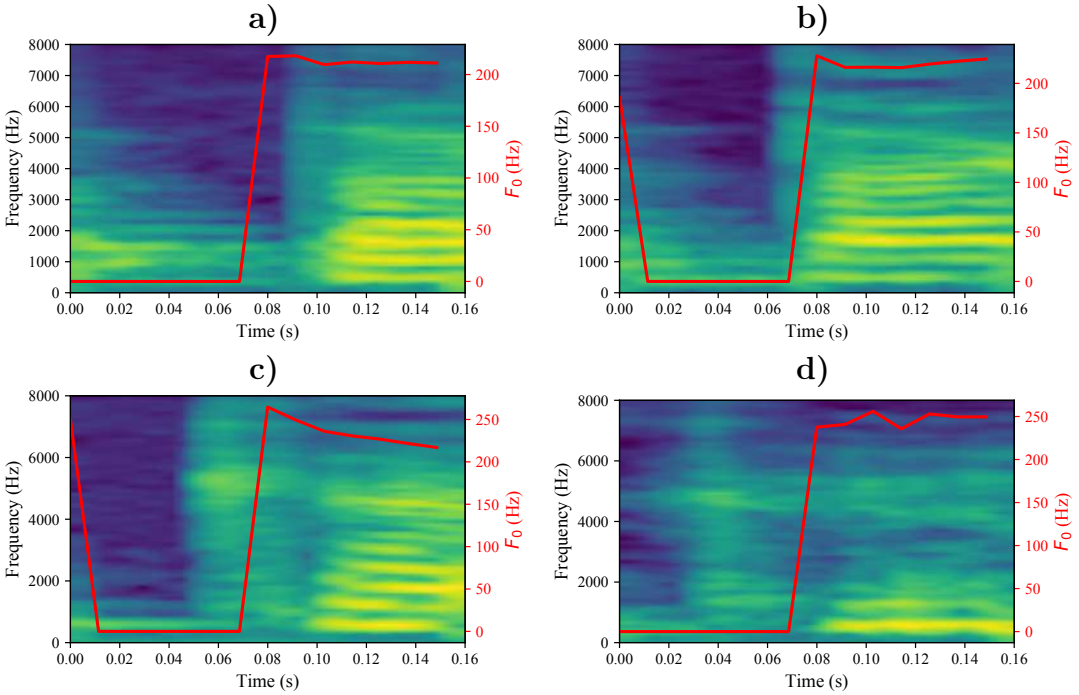

Figure 4.14: (**a)** Mel-spectrogram of an onset produced by a 75 years old female HC subject with m-FDA=3. **b)** Mel-spectrogram of an onset produced by a 73 years old female PD patient in with mild dysarthria severity state (m-FDA = 21). **c)** Mel-spectrogram of an onset produced by a 72 years old female PD patient with intermediate dysarthria severity (m-FDA = 31). **d)** Mel-spectrogram of an onset produced by a 75 years old female PD patient with severe dysarthria severity (m-FDA=47). All figures correspond to the syllable /ka/.

the final decision. The number of feature maps on each convolutional layer is twice the previous one in order to get more detailed representations of the input space in the deeper layers. The CNN is trained using the cross-entropy loss function, using an Adam optimizer. Figure 4.15 summarized this first architecture to process the input Mel-spectrograms.

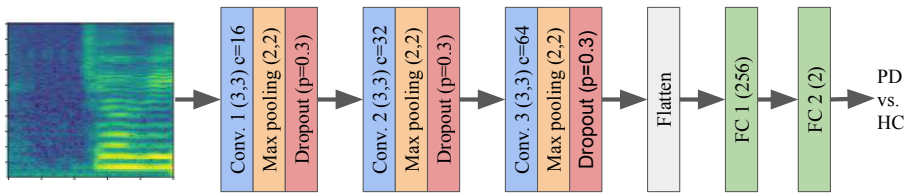

Figure 4.15: Architecture of the first CNN to process the Mel-spectrograms of the speech signals. **FC**: Fully connected layers. **c**= number of output channels in the convolutional layers. The values in parenthesis indicate the size of the convolutional filters and the number of neurons in the fully connected layers.

The second architecture is based on the ResNet models [He 16]. We consider a ResNet18 model, which has three residual blocks and 18 convolutional layers (see Figure 4.16). The skip connections helps to control the vanishing gradient problem

when we have deeper models, as it was explained in Section 2.2.2. Dropout layers were also considered to regularize the output of the residual blocks. The final decision is made by a fully connected layer with a Softmax activation function, similar to the first architecture.

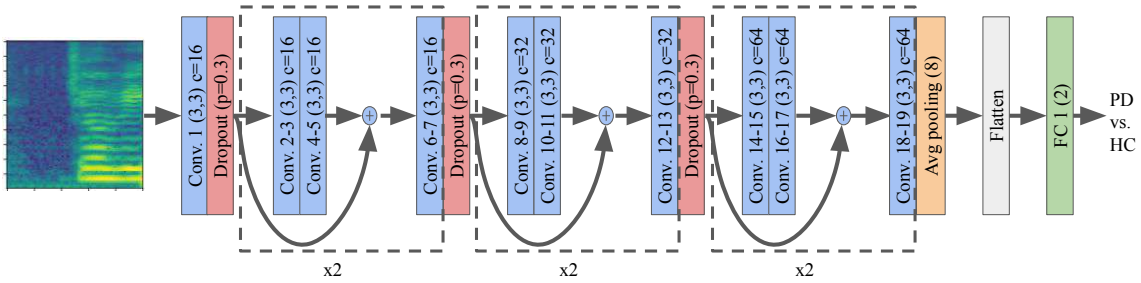

Figure 4.16: Architecture of the second CNN based on ResNet to process the Mel-spectrograms of the speech signals. **FC**: Fully connected layers. **c**= number of output channels in the convolutional layers. The values in parenthesis indicate the size of the convolutional filters and the number of neurons in the fully connected layers.

Chapter 5

Analysis of Parkinson's Disease from Handwriting

The main symptoms of PD in handwriting include micrographia, bradykinesia, and tremor [Smit 14]. On the one hand, micrographia is related to the reduction of the size in handwriting. The prevalence of micrographia in PD remains unclear. However, it has been observed that in general it affects less than 50% of the patients [Leta 14], which makes it not the most successful biomarker to evaluate the presence or severity of the disease. On the other hand, bradykinesia in PD patients causes longer times to complete a handwriting task than as usually required. Finally, tremor is related to involuntary movements, which produce irregular shapes in the drawings of the patients. The complete handwriting impairments in PD patients have been grouped and called PD dysgraphia, which is related to difficulties performing the controlled fine motor movements required to write [Leta 14]. The term dysgraphia implies that motor impairments related to the disease (e.g., tremor, rigidity, bradykinesia, akinesia, freezing of the upper limb) may affect handwriting kinematics without necessarily affecting writing size. In addition, it implies that dynamic and kinematic features are appropriate to diagnose and to monitor the disease severity of patients, or to evaluate the efficacy of a given treatment [Leta 14].

Handwriting assessment of PD can be divided according to the acquisition of the handwriting samples into two categories: online or offline. Online handwriting is captured from digitizer tablets, and contains information related to the dynamics of the handwriting process, the pressure of the pen, and the azimuth (orientation) and altitude (inclination) angles. Some tablets capture also information from the in-air movement, before the patient places the pen onto the tablet's surface. Conversely, offline handwriting can be captured also from tablets, and paper. Offline handwriting includes only spatial attributes from the drawings.

This chapter describes state-of-the-art methods and different proposed approaches to model handwriting in PD patients, both from a classical pattern recognition perspective and novel deep learning strategies. Section 5.1 shows a review of the literature about automatic handwriting evaluation of PD patients both to classify PD patients and HC subjects and to evaluate the neurological state of the patients. Then, Section 5.2 describes the classical kinematic features used in the literature to model handwriting in PD patients. Section 5.3 describes the geometric features proposed

initially in [Rios 19], and which are extended in this thesis to model Archimedean spiral shapes drawn by the patients. In Section 5.4 a feature set based on in-air analysis of handwriting is proposed to model the difficulties of patients to start or to stop new strokes. The chapter finishes in Section 5.5, where I describe methods to model the handwriting of PD patients in an end-to-end fashion using different configurations of CNNs, which process the reconstructed images from the online handwriting data.

5.1 A Review on Automatic Assessment of Handwriting in PD Patients

There is interest in the research community to automatically assess the handwriting process of PD patients. The handwriting assessment is commonly performed to discriminate between PD patients and HC subjects or to evaluate the neurological state of the patients. The overview presented here intends to provide the reader with a comprehensive and well structured view of methodological approaches and analyses regarding handwriting analysis of PD patients from a pattern recognition perspective. The survey is divided into two aspects: studies to classify PD vs. HC using handwriting, and those focused to evaluate the neurological state of the patients. Classical pattern recognition approaches and novel studies based on deep learning strategies will be covered.

5.1.1 Automatic Classification of PD and HC Subjects

Several studies have considered handwriting signals to classify PD patients and HC subjects. Most of these studies extracted kinematic features based on the dynamics of the velocity, acceleration, and jerk of the strokes [Rose 13, Drot 14, Drot 16, Kots 17, Much 18a, Jerk 18, Zham 18, Rios 19]. These kinematic features have been commonly combined with other ones based on the pressure of the pen, azimuth or altitude angles [Rose 13, Drot 16, Zham 18, Rios 19]. Recent studies have shown that the in-air movement is also important to characterize the handwriting impairments of PD patients [Rose 13, Drot 16, Jerk 18, Rios 19]. In [Rose 13] the authors computed kinematic features from 20 HC subjects and 20 PD patients, who were requested to write an address and their full name. The extracted features included the average on-surface and in-air times, the speed of the trajectory, and the average pressure of the pen. The authors used an LDA classifier, and reported accuracies close to 97%. In [Drot 14] the authors introduced the use of in-air movements in the handwriting of 37 PD patients and 38 HC subjects, who wrote a sentence in Czech language. The classification was performed with an SVM classifier. The results showed that the combination of features based on pressure, in-air, and on-surface times produces the highest accuracy (86%). This study was extended in [Drot 16], where the authors extracted kinematic and pressure-based features, and considered several classifier methods, reporting accuracies over 76.5%. In [Much 18a] the authors introduced a new strategy based on fractional derivatives to extract kinematic features from the handwriting of PD patients. Archimedean spirals from the PahaW database [Drot 16] were considered. The extracted features were classified using SVM and RF classifiers.

The authors reported an accuracy of up to 72.4% discriminating between PD and HC subjects. In [Jerk 18] the authors considered kinematic features from the in-air and on-surface trajectories to classify 10 HC subjects, 14 PD patients, 8 patients with progressive supranuclear palsy, and 11 patients with multiple system atrophy. The features from the four populations were classified with an LDA classifier. The authors reported an accuracy of up to 86% discriminating between the four groups, combining the in-air and the on-surface features. Similar kinematic, pressure, and angular-based features were considered in [Zham 18] to classify 31 PD patients and 31 HC subjects. The patients performed several tasks including the drawing of Archimedean spirals, the writing of an English sentence, and the repetition of the graphs *b*, *d*, and *bd*. The features were classified with a naive Bayes algorithm. The authors reported an AUC up to 0.933. New kinematic features were introduced in [Impe 19a] to characterize the velocity contour of the handwriting of PD patients. The features were extracted from different tasks from the PaHaW database [Drot 16]. The proposed features were based on the *Sigma-lognormal* model, which considers the state of the neuromuscular system to perform the handwriting process [ORei 09]. The extracted features were classified with a linear SVM. The authors reported an accuracy of up to 98.4% combining the features obtained for different tasks using an early fusion strategy.

Other studies have shown the importance of other groups of features based on geometric, spectral, and NLD analyses to characterize the handwriting impairments of PD patients. For instance, in [Sarb 13] the authors proposed features related to the power spectral density of the speed stroke to classify online handwriting from 17 HC subjects and 13 PD patients. The participants wrote a sentence in Persian language. The features were classified with a neural network, which achieved an accuracy of 86.2%. A different strategy was considered in [Pere 16b]. The authors classified offline drawings from 18 HC subjects and 74 PD patients, from the *handPD* database, by extracting geometric features based on the difference between the strokes drawn by the participants and the template that they have to follow. The classification was performed with different algorithms including SVMs and naive Bayes classifiers. The authors reported accuracies over 66% with the proposed approach. The authors in [Kots 17] classified handwriting samples from 24 PD patients and 20 HC subjects, who were instructed to draw several horizontal lines. The data were characterized with kinematic features based on the velocity of the trajectory, and NLD features based on the entropy on the horizontal and vertical movements. The features were classified with a naive Bayes algorithm, which achieved accuracies over 90.9%. Kinematic, geometric, spectral, and NLD features were considered in [Rios 19] to classify Archimedean spirals and sentences written by 39 PD patients, 31 HC subjects, and a set of 40 young HC subjects, The aim was to evaluate as well the influence of age in the classification problem. Accuracies over 94% were reported discriminating PD patients and young HC subjects, and over 89% classifying the patients and the HC subjects with similar age. In [Sena 19] the authors proposed the use of a classification strategy based on evolutive algorithms such as Cartesian genetic programming to classify the 18 HC subjects and 74 PD patients of the HandPD dataset [Pere 16b]. The authors considered the same geometric features extracted in [Pere 16b] based on the error between the spiral drawn by the participants and a template. Accuracies of up to 76.6% were reported. The same data and features used in [Sena 19] where

considered in [Ali 19a] to classify PD and HC subjects. The authors proposed a classification strategy based on AdaBoost to combine the output of several classifiers. The proposed strategy achieved an accuracy of up to 78.1% In [Cast 19] the authors considered a combination of kinematic, NLD and neuromotor features to classify data from 55 PD patients, 49 elderly HC subjects, and 45 young HC subjects. The participants performed a total of 17 exercises, divided into writing and drawing tasks. Kinematic features included the velocity, acceleration, and duration of the strokes. NLD features considered the CD, the LZC, the LLE, the HE, EMD decomposition, and entropy measures. Neuromotor features were based on the *Sigma-lognormal* model. The features were classified using different classifiers. The authors reported accuracies of up to 96.9% classifying PD patients vs. the young HC participants, and of up to 81.7% classifying the patients and the HC subjects with similar age.

Recent deep learning approaches have also been used to classify the handwriting of PD and HC subjects. In [Pere 16a] the authors classified offline Archimedean spiral drawings from 14 PD patients and 21 HC subjects using a CNN. The authors reported accuracies of up to 89.6% when the hyper-parameters of the CNN were optimized with a meta-heuristic optimization technique. In [Pere 18a] the same authors modeled the handwriting dynamics of 18 HC and 74 PD subjects from the newHandPD database. The authors proposed a model based on CNNs to classify the PD patients and the HC subjects. The signals from the smart pen were transformed into images by concatenating and reshaping the time series and the sensors into a square image. Several CNN configurations were considered and trained for each exercise performed by the patients. Afterwards, a majority voting scheme was implemented to make the final decision. Accuracies of up to 95% were reported. In [Gall 18] the authors considered a model called deep echo state network to classify 67 PD patients and 15 HC subjects who drew Archimedean spirals on a digitizer tablet. The deep learning model is based on RNNs, which process the time series of horizontal and vertical movements, the grip angle, and the pressure of the pen when the patients draw the spirals. The authors reported accuracies of up to 88.3% with the proposed model. A combination of NLD analysis and CNNs was proposed in [Afon 19] to classify handwriting samples from 21 HC subjects and 14 PD patients from the HandPD dataset [Pere 16b]. The authors considered a spatial representation based on recurrent plots to visualize the temporal dynamics of the handwriting samples. The images obtained from the recurrent plots were characterized and classified with a CNN. The authors reported accuracies over 87%. The same subjects of the HandPD dataset [Pere 16b] were classified in [Ribe 19] using a model based on bidirectional GRUs with an attention mechanism to process the raw data captured with the smart-pen. The authors reported an accuracy of up to 92.2%. A combined analysis of NLD analysis and deep learning was also proposed in [Cant 20]. The authors computed fuzzy recurrent plots to convert time series from online handwriting samples into gray scale texture images. The fuzzy recurrent plots were used to train a CNN based on AlexNet pre-trained with the Imagenet corpus. The proposed model was tested with data from Archimedean spirals collected from 25 PD patients and 15 HC subject. The authors reported an accuracy of 94%. In [Diaz 19] the authors transformed online drawings of the PahaW database [Drot 16] into offline images. The obtained images are passed trough a pre-trained version of a CNN based on VGG16 to extract features from the drawings. The extracted features

were classified with an AdaBoost algorithm. The authors reported an accuracy of up to 86.7% In [Moet 19] the authors modeled offline images of the handwriting of 37 PD patients and 38 HC from the PahaW database [Drot 16]. The authors proposed the use of parallel CNNs to extract different features from the hand drawn shapes of the patients. The CNN was based on a pre-trained version of AlexNet [Kriz 12] trained with the ImageNet dataset. The features obtained from the CNNs are combined and classified with an SVM. The highest accuracy reported by the authors is 83%. A pre-trained version of AlexNet was also considered in [Nase 20] to classify the handwriting samples from the PahaW database [Drot 16]. The authors applied a transfer learning strategy on CNNs initially trained with ImageNet and MNIST databases. The authors reported an accuracy of up to 98.2% obtained with images from the Archimedean spirals. However, the reported results seemed to be optimistic and biased, since some hyperparameters of the networks were optimized based on the accuracy obtained in the test set. An additional approach based on CNNs was proposed in [Gil 19] to classify 62 PD patients and 15 HC subjects using the Spiral drawings from the data from [Isen 14]. The input to the CNN was the FFT obtained from the X, Y, pressure, and grip angle, forming a 4-channel at the input. The authors reported an accuracy of up to 96.5% with their proposed strategy.

5.1.2 Automatic Evaluation of the Neurological State of Patients

Some studies focused not only on the classification of PD patients and HC subjects, but also on the assessment of the neurological state of the patients. For instance, in [Agha 17] the authors performed a handwriting assessment using a smartphone application where the patients draw a spiral. The authors computed several features including the kurtosis of the speed stroke, the length of the spiral drawing curve, the area of the spiral in each loop and the time of the drawing. The authors evaluated different items of the UPDRS score related to the upper limbs, and reported correlations ranging from 0.47 to 0.52 combining handwriting features with finger-tapping measures. The assessment of the neurological state of the patients based on the H&Y score was addressed in [Much 18b]. The authors used a regression algorithm and the kinematic features based on fractional derivatives proposed in [Much 18a]. The experiments also considered the classification of the 33 PD patients and 36 HC subjects from the PaHaW database [Drot 16]. The classification and regression algorithms were based on gradient-boosting trees. The highest accuracy for the classification problem was obtained with classical kinematic features (97.1%) rather than with those based on fractional derivatives. For the prediction of the neurological state, the authors reported an equal error rate of 12.5% (MAE=0.6). The assessment of the neurological state of the patients performed in [Rios 19] was focused on the classification of PD patients in several stages of the disease according to the MDS-UPDRS-III score. The patients were divided according to their MDS-UPDRS score into three groups (initial, intermediate, and severe). The highest accuracy was obtained with the combination of kinematic and geometric features (F-score=0.64).

As the UPDRS-III score involves a complete evaluation of the motor symptoms of the patients, it might be less suitable to assess only the impairments in the upper

limbs of the patients. There are a few studies that consider only the assessment of the motor symptoms of the patients in the upper limbs. In [Smit 17b] the authors evaluate the dysfunction on the upper limbs of 14 PD patients according to the Purdue pegboard test [Desr 95], which aims to evaluate manual dexterity in rehabilitation processes. The patients performed seven exercises using a smart pen, including hand resting, drawing a spiral, a circle, a zig-zag figure, the repetition of the graph *le*, and a modified version of the Fitt's task [Fitt 54]. The authors extracted kinematic features based on the drawing time and the writing size, features based on the Fitt's law [Fitt 54] to measure the trade-off between the speed and the accuracy to perform the modified Fitt's task, and spectral features extracted from a gyroscope attached to the smart pen to measure resting tremor. The features were correlated with the score from the Purdue pegboard test and with the UPDRS-III subscore for bradykinesia. A Spearman's correlation of up to 0.65 was reported between the drawing time features and the Purdue pegboard test score. Non-significant correlations were reported between the features and the UPDRS-III subscore for upper limbs.

The effect of dopaminergic medication in the handwriting of the patients was addressed in [Zham 19]. The authors evaluated whether there are significant differences in kinematic features computed from 24 PD patients in ON vs. OFF states. The patients performed several handwriting tasks such as the drawing of the Archimedean spiral, the repetition of simple graphs, the writing of a sentence, and a fluency test by writing names of animals. The results showed that there were significant differences between the features computed for the patients in ON vs. OFF states, especially for the features computed from simple tasks like the repetition of graphs. Differences were not observed in complex tasks like sentence writing or fluency, which carry memory and cognitive loads. In [Dann 19] the authors considered kinematic features extracted from Archimedean spirals of 20 PD patients and 20 HC subjects. The patients were examined in ON and OFF states to evaluate the effect of medication in the handwriting process. Different constraints were included for the patients to draw the spirals, e.g., spontaneous, as fast as possible, small, big, among others. An ANOVA test was conducted to compare the features from HC subjects, patients in OFF state, and patients in ON state. The features were able to discriminate between the three groups (p-val$<$0.05). The results indicated that the number of velocity peaks and the variation of the altitude angle were the most relevant features to separate between HC subjects and patients; and between PD patients in ON and OFF states.

5.1.3 Main Outcomes from the Literature

Although several studies have been performed to assess the handwriting process of PD patients, there are some open issues to be addressed in future studies. First, there is an absence of a proper and well-designed database that can be used to compare the different proposed approaches [De S 19]. A proper database should include most of the important tasks addressed in the literature, such as the Archimedean spiral (with and without reference templates), the repetition of the *l* and *le* patterns, and continuous writing exercises, among others. The use of the PahaW database [Drot 14, Drot 16] or the data from the multimodal corpus described in Section 3.3.1 shown to be the most complete to evaluate different handwriting impairments of PD patients.

According to the literature, the drawing tasks such as the Archimedean spirals are more suitable to assess the handwriting impairments of the patients. Writing tasks like sentences can be highly influenced by the education level of the patients [Saun 08]. Other important tasks include the repetition of graphs like *e*, *l*, and their combinations [Drot 16, Impe 19a]. The repetition of these graphs includes both up- and down-velocity strokes, and involves the writing of the same character scaled in amplitude. These aspects are observed in more detail in Figure 5.1 and Table 5.1, which summarize the most common handwriting tasks performed by PD patients in the literature. The described handwriting exercises have shown to be important to assess the handwriting deficits of the patients [Impe 19b]. More complex tasks such as the sentence writing or the Rey Osterrieth figure may be suitable to evaluate other aspects in the handwriting process, since they carry additional information related to memory and cognitive load [Impe 19b, Zham 19].

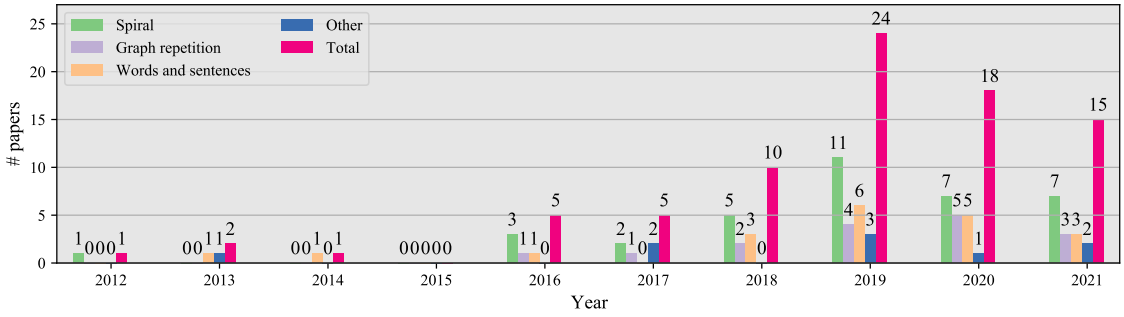

Figure 5.1: Different exercises considered in the literature for handwriting assessment of PD.

Table 5.1: Different exercises considered in the literature for handwriting assessment of PD.

References	Handwriting task
[Bart 12a], [Drot 16], [Pere 16a], [Pere 16b], [Agha 17], [Smit 17b], [Gall 18], [Much 18a], [Much 18b], [Pere 18a], [Zham 18], [Afon 19], [Ali 19b], [Cast 19], [Gil 19], [Impe 19b], [Moet 19], [Nase 20], [Ribe 19], [Rios 19], [Sena 19], [Vasq 19c], [Zham 19], [Cant 20], [Oroz 20b], [Arra 20], [Nomm 20], [Gupt 20], [Moet 20], [Gazd 21], [Cant 21b], [Lamb 21], [Diaz 21], [Parz 21], [Nola 21], [Impe 21]	Archimedean Spirals
[Drot 16], [Smit 17b], [Much 18a], [Zham 18], [Cast 19], [Impe 19b], [Moet 19], [Nase 20], [Zham 19], [Oroz 20b], [Tale 20], [Gupt 20], [Moet 20], [Gazd 21], [Diaz 21], [Nola 21]	Graph repetition
[Sarb 13], [Drot 14], [Drot 16], [Jerk 18], [Much 18a], [Zham 18], [Cast 19], [Impe 19b], [Moet 19], [Nase 20], [Rios 19], [Vasq 19c], [Zham 19], [Oroz 20b], [Tale 20], [Gupt 20], [Moet 20], [Gazd 21], [Diaz 21], [Ammo 21], [Nola 21], [Impe 21]	Words and sentences
[Rose 13], [Kots 17], [Smit 17b], [Cast 19], [Vasq 19c], [Zham 19], [Oroz 20b], [Alis 21], [Dent 21]	Other

Figure 5.2 and Table 5.2 summarize the most important methods considered for the handwriting analysis of PD patients. An important aspect to consider is that most of the studies consider only kinematic and pressure features, which only model some of the handwriting impairments of the patients. There are additional studies to

model the handwriting impairments of PD patients using spectral or NLD features. In addition, the use of deep learning methods has increased within the last years. Additional features should be proposed to assess other aspects in the handwriting process of the patients such as fluency, micrographia, or tremor.

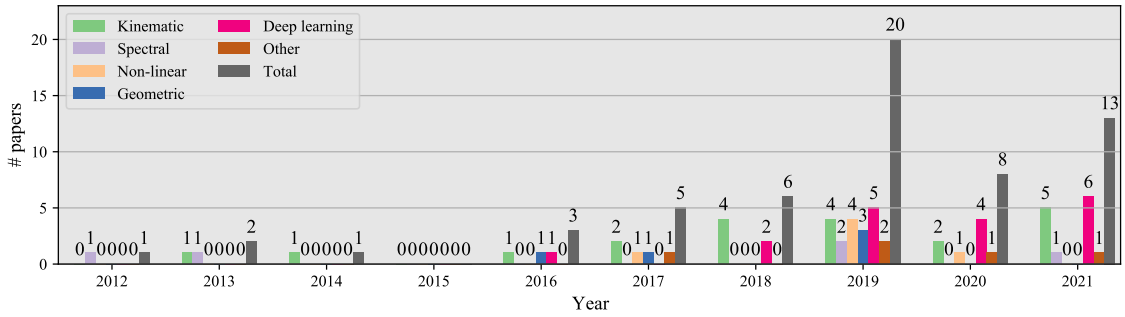

Figure 5.2: Different methods considered in the literature for handwriting assessment of PD.

Table 5.2: Different methods considered in the literature for handwriting assessment of PD.

References	Method
[Rose 13], [Drot 14], [Drot 16], [Kots 17], [Smit 17b], [Jerk 18], [Much 18a], [Much 18b], [Zham 18], [Cast 19], [Impe 19b], [Rios 19], [Zham 19], [Oroz 20b], [Gupt 20], [Lamb 21], [Diaz 21], [Ammo 21], [Parz 21], [Nola 21]	Kinematic
[Bart 12a], [Sarb 13], [Impe 19b], [Rios 19], [Nola 21]	Spectral
[Kots 17], [Ali 19b], [Afon 19], [Impe 19b], [Rios 19], [Cant 20]	NLD
[Pere 16a], [Agha 17], [Ali 19b], [Rios 19], [Sena 19]	Geometric
[Pere 16a], [Gall 18], [Pere 18a], [Afon 19], [Gil 19], [Moet 19], [Nase 20], [Ribe 19], [Vasq 19c], [Cant 20], [Tale 20], [Nomm 20], [Gazd 21], [Alis 21], [Cant 21b], [Diaz 21], [Dent 21], [Impe 21]	Deep learning
[Smit 17b], [Cast 19], [Impe 19b], [Arra 20], [Ammo 21]	Other

Regarding the applications addressed in the literature, most of the studies have focused on the classification of PD patients vs. HC subjects. There are few studies focused on evaluating other aspects of the patients such as the evaluation of the neurological state, the assessment of upper-limbs impairments of the patients, and the evaluation of the effect of dopaminergic medication in the handwriting process, among others. There is also need of longitudinal studies with the aim to understand and track the progress of the disease in the PD patients through time.

Finally, most of the studies consider only data from online handwriting, where the patients perform the exercises in a digitizer tablet. The handwriting process in a tablet may produce an uncomfortable feeling for the patients, and may introduce some bias in the results due to external factors such as education level or less contact with technology. Additional studies that compare online and offline handwriting should be proposed.

5.2 Kinematic Analysis of Handwriting

There are several kinematic features that can be extracted from online handwriting data. These features are based on the trajectory, velocity, and acceleration of the strokes, both in the horizontal, vertical, radial, and angular axes. These features also include measures based on the pressure of the pen and their derivatives, and those based on the azimuth and altitude angles of the pen (see Figure 3.4). The literature has also shown the importance of features based on in-air movement, i.e., before the participant places the pen on the tablet's surface [Drot 16].

The feature set used in this thesis to model the kinematic of the handwriting process of PD patients consists of 80 features, computed from 8 signals obtained from the tablet, according to Table 5.3. r and θ are the radial and angular trajectories, computed according to Equations 5.1 and 5.2, respectively.

$$r = \sqrt{x^2 + y^2} \tag{5.1}$$

$$\theta = \arctan\left(\frac{y}{x}\right) \tag{5.2}$$

A set of 10 functionals are computed from each signal from Table 5.3, including standard deviation, skewness, and kurtosis of the signal, its velocity, and its acceleration, and the average velocity of the signal. The complete description is observed in Table 5.3.

Table 5.3: Description of kinematic features for handwriting analysis. **Avg:** Average, **SD:** standard deviation. v: velocity, a: acceleration.

#	Signal	Description
1-10	x	SD., skewness, kurtosis, avg. v, SD v, skewness v, kurtosis v, SD a, skewness a, kurtosis a
11-20	y	
21-30	in-air	
31-40	r	
41-50	θ	
51-60	pressure	
61-70	azimuth	
71-80	altitude	

The velocity v is computed as the first derivative of the signal s, using second order accurate central differences[1], according to Equation 5.3. t_s is the sampling period of the signal. In a similar way, the acceleration is computed as the first derivative of the velocity. In addition, the velocity and acceleration are filtered with a moving average filter of 11th order to get a smoother representation of the signals and get more reliable kinematic features.

$$v_{(t)} = \frac{s_{(t+1)} - s_{(t-1)}}{t_s} \tag{5.3}$$

[1] https://numpy.org/devdocs/reference/generated/numpy.gradient.html

Figure 5.3 shows an example of the difference in the kinematic of handwriting between a PD patient and a HC subject. Figures 5.3a) and 5.3b) show an Archimedean spiral drawn by an HC subject and a PD patient, respectively. Note the tremor exhibited in the spiral drawn by the patient. The signals extracted from the pen when the subjects draw the spirals are shown in Figures 5.3c) and 5.3d) for the HC subject and the PD patient, respectively. For this case, note the difference in the stability of the pressure of the pen, the difference in the trajectory r followed by the patient and by the HC subject. Note also that the number of pen-up and pen-down movements (in-air) performed by the PD patient is larger than the number exhibited by the HC. Particularly, the analysis of those pen-up pen-down transitions is explored with more details in Section 5.4.

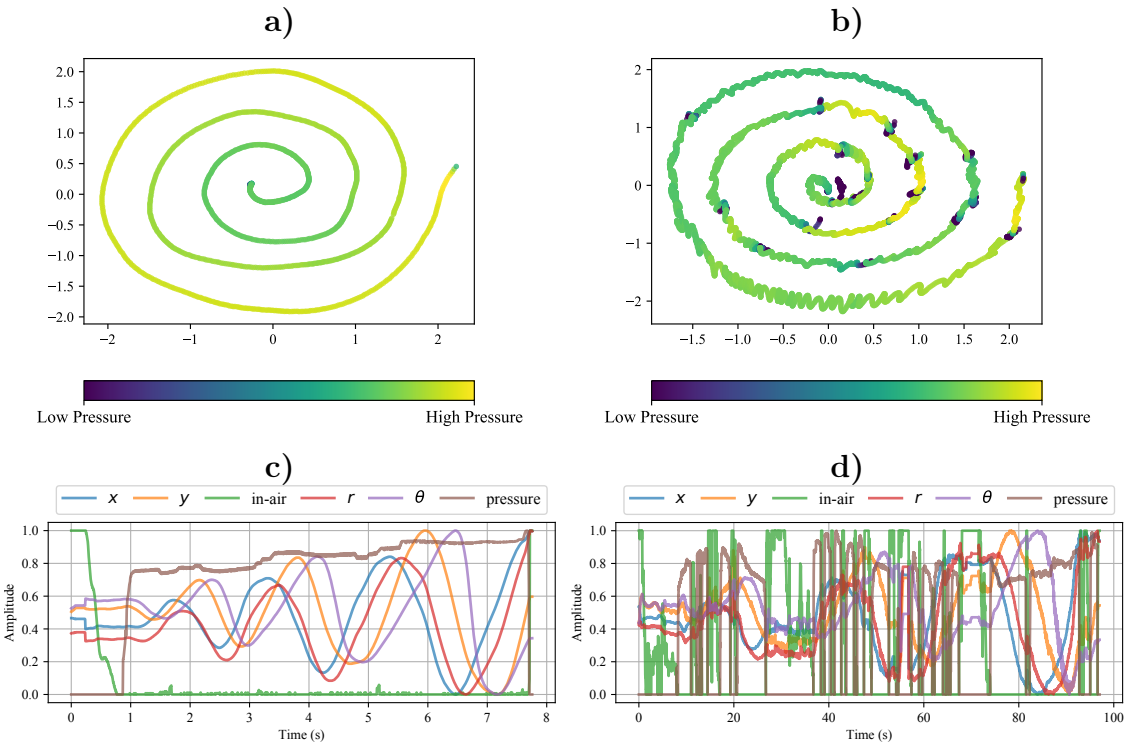

Figure 5.3: Difference in the handwriting kinematic between PD patients and HC subjects. **a)** Archimedian spiral drawn by a 71 years old HC subject. **b)** Archimedian spiral drawn by a 73 years old PD patient with MDS-UPDRS-III = 65. **c)** Signals extracted from the pen while the HC subject was drawing the spiral. **d)** Signals extracted from the pen while the patient was drawing the spiral. Dark blue indicates the pen in in the air.

5.3 Geometric Analysis of Handwriting

This feature set is inspired by [Rios 19] to model geometric and symmetry aspects in the Archimedean spirals drawn by the patients. In [Rios 19] the authors modeled the trajectory of the Archimedean spiral as an amplitude-modulated signal $\widehat{r}(t)$ defined by Equation 5.4.

$$\widehat{r}(t) = \left(a_3 t^3 + a_2 t^2 + a_1 t + a_0\right) \cdot \sin(2\pi f t) \tag{5.4}$$

The real trajectory $r(t)$ was modeled as a sinusoidal signal with increasing amplitude and frequency f. The amplitude coefficients (a_i) of the modeled trajectory values were estimated with a third-order polynomial regression based on the maximum peaks of the real trajectory. The third-order polynomial was chosen because it avoids an oscillatory behavior across the samples, and because it guarantees a smooth first derivative and a continuous second derivative across the trajectory. f is the fundamental frequency of the trajectory and it was estimated using the Fourier transform in the original implementation.

For this thesis, the original model is updated to consider two additional aspects not considered initially: (1) the initial phase of the trajectory to model the initial position and velocity of the pen, and (2) the frequency modulations that appear in the signal because the deceleration when the patients draw the outer loops of the spiral (it takes more time to complete the loop). With these two modifications, the model of the trajectory of the spirals changes to Equation 5.5.

$$\widehat{r}(t) = \left(a_3 t^3 + a_2 t^2 + a_1 t + a_0\right) \cdot \sin(2\pi f(t) \cdot t + \phi) \tag{5.5}$$

The frequency $f(t)$ is estimated according to the analytic signal $r_a(t)$ of the trajectory, shown in Equation 5.6. $\mathcal{F}$ is the Fourier transform, U the unit step function, and $h_r(t)$ the Hilbert transform of $r(t)$. The time derivative of the unwrapped instantaneous phase ξ is the instantaneous frequency $f(t)$, according to Equation 5.7.

$$r_a(t) = \mathcal{F}^{-1}\left(\mathcal{F}(r(t)) \cdot 2U\right) = r(t) + j h_r(t) = r_m(t) e^{j\xi(t)} \tag{5.6}$$

$$f(t) = \frac{1}{2\pi}\frac{d\xi}{dt}(t) \tag{5.7}$$

Finally, ϕ is estimated according to the cross-correlation between the real trajectory and an initial model that does not consider phase information. After ϕ estimation, both the trajectory and the model are phase-synchronized. Figure 5.4 shows an example of the real and modeled trajectories estimated for a PD patient and an HC subject. The figure also includes the instantaneous frequency estimated from the Hilbert transform for both trajectories. Note that the frequency is higher at the beginning for both subjects and then it starts to decrease. This behavior appears because at the beginning the inner loops of the spiral are smaller, thus it takes less time to draw them. Note also the oscillatory behavior of the instantaneous frequency for the PD patient, which could be an indicator about the presence of kinematic tremor.

Different features based on the geometry from the trajectory are extracted from the model, including the coefficients of the third order polynomial $(a_i, i \in \{0, 1, 2, 3\})$, the location and amplitude of the peaks in the trajectory, and the phase difference ϕ. The error between the real and modeled trajectories is computed using three different metrics: (1) the MSE, (2) the DTW distance to measure a time-aligned error between the real and modeled trajectories, and (3) the Frenchet distance, which takes into account the ordering of the points along the estimated and real trajectories. The Frenchet distance is defined as the shortest distance in-between two curves, where

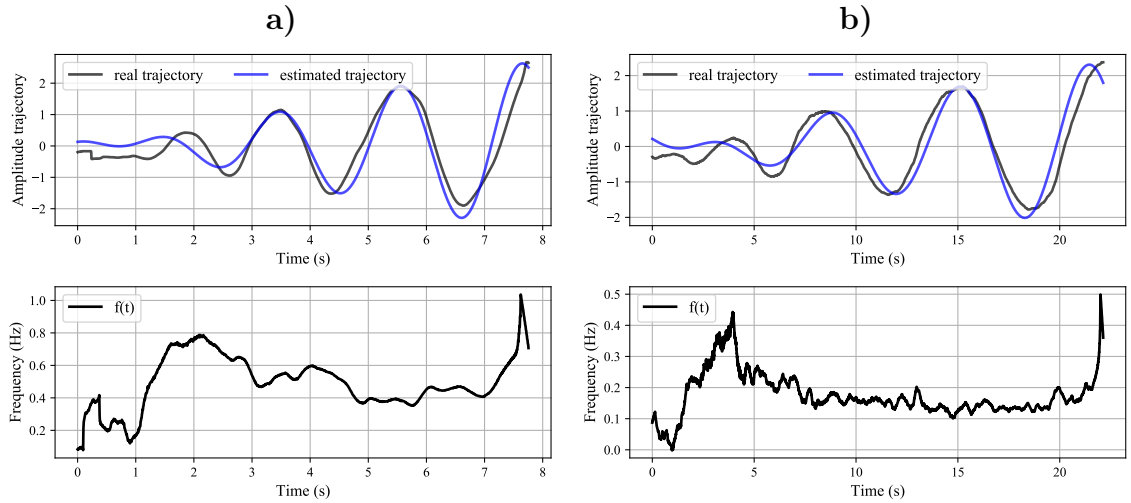

Figure 5.4: Difference in the real and modeled trajectories between a PD patient and an HC subject. **a)** Real and modeled trajectories (top), and instantaneous frequency of the real trajectory (bottom) for the Archimedean spiral drawn by a 71 years old HC subject. **b)** Real and modeled trajectories (top), and instantaneous frequency of the real trajectory (bottom) for the Archimedean spiral drawn by a 72 years old PD patient with MDS-UPDRS-III=44

it is allowed to change the velocity along each curve independently (walking dog problem) [Eite 94]. The feature set is completed with different statistical functionals (average, standard deviation, skewness, kurtosis, maximum, minimum, maximum position, and minimum position) computed over the instantaneous frequency $f(t)$ of the analytic signal.

5.4 In-air Analysis of Handwriting

The in-air movements of online handwriting has been considered particularly in the literature to model different handwriting impairments of PD patients [Drot 14, Drot 16, Vess 19]. The in-air trajectories corresponds to the movements performed by the hand while transitioning from one stroke to the next one. Most of the studies in the literature have explored a small set of in-air parameters, such as the in-air time, the velocity in-air, or different entropy measures over the in-air trajectories [Drot 14]. The analysis is extended for this thesis by including additional features to model other aspects of the in-air trajectories, not addressed previously. Particularly, the analysis is focused on the assessment of the transitions between in-air and on-surface segments, following the hypothesis that patients commonly exhibit difficulties to start/stop movements, both in the upper and lower limbs, and in the speech production system [Oroz 16b, Vasq 19c]. A set of 9 features is considered to model the in-air movement performed by PD patients: (1-2) the number of pen-ups and pen-downs per second to model mainly hesitations to start or to stop writing. (3-6) the average and the standard deviation of the slopes of the pen-up and pen-down transitions with the aim to model the stability of the hand when placing/lifting the pen from the tablet

surface. (7) the percentage of time in-air, similar to the included in [Drot 14]. (8) the Shannon entropy of the in-air trajectory to model the complexity of movements in-air, Finally (9) the LZC to model the complexity and repetitiveness of the binary sequence formed by in-air (1) on-surface (0) movements. The LZC reflects the rate of new patterns in the binary sequence. It ranges from 0 (deterministic sequence) to 1 (random sequence). Further details of the computation process can be found in [Kasp 87, Trav 17].

5.5 Deep Learning Models for Handwriting Analysis

Besides the previous models, which are based on different feature extraction strategies for a latter classification, end-to-end deep learning methods are also considered both to classify PD patients and HC subjects and to evaluate the motor state of the patients based on the MDS-UPDRS-III score. Two different models are proposed in this thesis to model both online and reconstructed offline handwriting data.

5.5.1 Deep learning for Online Handwriting Modeling

Online handwriting data of the patients is modeled in an end-to-end strategy using an improved version of the architecture introduced in [Vasq 19c]. The network is formed with a stack of 3 one-dimensional convolutional layers to process the raw signals collected from the tablet and, which include the horizontal (x) and vertical (y) positions, in-air movement, pressure of the pen, and azimuth and altitude angles. Figure 5.5 illustrates the considered network architecture. Other architectures were considered as well, like those based on combination of convolutional and recurrent layers; however, the performance of the current architecture was superior for the addressed experiments.

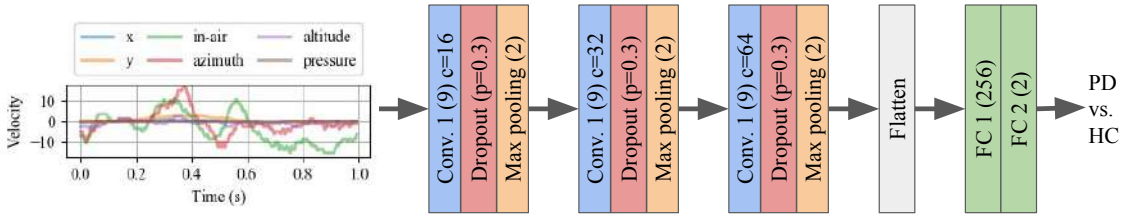

Figure 5.5: Deep learning model for end-to-end handwriting modeling of PD patients. **FC**: Fully connected layers. **c**= number of output channels in the convolutional layers. The values in parenthesis indicate the size of the convolutional filters and the number of neurons in the fully connected layers.

Two different inputs are considered for the CNN. The first one comprises the difference among consecutive samples of the sequences $\Delta S = \{s_2 - s_1, s_3 - s_2, \cdots, s_N - s_{N-1}\}$, computed for the six input channels (x, y, in-air, pressure, azimuth angle, altitude angle). The aim of using the difference is to transform the sequence from a point-level sequence, which depends on the position of the tablet, into a stroke-level sequence, which represents the direction of the pen movement [Zhan 16]. The second input sequence to the CNN corresponds to the handwriting transitions that occur

when a starting point of the stroke is detected (pen-down) or when the pen takes-off the surface of the tablet after drawing a stroke (pen-up). The aim is to evaluate the difficulties observed in the patients when they start/stop the handwriting movements. This is similar to the addressed in speech signals considering the onset and offset transitions to model the difficulties of patients to start/stop the vocal fold vibration, or in gait modeling the patient start or stop walking [Vasq 19c]. Once each pen-up or pen-down transition is detected, segments of 500 ms are taken to the left and to the right of the six signals captured with the tablet. Figure 5.6 shows the handwriting pen-down transitions of one HC subject and three patients in mild, intermediate, and severe states of the disease, respectively. Note that the dynamics of the in-air signals (black lines) is different for PD patients and HC subjects before starting the stroke (the first 0.5 seconds of the figure). Note that tremor is observed for the case of the patients, especially for the PD patient in Figure 5.6c), where oscillations around 7 Hz are observed when the pen is in the air.

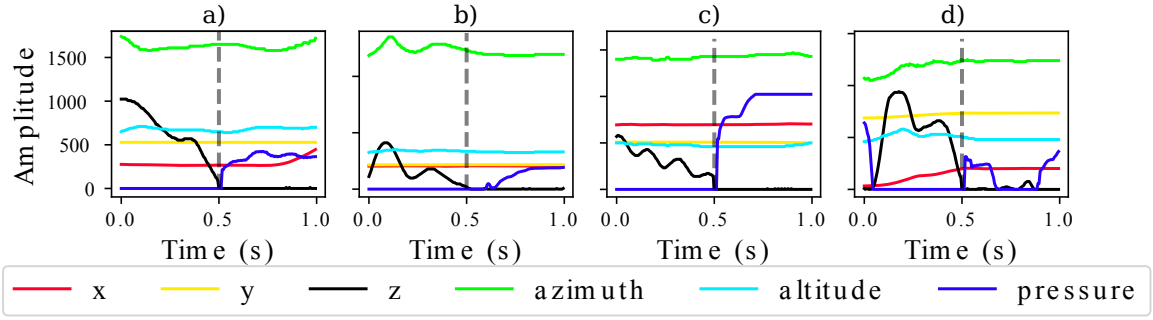

Figure 5.6: Handwriting pen-down transitions produced by: **a)** 68 years old male HC subject. **b)** 48 years old male PD patient in low state (MDS-UPDRS-III = 13). **c)** 41 years old male PD patient in intermediate state (MDS-UPDRS-III= 27). **d)** 75 years old female PD patient in severe state (MDS-UPDRS-III = 108).

5.5.2 Deep learning for Offline Handwriting Modeling

The aim of this model is to analyze and process the spatial patterns that appear in the drawings made by the patients and which reflect the presence of the disease. Those symptoms that are reflected in the images drawn by the patients may include the tremor observed in the drawing of Archimedean spirals (see Figure 5.3) or the micrographia that is present when patients write their name or a sentence.

The first step for this analysis is to reconstruct the images drawn by the patients from the online time-series. The following procedure was performed with the aim to reconstruct more realistic images, similar to what a patient would draw with a normal pen and paper. First a plot of the horizontal vs. the vertical positions of the pen is made, removing those points were the subjects have the pen in the air. The plot was made in gray-scale using the pressure of the pen as the gray level intensity for the reconstructed figure. The figure was then normalized to 8-bit integer values, similar to traditional RGB images. The images are then resized to 224x224 pixels to match the input to the networks used to process the ImageNet database. Finally, the background is removed and the colors are inverted in order to match the conditions of

the MNIST corpus to model handwriting digits. Figure 5.7 summarizes the performed procedure.

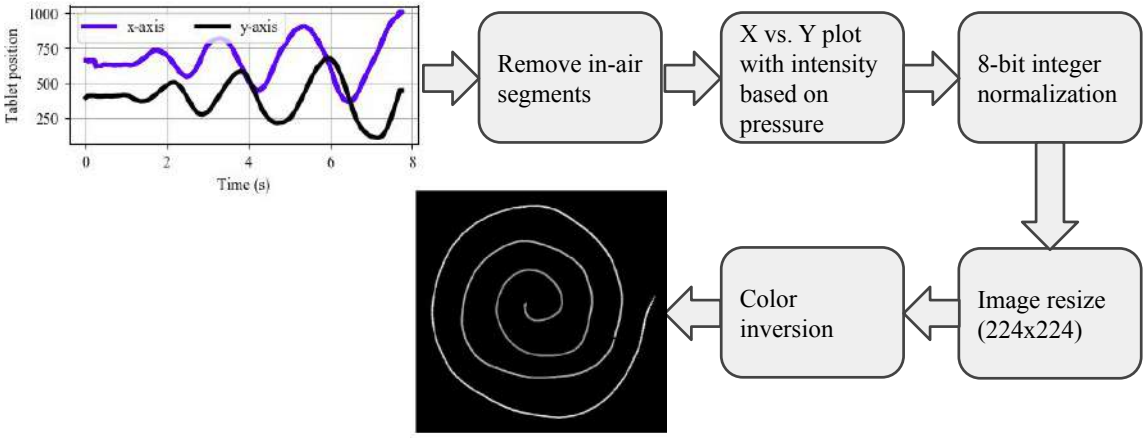

Figure 5.7: Pre-processing stages to reconstruct offline handwriting images from the online time-series.

The pre-processed images are used as input for a CNN model based on the SqueezeNet architecture [Iand 16]. The architecture has nearly 50 times less parameters than a CNN based on AlexNet [Kriz 12] but keeping a similar and competitive accuracy. The use of SqueezeNet makes the trained model available to be exported and used in low-power devices like smartphones, thus they can be included in further releases of Apkinson (see Section 8.2). A patient can perform one or several handwriting exercises in a normal pen and paper, and then take a picture with his/her smartphone, which will be processed locally to evaluate the upper motor skills of the patients. The main ideas/strategies of SqueezeNet include: (1) to make the network smaller by replacing the 3×3 filters used in the literature by that time, with 1×1 filters, which has 9 times fewer parameters. (2) to reduce the number of inputs for the remaining 3×3 filters, which is achieved by using only 1×1 filters prior to the 3×3 convolutional layer. (3) to make the downsample operation late in the network thus convolution layers have large activation maps. These three strategies are implemented into what the authors called the *Fire* module, which is the main building block used in SqueezeNet. The Fire module comprises *Squeeze* layers which are convolutional layers with 1×1 filters, and *Expand* layers which have a mix of 1×1 and 3×3 convolution filters. In addition, the number of filters in the squeeze layer must be less than the ones in the expand layer. Figure 5.8 shows an example of how the Fire module looks like.

The complete architecture of SqueezeNet with the Fire modules is shown in Figure 5.9. The network is formed with two parts: a feature extraction block formed with the first convolutional layer (Conv. 1) and 8 Fire modules, and the classification stage formed with a 1×1 convolutional layer with two channels (Conv. 2) and a global average pooling layer. The SqueezeNet model is trained using a transfer learning strategy, using pre-trained models from the ImageNet corpus. The weights of the layers corresponding to the feature extraction part were frozen i.e, they were

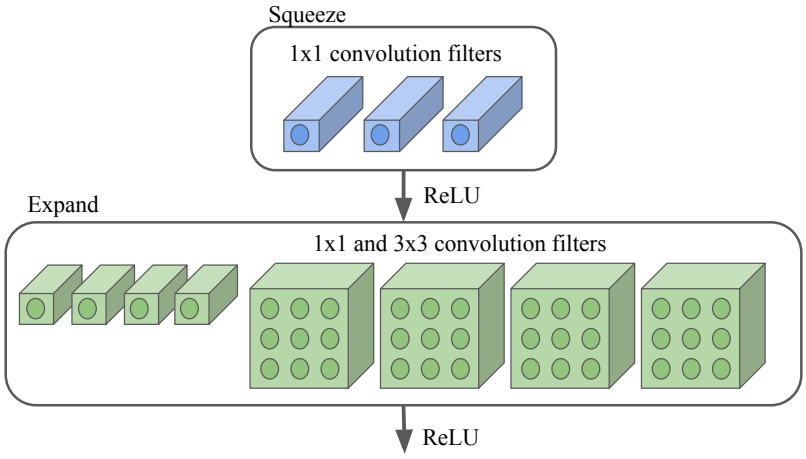

Figure 5.8: Organization of convolution filters in the Fire module of SqueezeNet. Adapted from [Iand 16]

kept from the pre-trained model. Only the weights for the layers of the classification part are updated.

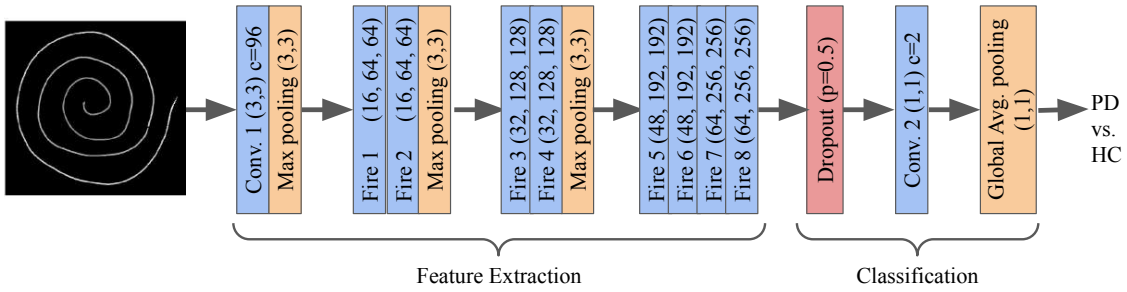

Figure 5.9: Full architecture of SqueezeNet. The values in parenthesis for the 8 Fire modules indicate the number of 1×1 filters in the Squeeze layer, the number of 1×1 filters in the Expand layer, and the number of 3×3 filters in the Expand layer, respectively.

Chapter 6

Analysis of Parkinson's Disease from Gait

One of the major manifestations of PD appears in gait, and typically causes disability of PD patients. More than 85% of PD patients develop gait impairments after three years of diagnosis [Kell 12]. The main symptoms in the gait of PD patients include speed reduction, smaller stride length, altered cadence, and increased gait variability. In the earliest stages of the disease bradykinesia is reflected in smaller arm swing, slower turns and reductions in step length [Yang 08]. In addition, although gait impairments are not clearly exhibited in early stages, their prevalence and severity increase with the disease progression [Kell 12], where gait becomes more unstable, freezing of gait (FoG) episodes occur, and falls are frequently reported [Galn 15].

The potential consequences of gait impairments in PD include increased disability, risk of falls, and reduced quality of life. As the disease progresses, PD patients typically exhibit shuffling gait with a forward-stooped posture and festinating gait. These characteristics make the patients to spend a lot of energy while walking, leading them to their maximum metabolic capacity every day [Coen 13]. PD patients consider mobility and walking limitations the most disabling aspects of the disease and consistently identify improvement in walking as the most relevant outcome when rating the success of the treatment [Hass 12b].

One of the most common symptoms in the gait of the patients is FoG, which is defined as *an absence or marked reduction of forward progression of the feet despite the intention to walk* [Nutt 11]. Patients describe FoG as a feeling of having the feet glued to the ground and being temporarily unable to re-initiate gait. FoG is context-dependent, that is, it triggers when patients walk through narrow spaces, initiate or end gait, an obstacle avoids patients to follow their gait trajectory, or during turns [Scha 03]. It has been observed that FoG produces harmonics in acceleration signals between 3 and 8 Hz [Moor 08]. Another important symptom of the patients is tremor, which is defined as *a rapid back-and-forth movement of a body segment* [Bach 89]. Tremor in PD patients appears mainly at rest, and tends to disappear during posture or movement [Deus 96]. However, in severe stages of the disease, the tremor may remain present during movement, and it is called kinetic tremor [Wenz 00]. The frequency associated to resting tremor typically ranges between 3.5 and 7.5 Hz [Clee 87], and the frequency for kinetic tremor ranges from 4 to

12 Hz [Wenz 00]. Additional symptoms in the gait of PD patients include increased double support time [Sala 04], reduced stride length and speed [Chan 16], decreased gait symmetry and regularity [Demo 15], reduced range of rotation of thigh, knee, trunk, and foot [Sala 04], and reduced knee symmetry [Toos 15].

This chapter describes state-of-the-art methods and different proposed algorithms to model lower limbs movement in PD patients, both from a classical pattern recognition perspective and novel deep learning strategies. Section 6.1 shows the reader a review of the literature about automatic gait assessment of PD patients using inertial sensors both to classify PD patients and HC subjects and to evaluate the neurological state of the patients. Studies based on video analysis of gait of PD patients are left out of the review. Then, Section 6.2 describes the classical kinematic features used in the literature to model gait in PD patients, and based mainly on the models proposed in [Bart 17]. Section 6.3 describes a set of spectral features to model the harmonic structure and spectral wealth of gait signals from PD patients. Section 6.4 described the application of NLD features to model gait impairments in PD patients. The chapter finishes in Section 6.5 with the description of methods to model the gait of PD patients in an end-to-end strategy using different configurations of convolutional and recurrent neural networks.

6.1 A Review on Automatic Assessment of Gait in PD Patients

The research community has shown a growing interest in the automatic gait analysis of PD. The studies are focused on the classification of PD patients and HC subjects, the assessment of the neurological state of the patients, and the detection of specific walking impairments in the patients, such as FoG episodes. The analyses have been performed commonly with inertial sensors e.g., accelerometers and gyroscopes attached to the body of the patients [Kluc 13, Shul 14, Oung 15, Hann 17], and with force-sensitive sensors placed inside the shoes of the participants [Xia 15, Ren 16]. There are also some studies that consider gait acquisition using walkway paths [Hass 12a, Rehm 19]; however, their use is restricted only to clinical environments. Gait analysis using wearable sensors is expected to play an increasingly important role in the assessment of PD. By using inertial sensors, it is possible to detect and characterize specific movements and register variations in the clinic and to monitor activities of daily living of PD patients at their own home [Brog 19].

Most of the studies have considered kinematic features based on the duration and velocity of the steps [Kluc 13, Pari 15, Ren 16, Hann 17, Djur 17, Kim 15, Cara 18]. Other studies have considered spectral features [Mazi 12, Das 12, Sanc 18, Sama 18] to evaluate the harmonic structure of the gait process of PD patients, compared to HC subjects. Despite the majority of studies which use kinematic, spectral and statistical features to model gait impairments of PD patients, there are some studies to model non-linearities that appear during the walking process [Sejd 14, Prab 20, Trip 13, Xia 15, Pere 18c, Chom 19]. For instance, a higher complexity and randomness has been observed in the accelerations of healthy gait in the sagittal plane, compared to PD patients due to the step-to-step adjustments for an effective bal-

ance control [Sejd 14, Baub 00]. Conversely, in the frontal and transversal planes, higher complexity and chaotic behavior can be observed in the gait of PD patients due to the tremor condition and the FoG episodes [Sejd 14]. Other studies that have considered the raw data as input for machine learning pipelines based mainly on sequence learning strategies via hidden Markov models [Cuzz 17], or novel deep learning algorithms [Camp 18].

6.1.1 Automatic Classification of PD and HC Subjects

Regarding the gait analysis using inertial sensors attached to the body of the patients, one of the first studies for automatic assessment of PD patients was performed in [Kluc 13]. The authors classified gait signals captured from 42 PD patients and 39 HC subjects using the eGaIT system[1]. The participants performed several exercises, including walking 10 meter 4 times at a comfortable walking speed (4× 10 task), heel-toe tapping, and circling foot movements. The authors computed several spectral and statistical features, including the energy content in several frequency bands, the variance, the root-mean square energy, among others. The authors reported accuracies up to 82% using different classification strategies. The accuracy improved up to 91% when only considering PD patients in severe state of the disease. In [Sejd 14] the authors classified gait signals from 10 PD patients and 14 HC subjects using several statistical, spectral, and NLD features. The signals were captured with inertial sensors placed at the L3 segment of the lumbar spine when the participants walked on a treadmill. Statistical features included the variance, skewness, and kurtosis of the signals. Spectral features included the peak frequency, spectral centroids, the bandwidth of the signal, and the energy content distributed in several bands according to the discrete wavelet transform. NLD features included the LLE, the LZC, and several entropy measures. The results from different statistical tests showed that features such as the skewness, kurtosis, LZC, entropy, the centroid frequency and the wavelet bands were able to discriminate between the HC subjects and the PD patients. In [Bart 17] the authors proposed a set of kinematic and statistical features to classify 190 PD patients and 101 HC subjects with gait signals collected using the eGaIT system. Kinematic features are computed over the single steps, which are segmented by comparing the strides of the participants with a template, using DTW. Accuracies of up to 82% were reported with the proposed approach, using an AdaBoost classifier. In [Djur 17] the authors aimed to identify the most accurate kinematic features to classify 40 PD patients and 40 HC subjects. The authors extracted several kinematic features and proposed a feature selection method based on an affinity propagation clustering and a random forest classifier. The selected features were classified using an SVM. The selected features included: the stride length, the stride time, the swing time, and the step time asymmetry. The selected features showed an accuracy of up to 85%. The authors in [Cuzz 17] considered data from inertial sensors attached to the lower spine of 24 HC subjects and 156 PD patients, who performed a 10 meter walking test. The raw gait data were used to train a hidden Markov model to discriminate between the patients and the healthy subjects. An F1-score of up to 0.81 was reported by the authors. In [Kuhn 17] the

[1] https://www.egait.de/

authors classified 26 HC subjects and 14 PD patients in OFF state by turning off their implanted DBS device. A set of 17 inertial sensors were placed in different parts of the body, covering both the upper and lower limbs. The patients performed several exercises, including a TUG test, and a 10 meter walking. The authors computed kinematic features based on the velocity of the strides. The authors reported an accuracy of up to 94.6% in the classification task using a random forest classifier. Results covering the classification of HC subjects and patients in ON state, which is a more challenging problem were not reported by the authors. In [Cara 18] the authors classified 25 PD and 25 HC subjects using information from 8 inertial sensors placed in different parts of the body. The participants performed a 15 meter walking exercise. The authors computed two different types of kinematic features: (1) range of motion features, which are defined as the difference between the maximum and minimum angle drawn in the sagittal plane between two adjacent articular segments within one gait cycle; and (2) standard kinematic features such as stride time, stride length, among others. The extracted features were classified with different algorithms. In addition, a majority voting strategy was considered to combine the information from the different sensors and feature sets. The authors reported accuracies over 90%. The authors in [Pere 18c] considered different NLD features to classify a set of age-balanced 45 PD patients and 45 HC subjects. The NLD features include the CD, LLE, HE, and the LZC, in addition to six entropy measures to quantify the general regularity of the gait process. Gait signals were collected with the eGait system, described previously. The authors reported an accuracy up to 85% using RF and SVM classifiers. The study from [Pere 18c] was extended in [Pere 20b], where the authors proposed a novel NLD analysis based on spatial clustering of Poincaré sections using GMMs. The proposed models were used to classify PD patients and HC subjects and to discriminate patients in different stages of the disease according to three levels of the MDS-UPDRS-III score. Accuracies of up to 86.7% were obtained in the 2-class problem. The automatic classification of three different stages of the disease shows accuracies of around 65%. In [Rehm 20], the authors aimed to compare different sets of kinematic, spectral, and NLD features to classify 81 PD patients and 61 HC subjects. The subjects wore inertial sensors attached to the lower back and performed a 2 minute walking test. The authors reported an accuracy of up to 87.3% using a partial least square discriminate analysis classifier. The best result was reported with the combination of spectral and NLD features.

There are some studies that have considered gait signals collected with pressure sensors placed in the sole of the feet of the patients. Those studies mainly considered the Physionet corpus, described in Section 3.2.3. For instance, the authors in [Xia 15] classified gait signals from 15 PD patients and 16 HC subjects from the Physionet corpus. The authors considered several statistical and NLD features computed from the raw signals, including the LZC, the fuzzy entropy, and the Teager-Kaiser energy. The features were classified using several algorithms. The authors reported accuracies close to 100% discriminating between PD patients and HC subjects. The Physionet corpus was also considered in [Ren 16]. The authors computed several kinematic features such as the stride time, the swing time, the stance time, among others. The features extracted from each foot were used to compute a phase synchronization coefficient and the conditional entropy to analyze the gait rhythm fluctuations

between both feet. The authors considered several classifiers, and reported an AUC of up to 0.928. The authors in [Prab 20] considered also the Physionet database to classify 13 PD patients and 13 HC subjects. The authors proposed NLD features based on recurrence quantification analysis. The classification was performed with an SVM and with a probabilistic neural network, which achieved an accuracy of up to 94.4% In [Ertu 16] the authors proposed a statistical transformation called shifted one-dimensional local binary pattern domain to classify the 93 PD patients and 73 HC subjects from the Physionet corpus. The authors computed several statistical features, which were transformed with the proposed method, and then classified using eight different algorithms. The highest accuracy was obtained with a neural network classifier (88.9%). The Physionet corpus was considered in [Zeng 19] to classify PD and HC subjects. The authors proposed a set of non-linear features extracted from the phase space using the empirical mode decomposition, which was applied to the euclidean distances computed between the trajectory of the attractor and the origin. The extracted features were classified with a neural network with a Gaussian kernel. The authors reported an accuracy of up to 97%. Novel deep learning strategies have also been considered to classify the gait of PD patients and HC subjects. In [Zhao 18] the authors classified data from the Physionet database using a combination of a two-layer CNN with a two-layer LSTM networks, which process the raw signals from the pressure sensor. The authors classified HC subjects and PD patients in three levels of the disease distributed based on the H&Y scale. The authors reported an accuracy over 98% for the four-class problem.

6.1.2 Automatic Evaluation of the Neurological State of Patients

One of the first studies that considered inertial sensors to predict the neurological state of PD patients was reported in [Sala 04]. The authors attached several gyroscopes to the lower and upper limbs of 10 PD patients with an implanted DBS and 10 HC subjects. The gait signals for the patients were captured both when the DBS was ON and OFF. The authors computed several kinematic features, including the stride length, stride velocity, stance time, double support time, and gait cycle time. The results indicated that the DBS significantly improved the gait performance. In addition, some of the features such as the stride length were highly correlated with the UPDRS sub-score for lower limbs (Pearson correlation of 0.87). In [Pari 15] the authors considered inertial sensors attached to the chest and to the kneels to assess the neurological state of 34 PD patients according to the UPDRS score. The participants performed several tasks, including 20 meter walking, TUG, and foot tapping. The authors computed kinematic features such as the standing time, the stride length, the stride velocity, among others. The regression algorithm was based on a KNN to predict the UPDRS score of the patients. A Spearman's correlation coefficient of 0.60 was reported for the prediction.

There are studies to assess specific items of the MDS-UPDRS score of the patients. In [Riga 12], the authors evaluated the severity of kinetic and resting tremor in 18 PD patients and 5 HC subjects, who performed several exercises using inertial sensors attached to different parts of the body. The authors computed several

spectral features such as the dominant frequency, the energy in different frequency bands, the entropy of the spectrum, among others. The participants were classified into four classes according to the tremor severity score from the UPDRS scale. The authors reported an accuracy of 87% with the proposed approach. In [Orne 17] the authors proposed a model to quantify the turning capacity of 46 PD patients. The patients were asked to walk 10 meter in a straight line, turn 180 degrees, and return to the starting point using inertial sensors placed in their ankles. The proposed model consisted of kinematic features such as the number of strides, the turning time, the number of continuous strides, and the number of hesitations. Those features were used to train a fuzzy inference system to map the kinematic features into a continuous scale related with the turning capabilites of the patients. The output of the model was consistent with the MDS-UPDRS-III item for gait; however, the accuracy was not reported. In [Sanc 18] the authors estimated the resting tremor severity using inertial sensors placed on top of the dorsal side of the hand, and in front of the shank right on top of the ankle. Data was collected from 57 PD patients, who performed a complete motor examination test based on the MDS-UPDRS-III scale. Kinematic and spectral features were considered to detect the presence of resting tremor in the exercises performed by the patients. Then, the estimation of the tremor severity was performed with a fuzzy inference system, which transforms the extracted features into five categories, namely: normal, slight, mild, moderate, and severe. Unfortunately, the authors did not correlate the output of the proposed inference model with the clinical score assigned to the patients. In [Agha 20] the authors predicted different subscores of the MDS-UPDRS-III scale for lower limbs, including the bradykinesia sub-score, and the sum of the scores for leg agility, rising from a chair, and gait. The proposed models were evaluated with data from 19 PD patients, who used inertial sensors placed in their ankles when they performed a heel tapping exercise. The authors computed spectral features based on the DWT, and NLD features such as the approximate entropy. The clinical scores were predicted using an SVR. The authors reported the results in terms of the intra-class correlation coefficient, and reported a correlation of 0.83 for the bradykinesia subscore, and of 0.78 for the combination of the other sub scores for lower limbs impairments. The authors in [Orne 19] proposed a model to quantify the leg agility of 50 PD patients, who performed a heel tapping exercise with inertial sensors attached to their ankles. The authors computed features related to the heel tapping exercise such as the amplitude when rising the leg, number of hesitations, the amplitude trend, and the speed trend. These features were used to create a fuzzy inference system to quantify the leg agility. The authors reported accuracies over 92.4% with respect to the MDS-UPDS-III item for leg agility. The authors from [Borz 20] aimed also to quantify the leg agility from PD patients by predicting its item from the MDS-UPDRS-III. Data from 93 PD patients were collected using the inertial sensors embedded in a smartphone. The patients performed a heel tapping test with the smartphone attached to the leg with an elastic band. The authors computed several kinematic and spectral features from the collected data and reported an accuracy of 77.7% predicting the four classes of the leg agility item using an MLP classifier. The authors in [Rava 20a] aimed to predict the progression rate of 160 PD patients over two years. The authors defined patients with fast progression as those ones who their MDS-UPDRS-III increases by more than 20% over

two years. The authors considered inertial sensors attached to different parts of the body, including ankles, wrist, lower back, and chest, and compute different kinematic features over the collected data. The patients performed the TUG test and the sway test, i.e., subjects stand still with their feet separated a certain distance and their hands across their chest for 30 seconds. The authors were able to detect the fast progression patients with a positive predictive value of 71%. A different application was considered in [Pfis 20]. The authors classified continuous states of ON, OFF, and dyskinesia events based on the bradykinesia item of the MDS-UPDRS scale. The authors trained an end-to-end CNN using raw signals with 1 minute length. The signals were collected with accelerometers placed in the wrist, and data were collected from 30 PD patients who performed different activitis of daily living. The authors reported an accuracy of 65.3% for the three-class problem.

6.1.3 Automatic Detection of FoG and other Gait Impairments

One of the most common symptoms addressed by the research community interested in gait assessment of PD patients is FoG [Silv 17]. Most of the algorithms considered statistical features [Mazi 12, Rodr 17, Sama 18], NLD features [Trip 13, Chom 19], the Freeze index (FI) [Kim 15, Zach 15, Rezv 16], or novel deep learning strategies [Camp 18]

One of the first studies to detect FoG episodes was performed in [Mazi 12], where the authors detected FoG episodes during walking, for a set of 10 patients. Several accelerometers were placed in the ankle, knee, and hip of the patients, who simulated walking activities of daily living like raising from bed and go to the kitchen. Statistical and spectral features were computed from the signals. The extracted features were used to discriminate between FoG episodes and normal walking, with accuracies of up to 95%. In [Trip 13] the authors detected FoG events considering six accelerometers and two gyroscopes placed in the lower and upper limbs of 5 HC subjects and 11 PD patients. The participants performed activities of daily living such as rising from a chair, free walking, opening and closing doors, making stops to drink water, among others. The collected signals were characterized with entropy measures. The FoG events were detected using several classifiers, which achieved accuracies of up to 96.1%. In [Kim 15] the authors considered only the inertial sensors from a smartphone to detect FoG events from 15 PD patients. Subjects were instructed to walk straight 3 meters. Then, they turned around and returned to their place. The smartphone was placed in several parts of the body when the patients performed the exercises. The authors defined the FI as the ratio between the acceleration power in the band 3-8 Hz and the power in the band 0.5-3 Hz. The FI was combined with other statistical and spectral features to classify the FoG episodes using an Adaboost classifier. The highest accuracy was reported when the smartphone was placed in the waist (86%). The FI was also considered in [Zach 15] to detect FoG episodes in 23 PD patients, who performed several walking activities to provoke the FoG episodes using a single inertial sensor placed in the wrist. The authors reported an accuracy over 75% using only the FI and a classifier based on a threshold. A modified version of the FI, computed using the continuous wavelet transform (CWT) was proposed

in [Rezv 16] to detect FoG episodes in 10 PD patients, who use a single inertial sensor placed in the shank. The detection of the FoG episodes was performed with a threshold. The reported results indicate a sensitivity of 82.1% and a specificity of 77.1%. In [Ahlr 16] the authors detected FoG episodes in 20 PD patients, who use an inertial sensor placed in the wrist. Several spectral features were computed from the collected signals. The detection of the FoG events was performed with an SVM classifier, which provided an accuracy of up to 96.1%. In [Rodr 17] the authors detected FoG episodes in 21 PD patients using several statistical features and an SVM classifier. The features were extracted from signals collected with an inertial sensor placed in the wrist of the patients, while they performed several walking exercises at home. The proposed approach achieved a sensitivity of 74.7% and a specificity of 79%. Recently, in [Sama 18] the authors detected FoG episodes of 15 PD patients from patients at home. The patients used an inertial sensor at the waist, while performing four activities: (1) showing the house, (2) a FoG provocation test by walking in a narrow space with several turns, (3) going outdoors for a short walk, and (4) a dual task activity (walking while carrying and object). The authors computed several statistical and spectral features from the raw signals, and used several classifiers to detect the FoG episodes. The features were computed in three frequency regions: (1) from 0.04 to 0.68 Hz, which corresponds to the posture transition band, (2) from 0.68 to 3 Hz, which corresponds to walking frequency content, and (3) between 3 and 8 Hz, which is related to FoG episodes. The results reported showed that the proposed approach was able to detect the FoG episodes with a sensitivity of 91.7% and a specificity of up to 87.4%. In [Chom 19] the authors detected FoG events in 21 PD patients and 9 HC subjects. The participants performed several activities such as turning, carrying a cup while walking, and other dual-task activities to provoke FoG episodes. Gait data were collected with inertial sensors from an i-Pod touch that the participants have inside their pockets. The proposed model consisted of the use of NLD features based on recurrent quantification analysis and an SVM classifier. The authors reported an accuracy of up to 99.3% with the proposed approach.

Deep learning strategies have also been considered to detect FoG episodes in PD patients. In [Camp 18] the authors detected FoG episodes in PD patients using a deep learning approach. The data were collected from 21 PD patients with FoG using a waist-placed inertial sensor. The tasks performed by the participants included free walking inside an apartment, walking 10 meters outdoors, and a TUG test. The authors considered a six-layer one-dimensional CNN. In total, the input of the CNN consisted of 64 frequency bins obtained from 9 inertial sensors (3-axis accelerometer, gyroscope and magnetometer) along with those obtained from the previous time interval, forming a tensor $\mathrm{X} \in \mathbb{R}^{64\times1\times18}$. Accuracies of up to 90% were reported for the addressed problem. The authors from [Torv 18] proposed a method based on LSTM units to detect the time instant before the appearance of the FoG events in a set of 5 PD patients from the Daphnet FoG dataset [Bach 10]. The patients performed several walking activities with inertial sensors attached to their ankles. The raw gait data was used to train the LSTM network. A transfer learning strategy was applied to adapt the model obtained from different subjects into a specific model for the target subject. Accuracies over 85.5% were reported by the authors. The accuracy increased up to 93% with the transfer learning strategy, leading to a patient dependent

model. In [Xia 18] the authors detected FoG episodes in the 10 PD patients from the Daphnet FoG corpus [Bach 10], using a 1D-CNN to process the raw signals captured with the inertial sensors placed in different parts of the body. The authors reported an accuracy of 80.7% in a patient independent strategy, while the accuracy for a patient dependent model improved up to 99%. The Daphnet FoG corpus [Bach 10] was also considered in [Moha 18] to predict the FoG of PD patients. The authors proposed an unsupervised learning approach based on a novelty detection strategy. Data from HC subjects were used to train a convolutional autoencoder with the aim to model the distribution of healthy gait and to obtain a reference model. Then, the predictions from the autoencoder for the PD gait were compared to the reference model using a distance measure based on the generalized extreme value distribution. The decision to detect the FoG episode was performed using a threshold of the distance. The authors reported an AUC of 0.77 with their proposed unsupervised strategy.

There are also studies to detect other movement impairments of the patients besides FoG. For instance in [Das 12] the authors detected dyskinesia episodes in 2 PD patients using five inertial sensors attached both to lower and upper limbs. The patients were continuously monitored during four consecutive days. The signals obtained from the inertial sensors were characterized using statistical and spectral features. Then, a multi-instance classification strategy was considered to detect the presence of dyskinesia events in the patients in a semi-supervised way. The authors reported accuracies up to 90% with the proposed approach.

6.1.4 Main Outcomes from the Literature

Although the increasing research in automatic gait assessment of PD patients using wearable sensors, there are several challenges that need to be addressed. Most of the proposed methods and systems are highly variable among themselves. There is no standard procedures about the place where the wearable sensors have to be attached, the exercises to be performed by the patients, or the features to be computed [Brog 19]. These aspects are seen in Tables 6.1, 6.2, 6.3, and 6.4.

Different applications have been considered for the gait assessment of PD patient. A summary about the considered applications within the last years is shown in Figure 6.1 and Table 6.1. Most of the papers are focused on the classification of PD vs. HC subjects, the assessment of the neurological state of the patients, or the prediction of FoG events on the patients. Particularly, regarding the assessment of the neurological state of the patients, the studies are focused on the prediction of the total MDS-UPDRS scale, or some specific items related to gait impairments such as bradykinesia, rigidity or tremor. There are also some studies focused on longitudinal evaluation of the patients [Abra 20, Rava 20a]. Other studies are focused on the classification between ON and OFF states [Pfis 20], the detection of dyskinesia events [Das 12], or the assessment of the turning capacity of PD patients [Orne 17].

Figure 6.2 and Table 6.2 show the most common methods used by the research community for gait assessment of PD. It can be noted that many of the current approaches have focused on extracting features based only on the kinematic analysis of the gait process, e.g., amplitude, frequency components, cadence, stride length, among others, which are not specifically designed for PD. This fact makes still unclear

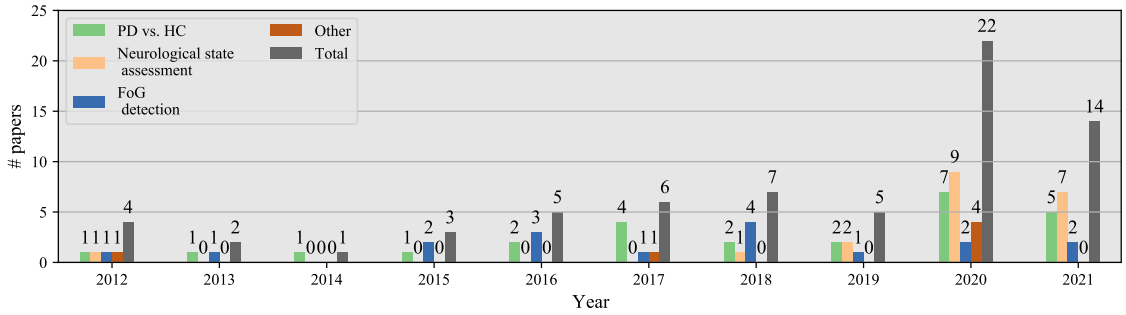

Figure 6.1: Different applications addressed in the literature for gait assessment of PD.

Table 6.1: Different applications addressed in the literature for gait assessment of PD.

References	Application
[Bart 12a], [Kluc 13], [Sejd 14], [Xia 15], [Ertu 16], [Ren 16], [Bart 17], [Cuzz 17], [Djur 17], [Kuhn 17], [Cara 18], [Prab 20], [Zeng 19], [Vasq 19c], [Pere 20b], [Rehm 20], [Oroz 20b], [El M 20], [Alkh 20], [Alha 20], [Seti 21], [Vasq 21a], [Varr 21], [Bala 21a], [Cant 21a]	Classification PD vs. HC
[Sala 04], [Riga 12], [Zhao 18], [Agha 20], [Orne 19], [Vasq 19c], [Abra 20], [Pere 20b], [Rava 20a], [Borz 20], [Oroz 20b], [Bala 20], [El M 20], [Alha 20], [Seti 21], [Vasq 21a], [Bala 21a], [Kher 21], [Vidy 21], [Cant 21a], [Bala 21b]	Neurological state assessment
[Mazi 12], [Trip 13], [Kim 15], [Zach 15], [Ahlr 16], [Mazi 16], [Rezv 16], [Rodr 17], [Camp 18], [Moha 18], [Sama 18], [Xia 18], [Chom 19], [Asho 20], [Zhan 20], [Borz 21], [Nagh 21]	FoG detection
[Das 12], [Orne 17], [Pfis 20], [Rava 20a], [Moon 20], [Pere 20a]	Other

which gait features are the most informative to assess gait in PD [Brog 19]. The only proposed feature designed specifically for gait assessment in PD is the FI [Kim 15]. The use of NLD-based features seemed to be suitable for the assessment of the disease, and should be further explored, according to the literature. In addition, the use of feature learning strategies using deep learning methods have not been enough explored in the assessment of gait impairments of PD patients. These features can be able to improve the performance of the current systems for assessment of gait disturbance, especially in PD.

Regarding the different gait exercises to be performed by patients, Table 6.3 and Figure 6.3 show the ones that are most common in the literature. Most of the studies considered straight walking or free walking exercises, which are designed and robust for unobtrusive monitoring of the patients, especially in at-home environments. More controlled exercises include heel tapping and TUG test, which are included mainly because they are part of the MDS-UPDRS evaluation. There is an increasing use of dual tasks, e.g., carrying objects while walking, which can be explored in further research to evaluate the effect of those secondary cognitive tasks in the walking process of PD patients.

Regarding the location of the sensors to evaluate the gait impairments of the patients, there is a high variability in the positions considered in the literature (see

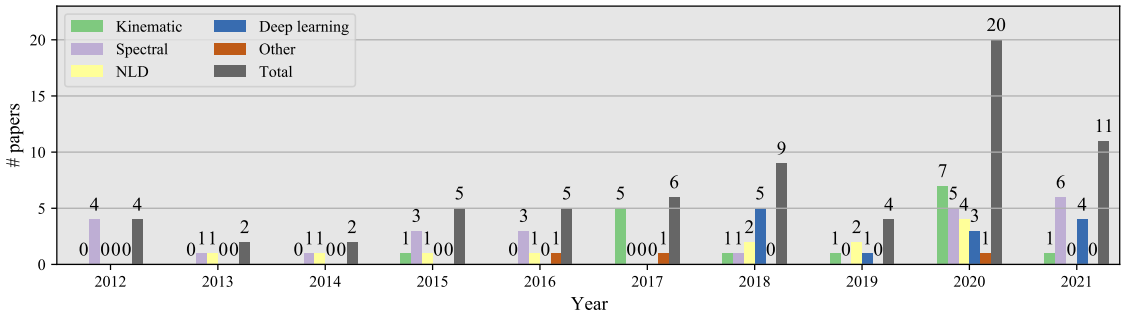

Figure 6.2: Different methods considered in the literature for gait assessment of PD.

Table 6.2: Different methods considered in the literature for gait assessment of PD.

References	Method
[Sala 04], [Bart 17], [Djur 17], [Kuhn 17], [Orne 17], [Rodr 17], [Pari 15], [Cara 18], [Orne 19], [Rava 20a], [Rehm 20], [Borz 20], [Oroz 20b], [Bala 20], [Alkh 20], [Moon 20], [Varr 21]	Kinematic
[Bart 12a], [Das 12], [Mazi 12], [Riga 12], [Kluc 13], [Sejd 14], [Kim 15], [Xia 15], [Zach 15], [Ahlr 16], [Mazi 16], [Rezv 16], [Sama 18], [Agha 20], [Rehm 20], [Borz 20], [Alha 20], [Seti 21], [Bala 21a], [Borz 21], [Zhan 20], [Kher 21], [Bala 21b], [Vidy 21]	Spectral
[Trip 13], [Sejd 14], [Xia 15], [Ren 16], [Prab 20], [Pere 18c], [Agha 20], [Chom 19], [Zeng 19], [Pere 20b], [Rehm 20], [Oroz 20b]	NLD
[Camp 18], [Moha 18], [Torv 18], [Xia 18], [Zhao 18], [Vasq 19c], [Pfis 20], [El M 20], [Asho 20], [Seti 21], [Nagh 21], [Vasq 21a],, [Cant 21a]	Deep learning
[Ertu 16], [Cuzz 17], [Abra 20]	Other

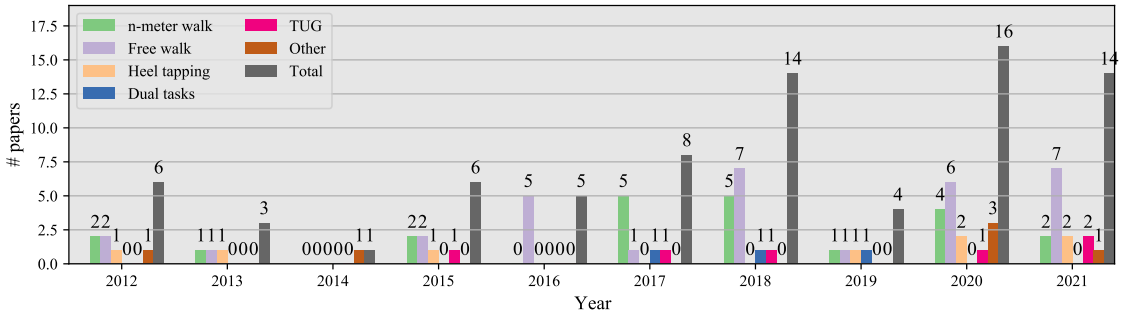

Figure 6.3: Different tasks considered in the literature for gait assessment of PD.

Table 6.4). Even some of the developed systems require multiple sensors located on different parts of the body, which is impractical for many clinical and especially home-based applications [Brog 19].

Table 6.3: Different tasks considered in the literature for gait assessment of PD.

References	Task
[Sala 04], [Bart 12a], [Riga 12], [Kluc 13], [Kim 15], [Pari 15], [Bart 17], [Cuzz 17], [Djur 17], [Kuhn 17], [Orne 17], [Camp 18], [Cara 18], [Pere 18c], [Torv 18], [Xia 18], [Vasq 19c], [Pere 20b], [Oroz 20b], [Moon 20], [Pere 20a], [Vasq 21a] [Vidy 21]	n-meter walk
[Bart 12a], [Kluc 13], [Pari 15], [Agha 20], [Orne 19], [Borz 20], [Vasq 21a] [Varr 21]	Heel tapping
[Das 12], [Mazi 12], [Trip 13], [Xia 15], [Zach 15], [Ahlr 16], [Ertu 16], [Mazi 16], [Ren 16], [Rezv 16], [Rodr 17], [Camp 18], [Moha 18], [Prab 20], [Sama 18], [Torv 18], [Xia 18], [Zhao 18], [Zeng 19], [Abra 20], [Pfis 20], [Rehm 20], [Bala 20], [El M 20], [Alkh 20], [Alha 20], [Seti 21], [Vasq 21a], [Bala 21a] ,[Kher 21], [Cant 21a], [Bala 21b]	Free walk
[Djur 17], [Sama 18], [Chom 19]	Dual tasks
[Pari 15], [Kuhn 17], [Camp 18], [Rava 20a], [Vasq 21a], [Borz 21]	TUG test
[Riga 12], [Sejd 14], [Rava 20a], [Asho 20], [Zhan 20], [Nagh 21]	Other

Table 6.4: Location of sensors considered in the literature for gait assessment of PD.

References	Location
[Bart 12a], [Kluc 13], [Bart 17], [Pere 18c], [Vasq 19c], [Pere 20b], [Oroz 20b], [Pere 20a], [Vasq 21a]	Shoes
[Sejd 14], [Cuzz 17], [Xia 18], [Rehm 20], [Zhan 20]	Back
[Xia 15], [Ertu 16], [Ren 16], [Prab 20], [Zhao 18], [Zeng 19], [Bala 20], [El M 20], [Alha 20], [Alkh 20], [Pere 20a], [Seti 21], [Bala 21a], [Kher 21], [Vidy 21], [Cant 21a], [Bala 21b]	Soles
[Sala 04], [Das 12], [Mazi 12], [Riga 12], [Trip 13], [Pari 15], [Kuhn 17], [Cara 18], [Rava 20a], [Varr 21], [Moon 20], [Asho 20]	Multiple positions in upper and lower limbs
[Orne 17], [Moha 18], [Torv 18], [Xia 18], [Agha 20], [Orne 19], [Nagh 21]	Ankles
[Zach 15], [Ahlr 16], [Mazi 16], [Rodr 17], [Camp 18], [Sama 18], [Abra 20], [Pfis 20]	Wrist
[Kim 15], [Chom 19], [Rezv 16], [Borz 21]	Other

6.2 Kinematic Analysis of Gait

This is the first feature set considered for the experiments in this thesis, and comprises several measurements to model different properties in the strides such as time, distance, and velocity. This feature set is based on the proposed in [Bart 17]. The individual strides are segmented based on the DTW algorithm, where the gait signals are compared with a template generated from HC subjects [Bart 11]. At the same time, each stride is divided into two phases that are analyzed individually: the stance phase when the foot is on the ground and the swing phase when the foot is in the air. The toe-off and the heel-strike angles are also analyzed. Figure 6.4 shows the main kinematic aspects that are analyzed during the walking process.

Several kinematic features are computed according to the described aspects of the gait process. The feature set includes the stride, stance and swing times, the

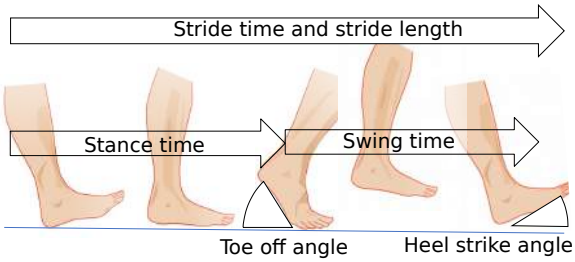

Figure 6.4: Different phases of the walking process. Source: [Oroz 20b]

stride length, the velocity of each stride, the toe-off angle, and the max clearance. The average and standard deviation are computed per each feature. In addition, these features are computed per foot (left and right) with the aim to evaluate the contra-laterality effect [Sade 00], i.e., right handed patients are more affected in the left lower limbs, while left handed patients may exhibit more impairments in the right parts of the body. The differences in the kinematic feature between PD patients and HC subjects is observed in Figures 6.5 and 6.6. Figure 6.5 shows the stride, swing, and stance duration of an HC speaker and three PD patients in mild, intermediate, and severe stages of the disease. Note how the variability of the duration within consecutive strides increases with the MDS-UPDRS-III score, particularly for the patient in severe stage (Figure 6.5d)).

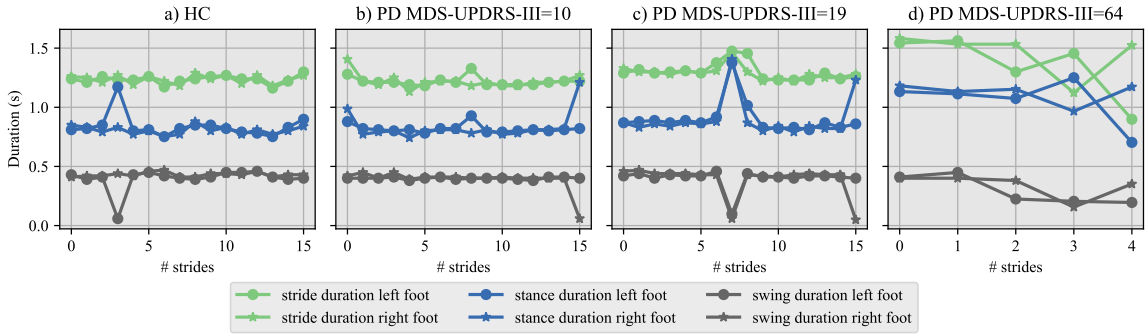

Figure 6.5: Stride duration when participants perform a 2x10 walking exercise: **a)** HC subject, **b)** PD patient in mild stage (MDS-UPDRS-III=10), **c)** PD patient in intermediate stage (MDS-UPDRS=19), **d)** PD patient in severe stage (MDS-UPDRS=64)

Figure 6.6 includes the difference between PD patients and HC subjects in other kinematic aspects of the gait process, such as the stride length, the stride velocity, and the toe off angle. Regarding the length of the strides, note that the HC subject and the PD patient with the mild stage of the disease perform longer strides than the remaining two patients. In addition, note the high variability of the stride length exhibited by the PD patient in intermediate stage of the disease in Figure 6.6c). The same differences are observed for the stride velocity (blue lines) comparing the HC subject and the three PD patients. Finally, regarding the toe-off angle, note that for the HC on the one hand the angle is centered around 0°. On the other hand, the angles for the PD patients shift towards either positive (feet looking outward),

e.g., PD patients in Figures 6.6b) and 6.6d), or towards negative values (feet looking inward) as for the case of the PD patient in Figure 6.6c).

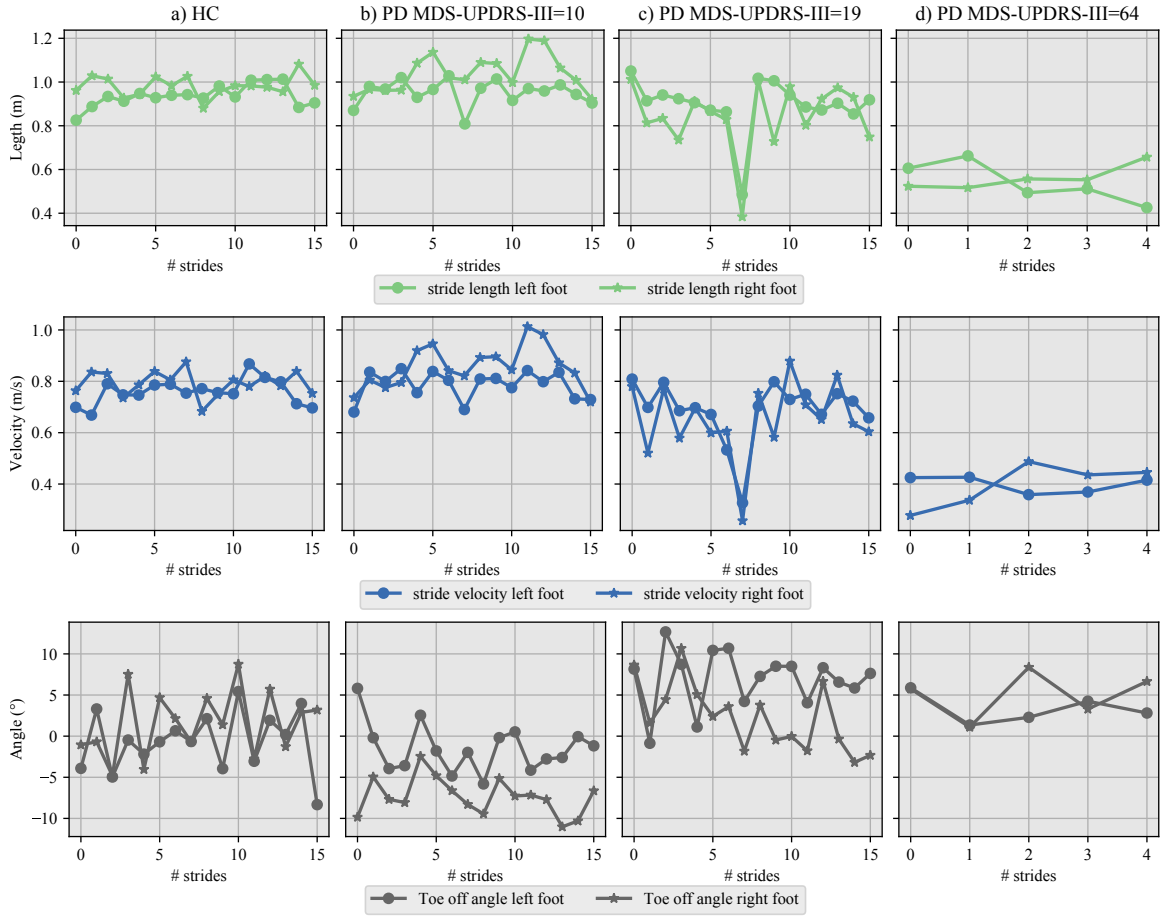

Figure 6.6: Stride length, stride velocity, and toe off angle when participants perform a 2x10 walking exercise: **a)** HC subject, **b)** PD patient in mild stage (MDS-UPDRS-III=10), **c)** PD patient in intermediate stage (MDS-UPDRS=19), **d)** PD patient in severe stage (MDS-UPDRS=64)

6.3 Spectral Analysis of Gait

This feature set is designed to model the spectral wealth and the harmonic structure of the gait signals obtained from the inertial sensors. The features are based on the CWT extracted from the accelerometer and gyroscope signals, obtained from each foot. The CWT for the gait signal $s(t)$ is defined according to Equation 6.1. $\psi(t)$ is the mother wavelet (Morlet), a is the scale factor, and b is the translational factor.

$$\mathrm{CWT}_s(a,b) = \langle s, \psi(t)_{a,b} \rangle = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(t) \psi \left(\frac{t-b}{a} \right) dt \tag{6.1}$$

The CWT is able to analyze non-stationary signals with a better resolution than the short-time Fourier transform (STFT) because the window size is adapted as a

function of the time and frequency of the signal. Lower frequencies are analyzed with longer window sizes to have a higher time-resolution. Conversely, higher frequencies are modeled with shorter window sizes to have better frequency resolution. On the contrary, for the STFT, all frequencies are analyzed with the same window size, independent on the frequency content of the signal. The magnitude of the CWT is known as the scalogram. The feature set is formed with the energy content from 8 different frequency bands from the scalogram, three spectral centroids, and the energy in the 1st, 2nd, and 3rd quartiles of the wavelet spectrum. The energy content in the locomotor band (0.5–3 Hz) is also computed, which is the frequency region where the normal gait process occurs; the energy content in the freeze band (3–8 Hz), which is related to the presence of FoG symptoms of the patients [Zach 15]; and the freeze index, which is the ratio between the energy in the locomotor and freeze bands [Zach 15, Rezv 16].

Figure 6.7 shows the difference in the spectral features computed for an HC subject and two PD patients in mild and severe stages of the disease, respectively. The figure shows the time domain gait signal from the accelerometer in the frontal plane (x-axis), the computed scalogram, the energy in the 8 frequency regions of the scalogram, and bar plots of the energy content in the locomotor and freeze bands, and the freeze index. Note that the energy content for the HC subject (Figure 6.7a)) and the PD patient in mild stage of the disease (Figure 6.7b) is centered around 1.6 Hz, which is part of the normal locomotor band. For the PD patient in severe stage of the disease, the energy is distributed across all spectrum, especially in higher frequencies, which are related with the presence of FoG symptoms. Note also that the pause performed by the subjects in the 2x10 exercise is well defined for the HC subject and the PD patient in mild stage of the disease, but not for the PD patient in severe stage of the disease in Figure 6.7c). In addition, note the difference in the energy in the motor band between the three participants: it is 10 times higher for the HC subject than for the PD patient in mild stage of the disease, and 2 times higher than for the PD patient in severe stage of the disease. Finally, note that the freeze index is lower for the HC subject than the values observed for both PD patients.

6.4 Non-linear Analysis of Gait

Gait signals are characterized as quasi-periodic time series, with autocorrelations in the stride intervals when considering walking on a long-time scale [Dier 17]. The origin of these autocorrelations may be attributed to neural central pattern generators [Haus 96], and/or to the biomechanics of walking [Gate 07, Ahn 13]. For many years, gait analysis has been studied with classical kinematic and biomechanical models in which variability was not of interest. Those traditional methods only provide estimates of the average variations within the strides, and therefore they are insufficient to characterize the local dynamic stability properties of the walking process. In order to characterize properly such underlying complexity during movement, more recent techniques derived from chaos theory and NLD are considered to model gait signals [Dier 17, Phin 20]. These novel techniques are well adapted to analyze time series with long-range autocorrelation. NLD features have shown to be also informative to evaluate the walking patterns of PD patients [Chom 19, Pere 20b].

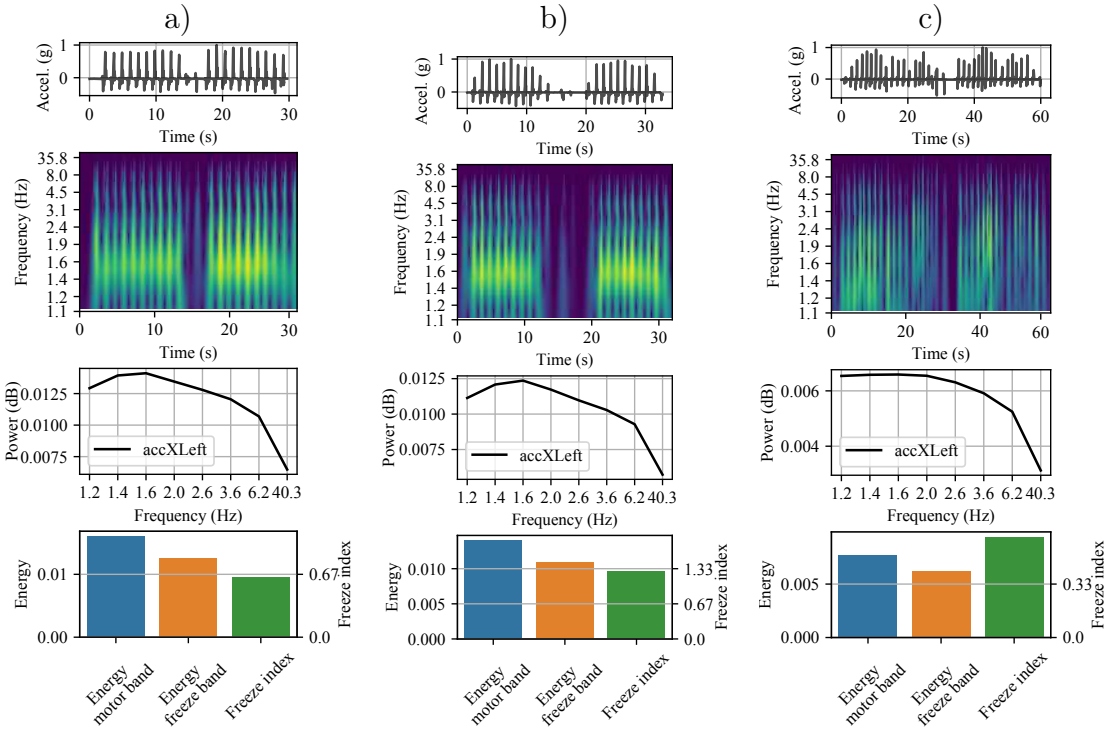

Figure 6.7: Spectral features from gait signals when participants perform a 2x10 walking exercise: **a)** HC subject, **b)** PD patient in mild stage (MDS-UPDRS-III=10), **c)** PD patient in severe stage (MDS-UPDRS=64).

The first step to extract these features is the phase space reconstruction. The process is performed following the Taken's theorem [Take 81], which indicates that there exists a dimension m such that a signal $s(t)$ is represented in a multidimensional space $\mathbf{S_t}$, known as phase space or attractor. The phase space is created using Equation 6.2, where the time-delay τ is computed by the first minimum of the mutual information function, and the embedding dimension m is found using the false neighbors method, proposed in [Kenn 92].

$$\mathbf{S_t} = \{s(t), s(t-\tau), \cdots, s(t-(m-1)\tau)\} \tag{6.2}$$

Figure 6.8 shows the phase space reconstructed from gait signals corresponding to a 10 meter walking test. The signals were captured from a gyroscope in the transverse plane (z-axis) for an HC subject and for two PD patients in mild and severe stages of the disease, respectively. The reconstructed attractor for the HC subject (Figure 6.8a) exhibits well defined trajectories with a clear recurrence. Conversely, the trajectories for the patients are more dispersed, especially when the neurological state is severe (Figure 6.8c).

Different NLD features are extracted from the reconstructed attractors to assess and compare the complexity, stability, and recurrence of the walking process performed by PD patients [Pere 20b]. The computed features extracted from those embedded attractor are briefly described as follows.

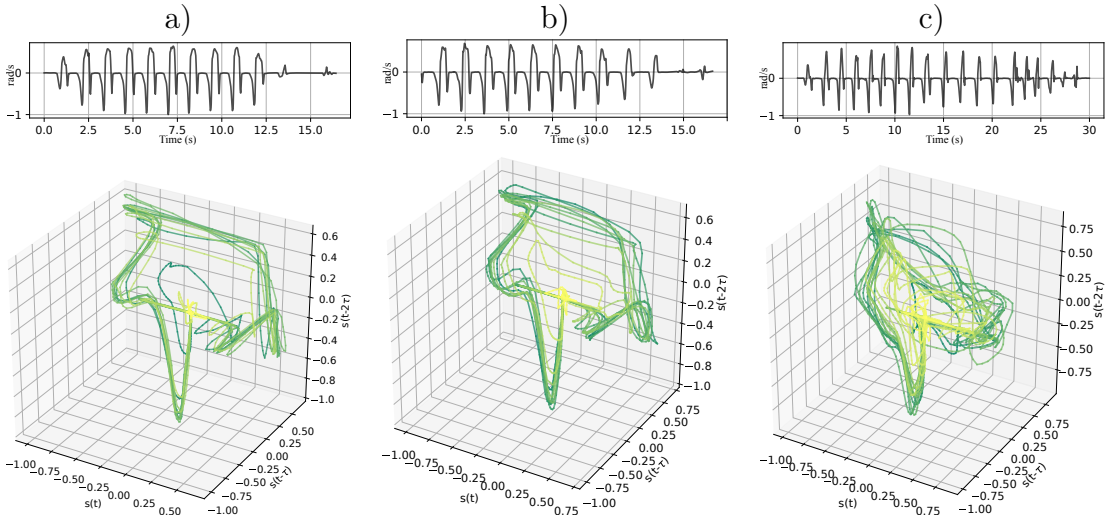

Figure 6.8: Phase space representation from gait signals captured with a gyroscope in the transverse (z) plane of **a)** HC subject, **b)** PD patient in mild stage (MDS-UPDRS-III=10), **c)** PD patient in severe stage (MDS-UPDRS=64).

Correlation Dimension

This feature measure the dimensionality of the phase space where the attractor is embedded [Gras 04]. The CD is an indicator about the complexity and dimensionality of gait signals and it is related to local instability during the walking process [Buzz 03]. The computation of CD starts with the estimation of the correlation sum $C(\varepsilon)$, using the Equation 6.3. ϑ is the Heaviside step function. $C(\varepsilon)$ can be interpreted as the probability to have pairs of points in a trajectory of the phase space inside the same hyper-sphere of radius ε.

$$C(\varepsilon) = \lim_{N\to\infty} \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \vartheta\left(\varepsilon - |s_i - s_j|\right) \tag{6.3}$$

For small values of ε, $C(\varepsilon)$ can be computed according to Equation 6.4, thus, the CD can be estimated using Equation 6.5 [Gras 04]. Then, CD is computed as the slope of the linear regression of $\log(C(\varepsilon))$ vs. $\log(\varepsilon)$.

$$C(\varepsilon) = \lim_{\varepsilon\to 0} \varepsilon^{\mathrm{CD}} \tag{6.4}$$

$$\mathrm{CD} = \lim_{\varepsilon\to 0} \frac{\log(C(\varepsilon))}{\log(\varepsilon)} \tag{6.5}$$

Largest Lyapunov Exponent

This feature measures the sensitivity of the system to changes in the initial conditions of the signal according to the rate at which the nearby trajectories of the phase space converge or diverge. The LLE gives information about the stability properties of the signal, which implies that a small perturbation introduced at any time, makes

the behavior of the signal unpredictable. The estimation of the LLE follows the algorithm introduced in [Rose 93]. After the phase space reconstruction, the nearest neighbor of every point in the attractor is located. The nearest neighbor $s_{\hat{j}}$ minimizes the euclidean distance $d_j(t)$ to the point s_j. $s_{\hat{j}}$ and s_j must be separated a distance larger than the average period of the signal to guarantee that the points are in different trajectories in the attractor. The average separation of all nearest neighbor points in the attractor at the time t is defined as $d(t) = Ce^{\lambda_1 t}$, where λ_1 is the LLE.

Hurst Exponent

This feature measures the long-term dependence of a time series. It is defined according to the asymptotic behavior of the re-scaled range of a signal as a function of time [Hurs 65]. The literature shows that patients with neurodegenerative diseases such as PD exhibit a decrease in HE compared to healthy young individuals [Jian 08], which indicates that long-term correlations of human gait from healthy young people are stronger than those from the elderly and the patients [Jian 08]. The randomness of the walking process has been observed to increase with age and by the presence of neurological disorders using the HE [Haus 97, Haus 00]. As a results, HE could be used as an indicator for gait adaptability, gait disorder, and fall risk [Phin 20]. The estimation process consists of dividing the signal into intervals of size L and calculating the average ratio between the range R and the standard deviation σ of the signal. HE is computed as the slope of the curve obtained from Equation 6.6.

$$L^{\mathrm{HE}} = \frac{R}{\sigma} \tag{6.6}$$

Lempel-Ziv Complexity (LZC)

This feature measures the degree of disorder of temporal patterns in a time series [Lemp 76]. In the computation process the signal is transformed into binary sequences according to the difference between consecutive samples, and the LZC reflects the rate of new patterns in the sequence [Trav 17]. It ranges from 0 (deterministic sequence) to 1 (random sequence). Further details of the computation can be found in [Kasp 87]. The LZC has been considered to model differences in the gait rhythm that appear due to the presence of neurodegenerative diseases [Xia 15]. The LZC also shows to be useful to quantify how the complexity of the fluctuation dynamics changes over time during walk, especially in PD patients [Kama 16]. The authors in [Kama 16] showed using the LZC that the gait signals from young people exhibits a less complex dynamical behavior than elder subjects, where the dynamical complexity increases, resulting in local instability. In addition, the authors showed that for PD patients the dynamical complexity is significantly increased resulting in increased risk of falls.

Sample Entropy (SampEn)

This feature is a regularity statistic to measure the average conditional information generated by diverging points on trajectories in the attractor. Gait signals with several repetitive patterns have smaller SampEn than signals with a complex dynamics.

This feature does not include self comparisons of points in the attractor, resulting in a more efficient computation than for the case of the Approximate entropy, which requires the comparison of all points in the phase space [Rich 00]. The SampEn has been related to the presence of bradykinesia in PD patients [Tzal 14, Hssa 19].

Detrended Fluctuation Analysis (DFA)

Similar to the HE, DFA evaluates the long-term dependency of a time-series, with the difference that DFA can be applied when the signal exhibits a non-stationary behavior. DFA is used to estimate the stochastic component of the gait process by looking at trends over time intervals in the signal. This feature has been highly applied to model the presence of non-stationarities in gait signals [Damo 10]. Particularly, the authors in [Haus 97] showed that the stride-interval fluctuations for patients with Huntington's disease are more random than for HC subjects, according to measures of the DFA. It is expected that a similar behavior will be present for PD patients. DFA has been also associated with gait stability [Herm 05], and gait velocity [Jord 07]. Details about implementation of DFA can be found in [Damo 10].

6.5 Deep Learning Models for Gait Analysis

Besides the previous models, which are based on different feature extraction strategies for a latter classification, end-to-end deep learning models are also considered both to classify PD patients and HC subjects and to evaluate the neurological state of the patients. A novel deep learning model is proposed. The neural network structure is based on one-dimensional convolutions to learn a filter bank from the raw gait signals, followed by a stack of two bidirectional GRU layers to model the temporal structure of the sequence. The proposed network includes at the end a layer with an attention mechanism with the aim to learn and give more importance to specific parts of the gait sequences e.g., the pauses, the swing phase, or the stance phase. Figure 6.9 illustrates the proposed architecture to model the gait signals of the patients.

The input corresponds to 3 second-length frames of the gait signals. The input is formed with 12 channels corresponding to the 3D-accelerometer and 3D-gyroscope attached to the left and right foot. The duration was chosen to guarantee at least 3 periods of the gait signals. The input then passes through a set of two one dimensional convolutional layers, which learn a filter-bank to process the gait sequence. The filtered signals then pass through a stack of two bidirectional GRU layers to model the temporal structure of the sequences. The last part of the network is an attention mechanism, which assigns more weights to specific parts of the gait sequence, such as pauses, the swing phase, the stance phase, or the beginning/stopping of the gait task. For the particular case of the data collected using Apkinson and due to the fact that the smartphone can be always placed in a different orientation when performing the gait tasks, a data augmentation strategy is proposed by randomly switching the axes of the inertial sensors in the input to the CNN-GRU model. With this approach a better generalization is guaranteed.

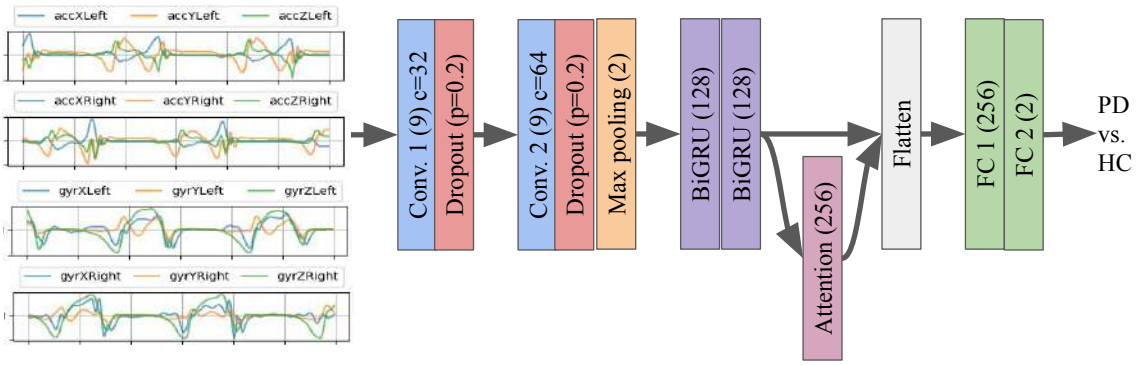

Figure 6.9: Deep learning model for end-to-end gait modeling of PD patients. **FC**: Fully connected layers. **c**= number of output channels in the convolutional layers. The values in parenthesis indicate the size of the convolutional filters and the number of neurons in the fully connected layers.

Chapter 7

Asynchronous Multimodal Analysis of Parkinson's Disease

Most psychological phenomena are complex and multifaceted, therefore requiring multimodal information to detect and to quantify them. The integration of multiple sources of data helps to get a more complete view about the neurological state of the patients. However, at the same time, the multimodal data imposes challenges in terms of information perception and data fusion strategies. Different modalities can be complementary, redundant, or even conflicting. For instance, It can be PD patients with a normal healthy handwriting, but a very impaired gait or speech.

This chapter is divided in two parts. Section 7.1 describes the most important studies that combine information from multiple modalities such as speech, handwriting, or gait to evaluate PD patients (other modalities such as medical images or electromyography are excluded from the review). Then Section 7.2 describes the fusion methods considered in this thesis to combine the information from speech, handwriting, and gait data from PD patients.

7.1 A Review on Multimodal Assessment of PD Patients

Although there are several works considering different bio-signals to assess motor impairments of PD patients, most of the studies consider only one modality. Multimodal analyses, i.e., considering information from different sensors, have not been extensively studied [Oung 15]. Although many improvements have been shown in several tasks, there is still an absence of a multimodal fusion system able to deliver an accurate prediction of the PD severity [Past 13] and to monitor the disease progression [Aria 18a].

One of the first studies that combined information from different modalities was developed in [Bart 12b]. The authors considered handwriting signals captured with a smart-pen with several gait signals collected with the first version of the eGait system. The signals were collected from 18 PD and 17 HC subjects, who performed several handwriting and gait exercises such as drawing an Archimedean spiral, heel-toe tapping, straight walking, among others. Several statistical and spectral features

were computed from the handwriting and gait signals. The features were combined following an early fusion strategy, and classified using an AdaBoost algorithm. The authors reported an accuracy of up to 97% with the proposed strategy. In [Oung 18], the authors extracted features based on energy and entropy from the empirical wavelet transform computed from speech and movement signals. Speech signals comprised utterances of sustained vowels. Movement signals were collected from inertial sensors attached to the waist, both wrist, and legs. 65 PD patients performed different movement exercises, including TUG, supination/pronation, toe tapping, among others. The authors classified the patients in four severity levels of the disease distributed according to the H&Y scale. The authors considered an extreme learning machine classifier, and reported accuracies of up to 95%; however, the approach considered by the authors is optimistic since the classifiers were optimized according to the accuracy obtained in the test set. A different approach to combine information from speech, handwriting, and gait was proposed in [Vasq 19c] by modeling the difficulty of patients to start/stop the movement of muscles in the upper and lower limbs, and in the vocal folds. The modeling considered the transitions between voiced and unvoiced segments in speech, the movement when the patient starts or stops a new stroke in handwriting, or the movement when the patient starts or stops the walking process. These transition movements were processed with time frequency representations and CNNs, trained with information from speech, handwriting, and gait. Data from 44 PD patients and 40 HC subjects were classified in two scenarios: (1) binary classification of PD and HC, which yielded an accuracy up to 97.6%, and (2) a 4-class problem classifying HC and patients in three stages of the disease, according to their MDS-UPDRS-III score, which resulted in an accuracy up to 55.6%. The best results were always obtained when information from the three modalities were combined. An additional model was proposed in [Garc 18b], where the authors computed i-vectors from features extracted from speech, handwriting, and gait signals from 49 PD patients and 41 HC subjects. The i-vectors for each modality were concatenated and classified using an SVM, which achieved an accuracy of up to 85%.

7.2 Fusion Methods for Multimodal Assessment of the Disease

Fusion of different modalities is a critical task, which can be implemented at data, feature, and decision levels. Each fusion scheme operates at a different level of analysis as illustrated in Figure 7.1 [Oung 15].

The data-level fusion gives the highest level of information details, as the signal is directly processed, but it can be highly susceptible to noise as there is an absence of pre-processing. An example of this type of fusion is the combination of accelerometer with gyroscope signals for the gait analysis, or the fusion between position and pressure for the handwriting analysis. Then, fusion at feature level, also known as early-fusion is a general type of fusion when closely-coupled modalities are combined. The typical example of this fusion strategy is when we stack together features from speech, handwriting, and gait; or features from different tasks within the same modality. This level of fusion produces a moderate level of information details, but it is

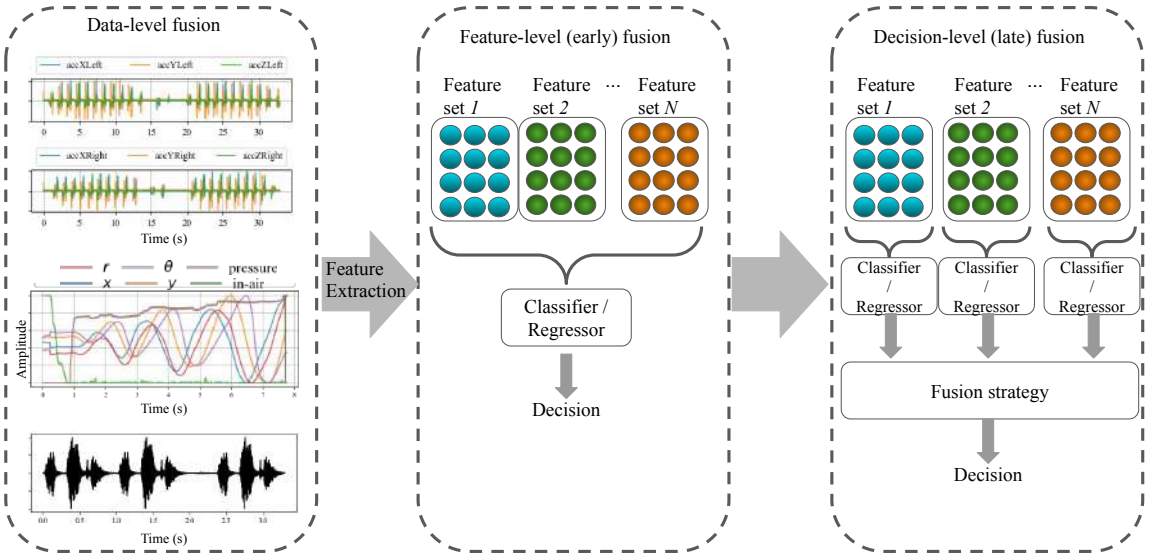

Figure 7.1: Different levels of fusion for multimodal data.

less sensitive to noise than the fusion at data level. Finally, the decision-level fusion or late fusion is the most common type of fusion in multimodal applications. The key reason is because its ability to manage loosely-coupled modalities, or when the number of feature sets to be combined is big. For instance, in the addressed problem, it is necessary to combine information from three different modalities, and for each modality the patients perform a set of different tasks. The fusion at decision level is highly robust against noise, improving accuracy and interpretation, because it is possible to check and evaluate the most important modalities and tasks in the global decision [Duma 09]. The experiments addressed in this thesis are carried out using three fusion strategies: the early fusion, were the different feature sets from different tasks and modalities are stacked together before the classification, and two late fusion strategies, which are explained as follows.

Late Fusion 1: Weighted Majority Voted Decision

This fusion strategy aims to map local predictions y_j made by the j-th classifier into a global decision Y, using a linear combination with weights $\widehat{\alpha}_j$, according to Equation 7.1, where C is the number of classifiers to combine. This approach is widely used because of its robustness, simplicity and scalability due to small computational costs [Atre 10]. It is also appropriate when there exist dependencies between the feature sets through the classifiers [Ma 12, Wu 04] e.g., when the output from different speech tasks is combined.

$$Y = \sum_{j=1}^{C} \widehat{\alpha}_j y_j \tag{7.1}$$

The goal of this fusion strategy is to find the most appropiate weights $\widehat{\alpha}_j$ to maximize the global decision score. In this thesis a method based on AdaBoost is considered. For this case, the weight associated to each classifier depends on the

errors committed by each model, following Equation 7.2, where ϵ_j is the error rate from each classifier on the training set. Hence, a classifier with an error rate of 0.5 will produce an $\widehat{\alpha}_j = 0$, thus not contributing to the global decision. Conversely, a classifier with a very low error rate will produce a very high $\widehat{\alpha}_j$, contributing the most to the global decision.

$$\widehat{\alpha}_j = \frac{1}{2}\log\left(\frac{1-\epsilon_j}{\epsilon_j}\right) \tag{7.2}$$

The previous Equation can be extended to multi-class problems, when the error rate for random guessing is not 0.5. For these cases when more than two classes is available, for instance to classify patients in different severity levels e.g, mild, intermediate, and severe, Equation 7.2 is extended to 7.3 [Hast 09], where $K > 2$ is the number of classes. The addition of the term $\log(K-1)$ is critical in the multi-class case. Now in order to $\widehat{\alpha}_j > 0$ there is only need that $(1-\epsilon_j) > 1/K$, or the accuracy of each classifier to be better than random guessing, rather than 0.5 as for the bi-class problems.

$$\widehat{\alpha}_j = \frac{1}{2}\log\left(\frac{1-\epsilon_j}{\epsilon_j}\right) + \log(K-1) \tag{7.3}$$

The estimation of the weights is also adapted to solve regression problems, following Equation 7.4. ρ_j is a performance metric for the $j-$th regressor, like the Spearman's correlation coefficient. For this case, a value of $\rho_j = 0$ will produce values for $\widehat{\alpha}_j = 0$, thus not contributing to the global decision. Conversely, positive correlations will produce high positive weights and negative correlations will produce slightly low negative weights to the global fusion.

$$\widehat{\alpha}_j = \frac{1}{2}\log\left(\frac{1}{1-\rho_j}\right) \tag{7.4}$$

Late Fusion 2: Dynamic Score Combination

In the previous fusion method, the computed weights for the fusion are *static* i.e., they are computed once based on the performance from the training set and they are equally applied to all samples in the test set. Conversely, in the dynamic score combination (DSC) methods the computed weights for the fusion are not fixed for all samples, but they change according to the sample to be classified [Tron 09]. The DSC strategy estimates dynamically a set of weights for each sample to be classified, thus Equation 7.1 changes to Equation 7.5 for the DSC, for the i-th sample in the test set. $s_{i,j}$ is the score obtained from the j-th classifier for the i-th sample in the test set e.g., the distance to the hyper-plane for the case of SVMs or the output of the Softmax activation for the case of DNNs.

$$Y_i = \sum_{j=1}^{C} \widehat{\alpha}_{i,j} s_{i,j} \tag{7.5}$$

Ideally the weights $\widehat{\alpha}_{i,j}$ are adapted both to each sample in the test set, and to the performance of the local classifiers for each sample, instead of having a set of weights

$\widehat{\alpha}_j$ based on the global performance of each classifier on the training set [Tron 09]. The formulation of the DSC establishes that if the scores for each classifier $s_{i,1} \leq s_{i,2} \leq \cdots \leq s_{i,C}$ are sorted, Equation 7.5 is equivalent to Equation 7.6.

$$Y_i = \widehat{\beta}_{i,1} \min_j(s_{i,j}) + \widehat{\beta}_{i,2} \max_j(s_{i,j}) \tag{7.6}$$

If the constrain that the weights should sum to one is included, then Equation 7.6 can be rewritten as Equation 7.7 [Tron 09]. Hence, $\widehat{\beta}_i$ will be an adaptive score, estimated for each particular sample of the test set.

$$Y_i = (1 - \widehat{\beta}_i) \min_j(s_{i,j}) + \widehat{\beta}_i \max_j(s_{i,j}) \tag{7.7}$$

Now, combining Equations 7.5 and 7.7, the value of $\widehat{\beta}_i$ is dynamically estimated for each sample from the test set, according to Equation 7.8 [Tron 09]. This is known as supervised DSC, because there is still necessary to estimate the weights $\widehat{\alpha}_j$ for each individual classifier. Equation 7.2 can be used to estimate the weights $\widehat{\alpha}_j$ for each classifier.

$$\widehat{\beta}_i = \frac{\sum_{j=1}^{C} \widehat{\alpha}_j s_{i,j} - \min_j(s_{i,j})}{\max_j(s_{i,j}) - \min_j(s_{i,j})} \tag{7.8}$$

The main advantage of the supervised DSC is that it is still possible to map the importance of each classifier to the final decision based on its performance on the training set, but adapting the scores individually to each sample of the test set by the computation of $\widehat{\beta}_i$. This is the approach considered in this thesis for the fusion of the different feature sets from the different modalities and from the different tasks performed by the participants in each modality.

Chapter 8

Analysis of Parkinson's Disease using Smartphones

Motor symptoms observed in PD patients progress differently among patients, thus it is important to monitor their symptoms individually and continuously. The continuous monitoring is not always possible for many patients, especially those with low accessibility to healthcare services. Hence, there is a need for a system to track the disease progression of the patients individually. A smartphone application that combines speech and movement analysis could be a suitable mechanism to monitor the disease progression. Such an application will be beneficial for patients and caregivers to be informed about the current stage of the disease; and for clinicians to make timely decisions regarding the medication and therapy of the patients.

This chapter is focused on the analysis and evaluation of PD patients using smartphone technologies. Section 8.1 describes existing mobile applications that are designed to evaluate different motor symptoms of PD patients. Then, Section 8.2 describes the *Apkinson* app, which was designed and developed in cooperation between researchers from the University of Antioquia (Medellín, Colombia) and the Friedrich-Alexander University Erlangen-Nuremberg (Erlangen, Germany).

8.1 A Review on Automatic Assessment of Parkinson's Disease using Smartphones

In the past years, several smartphone applications were developed to monitor the symptoms of PD patients [Post 19]. However, most of them only consider the evaluation of the upper and lower limbs using the inertial sensors embedded in the smartphone [Post 19, Stam 18]. Additionally, many of the existing applications designed for PD patients are focused only on the evaluation of specific aspects of the disease such as postural tremor [Frai 16], bradykinesia [Prin 14], fine motor skills [Lee 16], and FoG [Mazi 12, Kim 15, Cape 16].

One of the first studies to evaluate PD symptoms using smartphones was developed in [Mazi 12], where the authors proposed the use of a smartphone application and wearable accelerometers to detect FoG episodes during walking. Several accelerometers were placed in the ankle, knee and hip of 10 PD patients, who performed walking

simulating activities of daily living. Signals captured with the accelerometers were transmitted to a smartphone application via Bluetooth. The app computed statistical and spectral features from the collected gait signals. The extracted features where used to classify FoG episodes vs. normal walking. The authors reported an accuracy up to 95%. In [Grac 14] the authors developed an Android application to evaluate automatically three aspects of PD patients: (1) handwriting with the drawing of an Archimedean spiral, (2) finger tapping, and (3) gait. Different features were computed for each task, e.g., geometric features from the Archimedean spiral and kinematic features from the tapping and gait exercises. The extracted features were used to classify 17 PD patients and 18 HC subjects, using a Bayesian network that achieved an accuracy of up to 87.5%. In [Kim 15] the authors used the inertial sensors from smartphones to detect FoG events from 15 PD patients. Subjects were instructed to walk straight for 3 meters. Then, they turned around and returned to the starting place. The smartphone was placed in several parts of the body when the patients performed the exercises. The authors computed the FI and other statistical and spectral features to classify the FoG episodes using an Adaboost classifier. The highest accuracy was reported when the smartphone was placed in the waist (86%). In [Kost 15], the authors presented a smartphone application to capture data from postural tremor of PD patients. The authors extracted kinematic features such as the average and standard deviation of the acceleration signals. The features were used to classify a group of 25 PD patients and 20 HC subjects. The authors reported an AUC of up to 0.91 with an Adaboost classifier. The authors from [Cape 16] created a smartphone application to detect FoG episodes in real time. the proposed method consisted of the computation of the FI, the Energy index (EI), and the step cadence. EI is defined as the sum of the energies in the freeze and locomotor bands. The step cadence was computed based on the amplitude of the second harmonic of the power spectrum. The three features were computed for frames of 0.4 seconds length, and the FoG episodes were detected based on a set of rules and thresholds for EI and FI. The proposed model was tested with data from 20 PD patients that experienced FoG episodes when they performed a TUG test. The FoG episodes were detected with an AUC of up to 0.90. In [Agha 17] the authors performed a handwriting assessment using a smartphone application when patients draw a spiral. The authors computed several features including the kurtosis of the speed stroke, the length of the spiral drawing curve, the area of the spiral in each loop and the time of the drawing. The authors evaluated different items of the UPDRS scale related to the upper limbs, and reported correlations ranging from 0.47 to 0.52 combining handwriting features with finger-tapping measures. In [Stam 18] the authors developed the CloudUPDRS app to predict the UPDRS score of PD patients when they perform a set of movement exercises at home. The application includes a set of 17 exercises to evaluate postural, kinematic and resting tremor, finger tapping, and gait. The exercises are used to classify three levels of the UPDRS score of 12 PD patients recorded in several sessions during three months, using a deep learning approach. The authors reported an accuracy of 78%. However, it is not clear whether the results obtained are subject independent. In addition, no suitable feedback is shown to the patients in the app, but in a web application. This fact may affect the usability and interest of the patients to use the app. The authors from [Iako 19] collected data from the touchscreen keyboard

of smartphones using the i-Prognosis application[1]. The touchscreen patterns of 27 PD patients and 84 HC subjects were classified with a CNN to process the press and release time sequences of the keyboard. The authors reported an AUC of up to 0.78.

There are few applications to evaluate the speech symptoms of PD patients [Bot 16a, Zhan 18]. However, the analysis only considers the phonation of sustained vowels. In the *mPower* for iPhones [Bot 16a], the patients respond to a subset of questions from the MDS-UPDRS scale, and perform short activities such as finger tapping or the phonation of the sustained vowel /a/. In [Zhan 18] the authors introduced the HopkinsPD smartphone application, where the patients have to perform 5 exercises related to speech, finger tapping, gait, balance, and reaction time. The authors proposed the Parkinson's disease score (mPDS) based on the performed exercises by the participants. The mPDS aimed to detect intra-day symptoms fluctuation, even those related to the dopaminergic medication. The authors correlate the mPDS and the MDS-UPDRS-III scores of the patients, and report Pearson's correlations of up to 0.88; however, such correlation could be optimistic since the comparison is performed considering only three time periods in the longitudinal analysis. Additional studies also have shown that it is possible to evaluate the speech impairments of PD patients using signals captured with smartphones [Zhan 17, Aria 18b, Rusz 18a, Zhan 19]. However, such studies only consider the smartphone to record the speech data, without providing a feedback mechanism to the patient about their current state of the disease. A summary of the existing apps for the assessment of PD until February 2021 is available in Table 8.1. Note that there is no application specifically designed for speech assessment of PD patients. There are also apps designed only to collect data, which does not provide feedback to patients.

8.2 Apkinson

The need to monitor continuously the disease progression of PD patients, and to make a motor evaluation of the patients that includes specifically the speech production yields to the development of *Apkinson* [Oroz 20a], a mobile application designed with the aim to provide patients, caregivers and clinicians with a technological tool that supports them in the process of following the disease progression. Apkinson is an android application that records several signals using sensors embedded on the smartphone (microphone, accelerometer gyroscope, and the touch screen) and performs different analyses with the aim to model the disease progression of PD patients. The app incorporates exercises and models for speech, walking, hand movements and finger tapping, and the patient receives immediate and individual feedback with the results of the exercises. The individual feedback to the patients motivates them to continue using the app and trying to perform better every day. Apkinson is available to be downloaded by patients in the Google Playstore[2]. The source code is also available for the research community interested in performing updates or adaptations to other neurological or speech-related diseases[3].

[1] http://www.i-prognosis.eu/

[2] Apkinson: https://play.google.com/store/apps/details?id=com.sma2.apkinson

[3] Source code of Apkinson: https://github.com/jcvasquezc/SMA2

Table 8.1: List of existing mobile applications for assessment of PD patients

Name	Platform	Exercises	Evaluation	link	Cost (€)
PD Warrior	Android	Movement exercises	No feedback	https://pdwarrior.com/	Freemium
ARAT	Android	Exercises for upper limbs	Self evaluation		Free
MyTremorApp	Android	Postural tremor, balance, finger to nose, pronation/supination	Report of tremor and bradykinesia	https://medapplets.com/mytremor-app/	Free
Swallow Prompt	Android, iOS	Swallowing therapy	No feedback	https://speechtools.co/swallow-prompt	1.99
ListenMee	Android	Gait	Auditory feedback to improve gait		30
myParkinson's	Android	Postural tremor	No feedback		Free
Neurofit	Android	Movement exercises	No feedback	http://albertosanchez.net/neurofit.html	Free
AppTUG	Android, iOS	TUG test	No feedback	https://www.mon4t.com/	Free
StudyMyTremor	iOS	Postural tremor	Frequency and power of the postural tremor	http://studymyhealth.com/funktionen/studymytremor/	4.49
PD Me Tools	iOS	Balance, memory, reaction, time perception	Indicators for each of the four exercises	http://bellesfarm.com/	Free
TUG App	Android, iOS	TUG test	time to perform the test		Free
Tippy Tap	iOS	Finger tapping	Tapping scores		Free
mPower2	iOS	Finger tapping, gait, postural tremor, sustained phonation	History of performed exercises	https://parkinsonmpower.org/your-story	Free
MAP in PD	iOS	Finger tapping, balance	Self evaluation		Free
Motion in PD	iOS	Sustained phonation, finger tapping, balance read text	Self evaluation		Free
PD LifeKit	iOS	Finger tapping, cognition, sustained phonation, memory, postural tremor, singing	Global and individual indicators per task	http://connectedneuro.com/	14.99
Voice Analyst	Android, iOS	speech (not clearly defined)	Pitch and volume		10
DigiTap	iOS	Finger tapping	number of tapps	http://www.app-store.es/digitap	2.99
Lift Pulse	iOS	Postural tremor	Frequency of the tremor	http://www.app-store.es/lift-pulse	Free
Tremor12	iOS	Postural tremor	No feedback		Free
cloudUPDRS	Android	Movement of upper limbs, gait, finger tapping	No feedback	http://www.updrs.net/	Free
iPrognosis	Android, iOS	Usage interaction with the smartphone	Interaction of the patient with the smartphone	http://www.i-prognosis.eu/	Free

The app was developed within the framework of the project *Speech and Movement Analysis using your SMArtphone for neurological diseases (SMA)*[2], which was financed by the Ministry of Education and Research in Germany (BMBF). Different researchers from the Pattern Recognition Lab from the University of Erlangen-Nuremberg (Erlangen, Germany), the Machine Learning and Data Analytics Lab also from the University of Erlangen-Nuremberg and the GITA research Lab from the University of Antioquia (Medellín, Colombia) participated in the development of

the app during two *Sprints* within two years, following a SCRUM methdology. The list of contributors to the app is available online[4].

There is a set of 38 exercises included in Apkinson, and the patient is requested to do between six and eight of them every day (in its current version the set of exercises is repeated every week). The set includes tasks of different nature like speech production, hand movement, gait and finger tapping. The information from those signals is stored and processed on the phone, allowing Apkinson to compare the results with previous recording sessions, providing the patient with a direct and individual comparison. Those exercises that require more computation power due to more elaborated and complex algorithms are sent to a server. Specifically, the evaluation of pronunciation and intelligibility of speech need to be computed on the server side. The first one requires the use of Phonet (see Section 4.3) to compute phonological posteriors, and the second one is based on an ASR. Once all of the computations are performed, the result is sent back to Apkinson and included in the feedback that is provided to the patient.

When the patient uses Apkinson for the first time, it requests the patient to introduce several metadata including the birth date, gender, dominant hand, years of education, years of diagnosis, medication name, dose and intake time, weight, height and others. After the patient uses Apkinson for several sessions, s(he) can see the progress in the performed exercises, and move to the different modules of the app, including access to exercises, results and other settings. In the main screen, the patient can start to do the exercise session of the day. Figure 8.1a) shows these options. The name *Camilo* is a reference that corresponds to the name of the patient. Apart from the registration and the metadata information, Apkinson incorporates a settings module where the patient can manage general aspects of the app like updates of the demographics or medication information, or to change the time of the notifications to remind the patients to do their daily exercises. In addition, when the patient attends a medical appointment, Apkinson allows the medical examiner to export the information from the patients' smartphone, and also to update exercises that the patient has to perform. A screenshot of the settings module is shown in Figure 8.1b).

When the patients press the *QUICK START* (see Figure 8.1a)), they are moved to the module with the exercises that should be performed on a daily basis. The patient will receive a daily notification as a reminder to do the exercises. There are three groups of exercises, the first group has a total of 21 speech tasks including the sustained phonation of the vowels /a/, /i/ and /u/, six DDK exercises, ten different sentences that the patient has to read, and the description of images that appear on the screen. The speech tasks are thought to evaluate phonation, articulation and prosody impairments in the speech of the patients. Additional information about the data collection using Apkinson was described in Section 3.3.4. The other two groups of exercises contain a total of 17 tasks that are captured using the inertial sensors of the smartphone. The aim is to evaluate different abnormal aspects in movements including postural tremor, kinetic tremor, finger tapping, gait deficits, among others. The patient can access the instructions via video, voice and text on the App. Those instructions guide the patient to perform the exercises correctly.

[4] Apkinson contributors `https://github.com/jcvasquezc/SMA2/graphs/contributors`

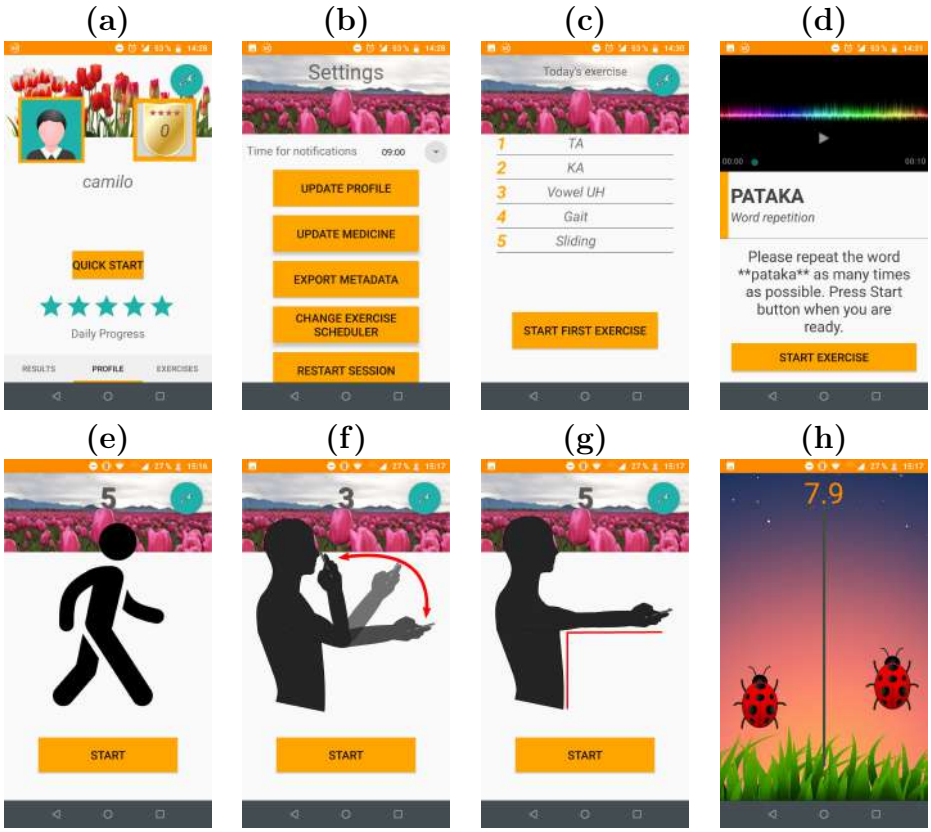

Figure 8.1: Different screens from Apkinson to monitor the disease progression of PD patients.

Figure 8.1c) indicates a list of exercises to perform in the current session; Figure 8.1d) shows the instructions (text and audio) to explain the patient how to perform a DDK exercise. Figure 8.1e) shows the screen that the patient can see when is about to start one of the walking tasks. Note that the patient is asked to put the smartphone in the pocket before starting the walking tasks. Figure 8.1f) shows the example that the patient sees when doing the finger-to-nose test. Figure 8.1g) shows the example of how to do the hand tremor exercise; and Figure 8.1h) shows the screen of the task where the patient has to touch the lady-bug alternating between the two finger thumbs. For each exercise, the patient can see a video with the example and can also read the instruction that is written below the video screen.

One relevant aspect when using sensors embedded in smartphones with an Android operating system is the sampling frequency. The smartphone is programmed to do the sampling when the scheduler allows it, which means it depends on whether the user or the device itself is running other applications at the same time and the sampling frequency can oscillate in a range between 20 and 200 Hz. There is no way to avoid that situation because it is programmed like that in the Android operating system. With the aim to avoid that potential problem, we implemented a resampling method that is always applied such that it assures a constant sampling frequency of 100 Hz. The method is standard and it is based on a linear interpolation.

8.2.1 Speech Assessment

For the evaluation of speech, Apkinson focuses on the analysis of stability, speech rate, intonation, intelligibility and pronunciation. The first three are computed directly on the phone and the last two are computed on the server. The sampling frequency for the speech signals in Apkinson is set at 16 kHz and it is important to mention that speech signals do not have stability problems in the sampling frequency. Once it is set, the value is constant for every recording. The *stability* of the vocal folds' vibration is measured by computing Jitter from the sustained phonation of the vowel /ah/, following the same procedure described in Section 4.2.1. The *speech rate* is considered to evaluate the speed of the articulation movements necessary to produce DDK exercises. The speech rate is computed considering the number of voiced sounds per second produced by one speaker during the rapid repetition of /pa-ta-ka/. The method used in Apkinson to identify the voiced frames is based on the presence of pitch in short-time speech frames of 40 ms extracted with a time-shift of 10 ms. The *intonation* in Apkinson is measured as the standard deviation of the pitch contour extracted from the longest sentence included in the speech protocol. The *pronunciation* is evaluated with the Phonet toolkit, described previously (see Section 4.3) by computing phonological posteriors of stop consonants (/p-t-k/) when the patients perform the DDK exercises. Finally *intelligibility* is measured based on the word error rate computed between the read sentences and the recognized ones using a pre-trained ASR system. The ASR was trained using the Kaldi framework, using a general GMM-HMM architecture, and it is installed on the web server. The model was trained with the ten sentences included in the protocol, read by 103 healthy participants. Each sentence was repeated ten-times, which gives us 10,300 recordings, for an approximate of 10 h duration of recordings to train the ASR.

8.2.2 Movement Assessment

The evaluation of the motor symptoms of the patients in the upper and lower limbs is performed in Apkinson with measures of: (1) regularity of movements, (2) FoG, (3) hand tremor, postural stability, and gait dynamics. All these features are computed locally on the phone.

Movement Regularity

Apkinson includes several movement exercises where the patients perform repetitive patterns that form quasi-periodic signals. These exercises are inspired on the MDS-UPDRS-III scale [Goet 08] and include the finger-to-nose test, the pronation-supination test, and the arm-circles exercise. Apkinson evaluates the regularity in the repetition of these exercises according to the temporal variability (TV) of the fundamental period of the acceleration signals in the z-axis. TV is measured according to the standard deviation of the fundamental period of the signal computed for windows of 400 ms length with a time-shift of 20 ms. The value of TV is normalized according to a sigmoid factor to get a regularity index (RI) score between 0 and 100% (see Equation 8.1). With the normalized score, a person with very regular movements

will get a regularity measure near to 100%, while a patient with irregular movements will get lower scores.

$$\mathrm{RI} = \frac{200}{1 + e^{2\mathrm{TV}}} \tag{8.1}$$

FoG

FoG is one of the most debilitating motor symptoms in advanced stages of PD, as it was described in Section 6. The computation of the FI [Kim 15] was included in Apkinson to measure how the patient is affected by FoG. The FI is computed in the gait exercises included in Apkinson.

Postural Stability

Posture stability is a common problem in PD patients and one of the main causes of falls. The postural stability is evaluated in Apkinson in the standing task, where the patient should be standing straight for 30 s with the smartphone in his/her pocket. The postural stability is measured based on the energy of the acceleration signals in the three axes, according to Equation 8.2, where a_x, a_y and a_z correspond to the acceleration measured in frontal, sagittal and transversal planes, and N is the length of the gait signal. The value of the energy is normalized according to a sigmoid factor to get a postural stability index (PSI) between 0 and 100%, following Equation 8.3. With the normalized score, a person with small movements will get a PSI near to 100%, while a patient with strong movements will get a lower score.

$$E_p = \frac{1}{N} \sum (a_x^2 + a_y^2 + a_z^2) \tag{8.2}$$

$$\mathrm{PSI} = \frac{200}{1 + e^{2E_p}} \tag{8.3}$$

Hand Tremor

Hand tremor in Apkinson is evaluated in the *postural tremor exercise*, where the patient extends the arm holding the phone, keeping such a position for at least 10 s. Apkinson computes the energy of the acceleration signals when the patient is holding the phone, using the same strategy considered for the postural stability in order to get a tremor performance between 0 and 100%.

Gait Dynamics

Apkinson has incorporated a step detection algorithm based on a peak detection method of the acceleration signals. With the number of detected steps and their location in the acceleration signal, we computed the duration of each step. The number of steps and the average duration are also included in the report section of Apkinson to show the patient their current performance when they perform the free walking test.

8.2.3 Fine Motor Assessment

Fine motor tasks aim to evaluate different dimensions of PD such as akinesia (inability to initiate movement), bradykinesia (slow movements), freezing (momentary loss of movement), deficit in space-visual ability, and loss of cognitive ability. Apkinson includes three finger tapping exercises based on those proposed in [Tava 05] to evaluate the patient performance in fine movements. The first one consists of tapping with the thumb of the dominant hand ladybugs that randomly appear on the screen for 10 s. The second one is a two-finger tapping test where the patient uses both thumbs to hit two ladybugs that appear randomly on the screen (see Figure 8.1h) The third task is to slide horizontally a bar until reaching a target point, which moves randomly every time once it is reached. This third task is inspired in the Fitt's test to evaluate human computer interaction systems [Fitt 54]. Each exercise requires rapid reaction, concentration, ability to associate, spatial location and repeated movements of extension and contraction of the fingers. The evaluation of the fine movement skills of the patients is performed with four features: (1) tapping accuracy, which is the number of lady-bugs the patient manages to capture during the time of the exercise, relative to the number of times the patient touches the screen. (2) Tapping velocity is computed as the number of taps performed, relative to the duration of the tapping exercise. (3) Tapping precision measures the distance between the point in the screen pressed by the patient and the real place of the lady bugs in the tapping exercises. Finally (4), the sliding velocity measures the number of times the patient is able to reach the target bar during the time of the sliding exercise.

8.2.4 Feedback to Patients

Patients can see their performance after doing the exercises, and also compare their results with respect to previous sessions. Figure 8.2 shows the different screens that the patient can visualize to get feedback about the current state. Figure 8.2a) shows the results obtained from the speech exercises. The five vertices of the radar plot correspond to the evaluation of stability, speech rate, intonation, intelligibility and pronunciation. The general speech performance is obtained from the area of the resulting pentagon. Figure 8.2b) indicates results obtained from movement exercises and the six vertices of the plot correspond to the computed features: regularity of movements, hand tremor tremor, average duration of the strides, number of steps, postural stability and the FI. As in the previous case, the area of the resulting hexagon is computed as a general performance for movement analysis. Figure 8.2c) shows the evaluation of fine motor skills including the tapping accuracy, tapping velocity, the tapping precision, and the sliding velocity. The general fine motor skill indicator is computed as the area of the resulting quadrangular. For all cases, the reference plot in cyan color is computed as the result of evaluating 60 HC subjects. The plot computed for the patient is in orange and when there is overlap between the reference subjects and the patient, the resulting plot is in light green. The global motor evaluation considers the areas of the geometrical figures obtained from each evaluation separately (speech, movement and fine motor). The result is composed by three area values that are used as the vertices for a triangle. As in the case of the individual evaluations, there is a reference plot computed with the results of 60 HC

subjects (see Figure 8.2d)). Finally, Figure 8.2e) indicates how the historic results are displayed for the follow-up evaluation of the patient.

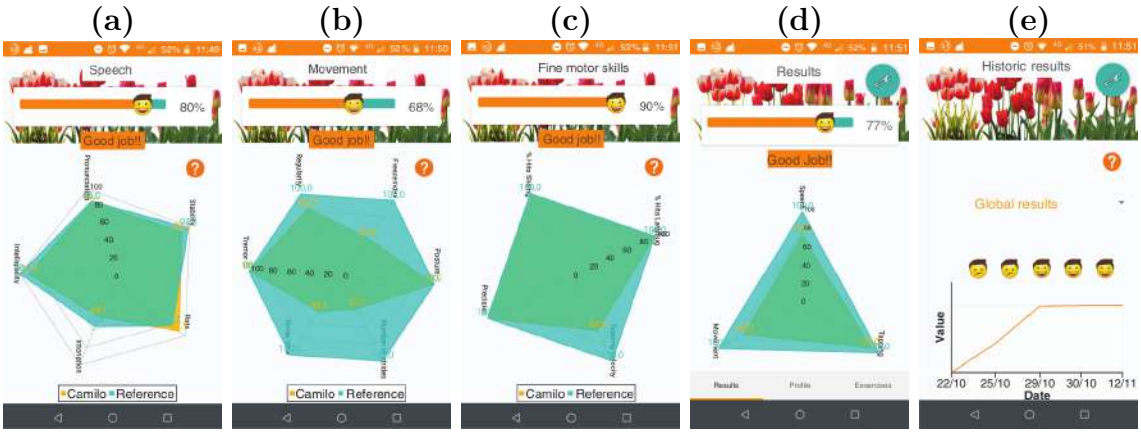

Figure 8.2: Different screens from Apkinson to show results and give feedback to the patients.

8.2.5 Communication between Apkinson and the Server

The communication between Apkinson and a server is guaranteed for different applications. On the one hand to perform the most complex speech analyses that cannot be computed locally on the smartphone and that include the ASR system trained on Kaldi to compute the intelligibility metrics, and the phonological analysis module to compute the pronunciation scores. On the other hand, to receive and store properly the data collected from the application and that include metadata and medication information of the patients, multimedia files with the speech and movement recordings, and results of the exercises performed by the patients. Our aim is that potential users like patients, caregivers, and medical doctors can access the data of the patients to track the evolution of the neurological state of the patients. This option is still under development and will be configurable according to the Ethical regulations of each country or institution. The data collection using the web server is right now available only for patients in Colombia because of privacy and ethical requirements. It is important to mention that the app senses the WiFi connection of the smartphone and files are only sent to the server when there is WiFi connection. This is with the aim to avoid consuming all of the mobile data of the user, which could potentially demotivate to use Apkinson.

8.2.6 Limitations of Apkinson

The current version of Apkinson is able to handle audio and movement resulting files from the exercises. Additionally, the system is able to compute the complex speech indicators on the server side. Although all features of the system are working properly, the current version of Apkinson does not include different profiles (patient, caregiver or clinician). Future work includes developing graphic user interfaces for caregivers and clinicians, such that they can remotely monitor the progress of the patient while

using the app. Another limitation of the app is that it only works on Android devices, thus we encourage the research community to work on the adaptation of Apkinson to make it usable on iOS devices as well.

Chapter 9

Experiments & Results

This chapter includes details of the experiments addressed during the development of this thesis. It is divided into five main sections. Section 9.1 is dedicated to the assessment of PD from speech, including the analysis of the high quality speech data from the Multimodal corpus, and the assessment of the speech collected from the smartphones using the Apkinson app, described in Section 8.2. The analysis is performed using the methods described in Chapter 4 for traditional feature analysis and deep learning techniques. Section 9.2 includes experiments and results about handwriting assessment of PD patients. The assessment is performed using all methods described in Chapter 5 to model the kinematic, geometric, and in-air aspects of online handwriting, as well as the deep learning methods to model both online and offline handwriting. Section 9.3 includes experiments and results about gait assessment of PD patients. The assessment is performed using all methods described in Chapter 6 to model the kinematic, spectral, and non-linear aspects of gait signals, as well as the deep learning methods. The experiments are performed with the data collected from both the high-quality eGait sensors and with those embedded in the smartphones via the Apkinson app. Section 9.4 shows the results obtained combining speech, handwriting, and gait analysis, described in Chapter 7. Finally, Section 9.5 summarizes and discusses the analysis of the most important results. The validation strategy for the results presented in this chapter follows the methodology described in Section 2.3.

9.1 Speech Assessment

The automatic assessment of the speech of PD patients covers three scenarios: (1) the automatic classification of PD patients vs. HC subjects, (2) the evaluation of the dysarthria severity of patients based on the m-FDA scale, and (3) the assessment of the motor-state severity of patients based on the MDS-UPDRS-III scale. The evaluation of the dysarthria severity includes also the longitudinal evaluation of the patients from the longitudinal and the At-Home corpus to evaluate the progress of the speech impairments in the patients. The speech assessment is addressed considering different modeling strategies. The focus is always to analyze different speech dimensions, speech tasks, and their different combinations. The speech dimensions include the feature sets to model phonation, articulation, prosody, phonological, among other speech aspects, in addition to the end-to-end modeling using deep learning techniques.

The speech tasks include read sentences, DDK exercises, read texts and monologues pronounced by the participants. The fusion of feature sets and speech tasks is performed with the three methods described in Section 7.2: (1) early fusion (referred as *Early*), (2) weighted majority voted decision (referred as *Late 1*), and dynamic score combination (referred as *Late 2*).

9.1.1 Automatic Classification of Parkinson's Disease Patients

The results classifying PD patients vs. HC subjects from the Multimodal corpus (Section 3.3.1) are presented in Table 9.1 using the different feature sets and speech tasks. The feature sets include phonation, articulation, prosody, phonological, OpenSMILE, those obtained from the RAE, and their combinations. The results are presented in terms of the unweighted average recall (UAR), which is used to avoid bias due to the unbalance in the groups and it can be interpreted as an average ratio of the true positives per class. The OpenSMILE features are included only as a baseline, and they are not included in the fusion strategies because the computational cost that they represent, and because the lack of clinical interpretability of the feature set. Results of the best model per task are highlighted in bold.

Table 9.1: Results classifying PD patients vs. HC subjects from the Multimodal corpus using different speech feature sets and SVM classifiers. Results in terms of unweighted average recall (UAR [%]).

	Feature sets						F. F.	F. F.	F. F.
Task	**Phonation**	**Articulation**	**Prosody**	**OpenSMILE**	**Phonological**	**RAE**	**Early**	**Late 1**	**Late 2**
DDK1	71.5 (9.6)	76.3 (12.1)	70.0 (8.0)	79.8 (6.3)	75.8 (7.4)	74.1 (7.6)	**83.5 (7.0)**	77.4	80.8
DDK2	69.2 (9.1)	74.5 (12.2)	68.0 (11.4)	82.3 (6.9)	78.3 (8.9)	75.3 (5.8)	**85.6 (7.1)**	78.5	82.8
DDK3	71.0 (10.2)	74.5 (6.6)	69.1 (12.1)	77.3 (11.1)	79.5 (9.1)	74.5 (7.7)	**80.8 (10.5)**	77.7	80.0
DDK4	64.3 (8.8)	70.9 (5.5)	64.1 (5.9)	**77.7 (8.5)**	75.8 (9.9)	66.8 (8.9)	75.5 (6.9)	75.1	76.5
DDK5	64.3 (9.1)	68.6 (4.8)	69.4 (11.1)	79.3 (9.6)	75.9 (8.7)	66.6 (12.3)	**80.5 (7.8)**	73.9	77.9
DDK6	68.1 (6.8)	76.1 (4.7)	67.3 (14.7)	**82.3 (8.1)**	75.3 (7.8)	64.2 (9.0)	78.8 (7.1)	77.5	78.8
Sentence 1	62.2 (6.3)	74.6 (10.4)	65.4 (12.1)	79.6 (6.5)	74.3 (6.6)	62.5 (7.9)	**79.6 (5.5)**	79.2	79.3
Sentence 2	66.3 (3.5)	75.1 (7.6)	67.6 (6.8)	82.0 (6.3)	73.8 (9.6)	68.1 (6.0)	**82.2 (7.9)**	78.2	81.2
Sentence 3	63.4 (5.9)	75.5 (4.1)	61.1 (9.6)	75.7 (8.9)	76.3 (8.4)	63.6 (11.4)	**78.6 (5.7)**	76.6	79.7
Sentence 4	70.7 (11.3)	74.8 (5.7)	71.6 (6.9)	83.3 (5.6)	78.1 (8.8)	68.4 (10.3)	84.3 (6.6)	82.4	84.4
Sentence 5	64.5 (6.9)	72.1 (6.6)	70.8 (8.2)	77.1 (7.7)	78.3 (5.9)	67.3 (8.7)	**79.9 (5.1)**	78.0	77.6
Sentence 6	69.9 (5.1)	76.4 (6.3)	71.4 (9.0)	76.9 (8.4)	**79.8 (9.8)**	74.4 (7.3)	77.8 (7.0)	78.9	78.0
Sentence 7	67.9 (6.9)	79.8 (6.8)	71.2 (8.9)	77.9 (7.5)	**81.6 (11.6)**	70.9 (5.3)	78.6 (6.6)	80.4	79.6
Sentence 8	69.4 (9.4)	73.4 (10.0)	68.3 (6.8)	79.0 (8.4)	74.9 (10.7)	70.8 (5.1)	78.4 (4.4)	**80.3**	80.0
Sentence 9	74.1 (4.8)	79.7 (8.4)	76.4 (4.7)	82.3 (9.8)	76.8 (10.5)	71.7 (7.0)	78.0 (9.8)	**83.7**	82.4
Sentence 10	65.5 (9.8)	77.7 (8.8)	69.9 (6.3)	78.9 (5.3)	78.5 (7.5)	69.6 (5.1)	**80.5 (5.4)**	79.3	80.1
Read text	65.0 (8.4)	78.4 (9.2)	73.4 (10.1)	82.7 (7.7)	83.3 (9.9)	75.5 (8.4)	**84.2 (7.4)**	82.5	81.7
Monologue	68.1 (7.5)	78.6 (8.1)	72.0 (10.7)	80.8 (8.5)	82.9 (5.8)	81.5 (6.7)	83.1 (7.5)	**86.0**	85.2
F. T. Early	70.8 (7.7)	80.1 (7.9)	79.1 (7.5)	**86.2 (6.7)**	85.1 (7.1)	83.9 (9.3)			
F. T. Late 1	75.3	79.8	80.5	85.8	**86.1**	80.6	84.8	85.2	
F. T. Late 2	74.5	81.1	82.2	85.3	86.0	83.3	**87.7**		86.0
Average	68.4	76.1	70.9	80.6	78.3	72.1	81.1	79.5	80.6

RAE: recurrent autoencoder. **F. T.**: Fusion of tasks. **F. F.**: fusion of feature sets.
Results of the best feature set per task are highlighted in bold.

The highest UAR is obtained by combining information from all speech tasks and all feature sets, using the *early* fusion for the feature sets and the *Late 2* fusion for the tasks (87.7%). The early fusion seems to be the most accurate method for most of the speech tasks, while late fusion strategies are the most accurate for the fusion at task-level. This is explained because early fusion is better to model different and complementary information produced by each feature set, and late fusion methods deal better with the redundant information that may appear when using the same

features computed from different speech tasks. When considering different models individually, the highest UARs (above 80%) are obtained mainly with the OpenSMILE and phonological features. OpenSMILE features comprises an extensive feature set with more than 6000 computed features, which can be difficult to interpret and to analyze in the clinical practice. Conversely, the phonological analysis was designed specifically for pathological speech modeling, and comprises only 108 features based on 18 phonological classes to model the pronunciation of the different phonemes of the language. This fact makes the phonological analysis more useful in the clinical practice. Phonation, articulation, prosody, and the RAE-based features are also accurate to classify the addressed population, and provide complementary information to improve the accuracy. The results obtained with phonation and prosody features are around 70%. This is likely because although these dimensions are affected in PD patients, they are not the most sensitive aspects to evaluate dysarthric speech signals.

The most accurate speech tasks for the classification in the early fusion are the different DDK exercises (up to 85.6%) and the read text (84.2%). On one hand, the DDK exercises are very easy to produce and are potentially useful to evaluate the speech of patients in almost every language. This is mainly because the production of these DDK tasks requires the speaker to move several articulators including lips, tongue, and velum. On the other hand, the read text is a more complete exercise that involves the pronunciation and assessment of the different phonemes of the language. Hence it helps to perform a controlled evaluation of the speech of the patients. In summary, each feature set and speech task is useful to characterize different aspects related to speech production. The improvements observed when models and tasks are combined indicate that the information is complementary and suitable to be used to assess the speech of PD patients.

The results in the Multimodal corpus suggest that the proposed methods are valid and accurate to model the speech of PD patients in a clinical setting. Now the aim is to evaluate whether those methods are also accurate to model the speech of PD patients in at-home environments using smartphone data, which can be used to monitor the state of the patients at-home. The results classifying PD patients vs. HC subjects from the Apkinson corpus are presented in Table 9.2. The best results per task are also highlighted in bold. The highest UAR in this case is obtained with the *Late 2* fusion strategy, both at feature and task-levels (89.5%). This result is similar to the obtained previously for the Multimodal corpus. The best result is slightly higher in this case probably because the size of the Apkinson corpus is smaller than the size of the Multimodal corpus. For most speech tasks the best result is obtained using any of the fusion strategies, similar to the previous case with the Multimodal corpus. The most accurate feature sets on average correspond to the RAE (76%), phonological (75.8%), and articulation (75.6%). This is also similar to the results previously reported. The only feature sets that are highly affected because of the smartphone data are phonation and prosody. The reduction of the accuracy for such feature sets is 6% for phonation and 7.9% for prosody. The reduction for the other feature sets is 2.5% for phonological and 0.5% for articulation. The RAE feature set experienced a slight improvement of 3.9%. In summary, the results indicate that there is not a visible impact in the classification when speech signals collected from

smartphones are considered. In addition, only phonation and prosody analyses seems to be affected because of the use of smartphone data.

Table 9.2: Results classifying PD patients vs. HC subjects from the Apkinson corpus using different speech feature sets and SVM classifers. Results in terms of UAR [%].

	Feature sets					F. F.	F. F.	F. F.
Task	**Phonation**	**Articulation**	**Prosody**	**Phonological**	**RAE**	**Early**	**Late 1**	**Late 2**
DDK1	69.8 (9.4)	81.7 (9.9)	60.3 (13.0)	78.1 (10.5)	83.3 (9.1)	74.7 (13.7)	**87.0**	85.0
DDK2	61.1 (16.6)	77.1 (9.0)	66.5 (14.6)	69.9 (11.7)	73.7 (14.6)	64.9 (12.3)	77.7	**78.7**
DDK3	67.6 (20.2)	76.3 (12.7)	60.4 (11.3)	72.3 (14.0)	71.1 (14.6)	76.0 (10.0)	78.0	**83.7**
Sentence 1	48.3 (11.2)	70.8 (11.5)	63.3 (19.0)	68.3 (17.5)	72.5 (11.5)	74.6 (9.4)	73.6	**77.9**
Sentence 2	65.0 (9.4)	69.6 (12.1)	67.1 (16.2)	68.8 (16.7)	70.0 (17.6)	**71.7 (15.8)**	71.6	72.1
Sentence 3	61.7 (12.5)	76.3 (12.4)	57.9 (16.4)	68.8 (11.2)	72.9 (14.3)	**77.1 (11.4)**	75.8	74.0
Sentence 4	53.8 (13.9)	66.7 (11.3)	59.2 (13.4)	68.3 (17.8)	76.3 (10.4)	**77.9 (14.8)**	76.3	75.8
Sentence 5	54.8 (16.2)	71.3 (14.7)	60.2 (19.7)	75.4 (10.8)	76.3 (12.4)	**78.3 (11.0)**	73.3	77.6
Sentence 6	65.8 (9.6)	70.0 (19.5)	59.2 (10.5)	**76.3 (16.3)**	66.3 (15.0)	72.5 (14.7)	75.8	75.8
Sentence 7	53.8 (14.0)	67.1 (15.4)	51.7 (16.9)	**76.3 (15.8)**	68.3 (12.9)	75.0 (10.5)	68.0	70.2
Sentence 8	54.2 (14.6)	76.7 (14.0)	53.8 (12.1)	73.3 (7.0)	79.2 (12.2)	79.6 (8.2)	79.7	**80.9**
Sentence 9	62.1 (14.8)	72.1 (6.7)	74.6 (18.0)	77.1 (14.3)	75.0 (12.8)	76.7 (14.7)	**83.5**	81.7
Sentence 10	57.1 (15.1)	**76.7 (12.8)**	55.4 (18.2)	70.4 (11.1)	64.2 (14.1)	73.8 (10.9)	74.1	75.4
Monologue	71.7 (16.9)	77.1 (12.3)	62.5 (13.4)	81.9 (11.1)	82.5 (15.8)	**82.7 (14.7)**	82.6	
F. T. Early	65.4 (11.8)	85.4 (7.7)	68.8 (14.6)	87.9 (9.4)	**86.3 (11.8)**			
F. T. Late 1	72.4	86.4	73.6	87.2	87.2	**88.4**	**88.4**	
F. T. Late 2	75.7	84.9	77.2	88.4	87.2	88.4		**89.5**
Average	62.4	75.6	63.0	75.8	76.0	77.0	77.7	78.5

RAE: recurrent autoencoder. **F. T.**: Fusion of tasks. **F. F.**: fusion of feature sets.
Results of the best feature set per task are highlighted in bold.

Additional to the analysis based on feature extraction and the later SVM classifier, the classification is performed using the different deep learning methods explained in Section 4.5. The results are seen in Tables 9.3 and 9.4, for the Multimodal and Apkinson corpus, respectively.

For the Multimodal corpus, no high differences in the accuracy are observed among the different models. The highest UAR for the Multimodal corpus is obtained using the ResNet architecture when processing the Mel-spectrogams of the onset transitions and when all speech tasks are combined using the *Late 1* fusion strategy (96.2%). This is particularly positive because the network is able to learn about the articulatory impairments present in the onset transitions, and which are related to the difficulties of the patients to start the vibration of the vocal folds [Vasq 17a, Oroz 16b]. The best results for the Apkinson corpus is obtained also with the ResNet model and with the combination of the different tasks (97.2%). These results indicate that there is not a visible impact in the classification when speech signals collected from smartphones are considered. In this case the most accurate results were obtained when using the Mel spectrogram from the full chunks rather than only the onset transitions. This result could be likely explained due to the small size of the Apkinson corpus. The number of samples in the Multimodal data is big enough to train an accurate model using only the spectrograms of the transitions, while the Akinson corpus needs to be trained with the full set of chunks to get accurate results. This explanation can be validated based on Table 9.5, which shows the number of training tensors for each corpus and model.

A summary of the best results obtained classifying PD patients vs. HC subjects using the different methods is shown in Table 9.6. The results include additional performance metrics like sensitivity, specificity, F-score, and the area under the ROC

Table 9.3: Results classifying PD patients vs. HC subjects from the Multimodal corpus using deep learning methods to model speech signals. Results in terms of UAR [%].

Task	CNN Transitions	CNN Full	ResNet Trasitions	ResNet Full
DDK1	90.0	88.8	88.8	**90.1**
DDK2	85.7	90.8	90.8	**92.0**
DDK3	90.5	**91.3**	**91.3**	90.1
DDK4	85.1	89.2	**90.1**	86.3
DDK5	89.6	89.9	**91.0**	86.2
DDK6	91.3	**91.9**	91.5	87.7
Sentence 1	87.4	90.1	**91.8**	87.9
Sentence 2	91.4	90.0	**93.3**	86.3
Sentence 3	88.2	**91.3**	89.7	86.3
Sentence 4	89.8	**92.8**	91.4	87.5
Sentence 5	87.9	**91.0**	89.6	88.7
Sentence 6	90.6	**92.8**	90.6	88.7
Sentence 7	87.1	**92.1**	90.2	87.2
Sentence 8	88.8	**94.1**	92.1	87.7
Sentence 9	89.2	**92.0**	91.4	87.2
Sentence 10	91.0	**91.7**	87.7	88.0
Read Text	90.7	**92.4**	90.0	**92.4**
Monologue	75.6	81.6	75.5	**88.8**
F. T. Late 1	93.2	94.9	**96.2**	93.5
F. T. Late 2	93.2	94.2	**95.5**	91.7
Average	88.3	90.8	89.8	88.3

Results of the best feature set per task are highlighted in bold. **F. T.**: Fusion of tasks.

Table 9.4: Results classifying PD patients vs. HC subjects from Apkinson corpus using deep learning methods to model speech signals. Results in terms of UAR [%].

Task	CNN Transitions	CNN Full	ResNet Trasitions	ResNet Full
DDK1	71.7	73.7	71.2	**93.8**
DDK2	77.5	69.4	61.4	**89.1**
DDK3	78.2	68.9	69.8	**93.5**
Sentence 1	63.5	68.3	68.5	**83.6**
Sentence 2	74.8	67.4	72.5	**86.2**
Sentence 3	72.2	71.2	71.6	**85.7**
Sentence 4	53.5	67.8	56.4	**87.9**
Sentence 5	69.5	70.4	67.7	**87.8**
Sentence 6	82.4	63.2	62.6	**86.6**
Sentence 7	77.8	71.2	74.2	**83.7**
Sentence 8	70.4	72.9	69.2	**88.7**
Sentence 9	72.4	69.6	63.5	**88.7**
Sentence 10	60.0	72.5	66.2	**87.9**
Monologue	70.9	78.1	65.5	**87.2**
F. T. Late 1	74.0	79.7	67.1	**97.2**
F. T. Late 2	76.2	79.7	71.3	**96.8**
Average	71.6	71.5	67.4	89.0

Results of the best feature set per task are highlighted in bold. **F. T.**: Fusion of tasks.

curve (AUC). The best result obtained for each corpus is highlighted in bold. They were obtained with the deep ResNet models. In addition, for almost all cases, the best results are obtained using late fusion strategies to combine the speech tasks. UARs up to 96.2% and up to 97.2% are obtained for the Multimodal and Apkison corpus, respectively. After the results using the ResNet models, the best results are

Table 9.5: Number of training tensors for the Multimodal and Apkinson corpus to train the ResNet models.

Data	Segments	
	Full	Transitions
Multimodal	208220	92374
Apkinson	37880	14000

obtained with the fusion of the feature sets using the *Early fusion* for the Multimodal corpus (UAR=87.7%) and the *Late 2* fusion for the Apkinson corpus (UAR=89.5%).

Table 9.6: Best results obtained for each method classifying PD patients and HC subjects in the Multimodal and Apkinson corpus using speech signals.

Feature set	**Task**	**ACC [%]**	**F-score**	**UAR [%]**	**SENS [%]**	**SPEC [%]**	**AUC**
Multimodal corpus							
Phonation	F. T. Late 1	68.5	0.675	75.3	59.8	90.8	0.855
Articulation	F. T. Late 2	81.3	0.785	81.1	81.7	80.5	0.905
Prosody	F. T. Late 2	82.5	0.797	82.2	83.0	81.4	0.902
OpenSMILE	F. T. Early	86.2	0.845	86.2	86.3	86.1	0.934
Phonological	F. T. Late 1	86.1	0.837	86.1	86.2	86.0	0.931
RAE	F. T. Early	85.5	0.830	83.9	87.8	80.0	0.925
F. F. Early	F. T. Late 2	85.8	0.838	87.7	83.5	91.9	0.950
F. F. Late 1	Monologue	86.0	0.837	86.0	86.0	86.0	0.918
F. F. Late 2	F. T. Late 2	84.9	0.827	86.0	83.5	88.5	0.941
ResNet Transitions	**F. T. Late 1**	**94.4**	**0.935**	**96.2**	**92.3**	**100.0**	**0.999**
Apkinson corpus							
Phonation	F. T. Late 2	78.6	0.765	75.7	58.1	93.3	0.795
Articulation	F. T. Late 1	88.4	0.875	86.4	74.4	98.3	0.950
Prosody	F. T. Late 2	79.6	0.780	77.2	62.8	91.7	0.816
Phonological	F. T. Late 2	90.3	0.896	88.4	76.7	100.0	0.959
RAE	F. T. Late 1 & 2	89.3	0.885	87.2	74.4	100.0	0.919
F. F. Early	F. T. Late 1 & 2	90.3	0.896	88.4	76.7	100.0	0.966
F. F. Late 1	F. T. Late 1	90.3	0.896	88.4	76.7	100.0	0.971
F. F. Late 2	F. T. Late 2	91.3	0.907	89.5	79.1	100.0	0.988
ResNet Full	**F. T. Late 1**	**97.0**	**0.970**	**97.2**	**97.7**	**96.7**	**0.997**

ACC [%]: accuracy, **UAR [%]**: unweighted average recall, **SENS [%]**: sensitivity, **SPEC [%]**: specificity
AUC: Area under the ROC curve, **RAE**: recurrent autoencoder. **F. T.**: Fusion of tasks. **F. F.**: fusion of features. The best result for each corpus is highlighted in bold

The confusion matrices, ROC curves, and the histograms obtained with the scores of the predictions of the ResNet models are shown in Figures 9.1 and 9.2 for the Multimodal and Apkinson corpus, respectively. The scores for the histograms are obtained as the difference between the probabilities of a speech sample to be classified as PD patient or as an HC subject. These probabilities are computed at the output of the Softmax activation in the neural network. Note that almost a perfect classification is obtained for both corpora.

The results are compared to the ones obtained with the classical feature extraction and classification using the SVM. The confusion matrices, ROC curves, and the respective histograms obtained using the traditional techniques are shown in Figures 9.3 and 9.4 for the Multimodal and Apkinson corpus, respectively. For this case the scores used to compute the histograms correspond to the distance to the hyperplane of the SVM classifier. The results for the Multimodal corpus correspond to the

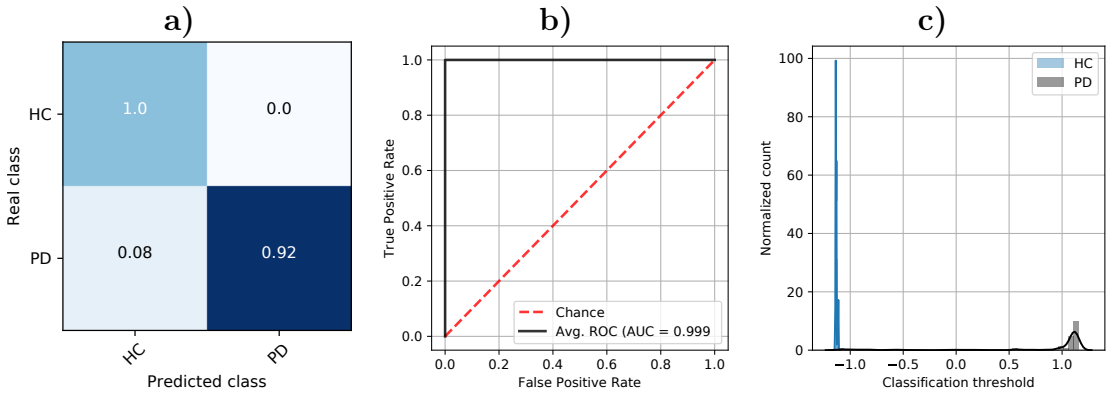

Figure 9.1: Details of the best result obtained classifying PD patients and HC subjects from the Multimodal corpus using the ResNet models. **a)** Normalized confusion matrix. **b)** ROC curve. **c)** Distribution of the classification scores.

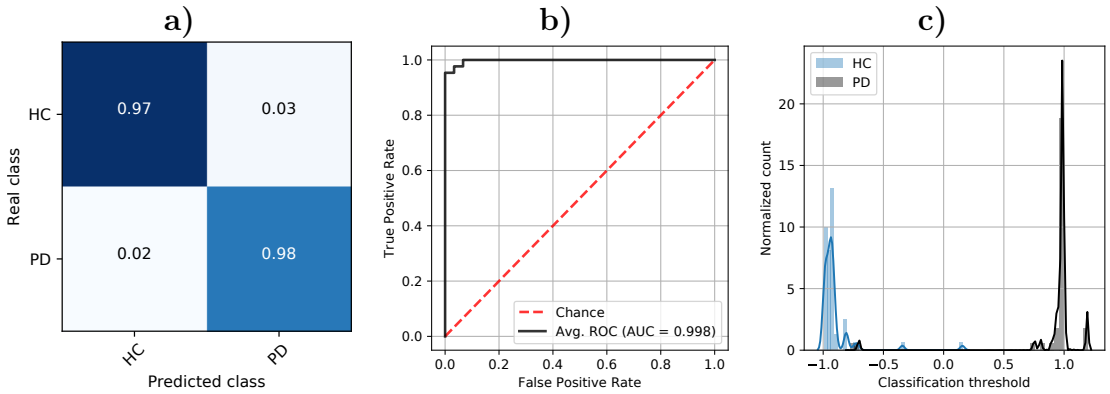

Figure 9.2: Details of the best result obtained classifying PD patients and HC subjects from the Apkinson corpus using the ResNet models. **a)** Normalized confusion matrix. **b)** ROC curve. **c)** Distribution of the classification scores.

ones obtained with the *Early* fusion at feature level and *Late 2* fusion at task-level. The results for the Apkinson corpus correspond to the *Late 2* fusion at feature- and task-levels.

Besides the results obtained with the different methods, and with the aim to evaluate how such models are affected by the phonetic content pronounced by the patients, Figure 9.5 shows the ranking of the predictions for the 10 sentences of the Multimodal corpus, using the different methods. The sentences are sorted according to their average accuracy. Sentences 9 and 7 are the most accurate on average to classify PD vs. HC subjects. Conversely, sentences 1 and 5 show to be the less accurate for the addressed problem. This could be explained because such sentences can contain phonemes that are more difficult to pronounce by PD patients than for HC subjects. It is important to find which are those groups of phonemes in order to design more proper and standardized tests for the evaluation of the patients.

The correlation between the UAR obtained with each method and different phonetic aspects of the sentences is shown in Table 9.7. Those values with *strong* correlations are highlighted in bold. The results indicate that specially the accuracy

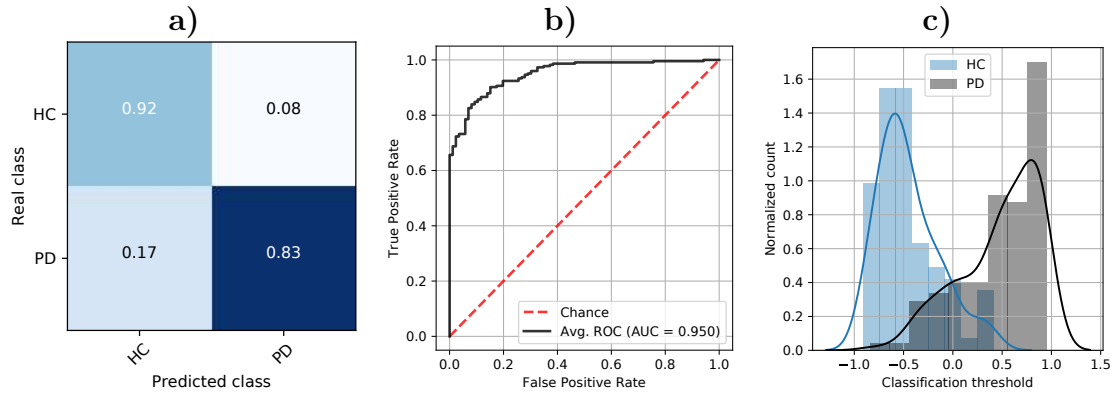

Figure 9.3: Details of the best result obtained classifying PD patients and HC subjects from the Multimodal corpus using the different features and SVM classifiers. **a)** Normalized confusion matrix. **b)** ROC curve. **c)** Distribution of the classification scores.

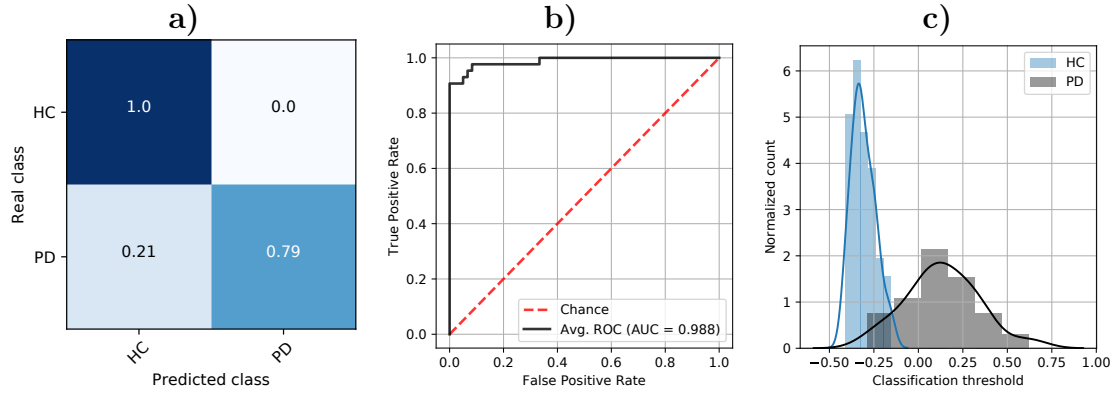

Figure 9.4: Details of the best result obtained classifying PD patients and HC subjects from the Apkinson corpus using the different features and SVM classifiers. **a)** Normalized confusion matrix. **b)** ROC curve. **c)** Distribution of the classification scores.

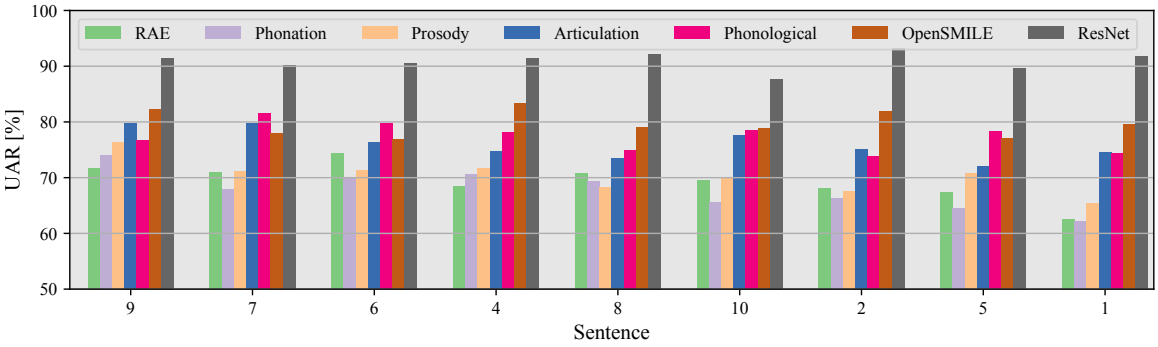

Figure 9.5: Ranking of the sentences of the Multimodal corpus. The sentences are sorted according to their average accuracy.

obtained with the phonological features is highly correlated with different aspects of the sentences like the average duration, the standard deviation of the duration,

the number of words, the number of phonemes, the number of plosives, the number of fricatives and the number of vowels. A high correlation is observed between the accuracy of the prosody features and the number of voiced phonemes in the sentence. These results suggest that such a group of phonemes has to be considered with special attention when designing speech protocols for the assessment of the disease, depending on the method used for the analysis. In addition, novel characterization strategies should be designed focused on the assessment of such groups of phonemes, like the use of specific phonological features designed only for plosives, vowels, fricatives, among others. For phonation, articulation, OpenSMILE, and the ResNet models there are no visible correlations between the accuracy obtained and the phonetic content present in the sentences, which indicate that these methods are text-independent and they can be used for non-intrusive evaluation of the patients.

Table 9.7: Spearman's correlation between the accuracy obtained with each feature set in the 10 sentences and different phonetic aspects of the sentences.

Aspect	Phonation	Articulation	Prosody	OpenSMILE	Phonological	RAE	ResNet
Avg duration	0.234	0.405	0.451	-0.236	**0.762**	0.514	-0.421
Std Duration	0.292	0.343	0.489	-0.233	**0.751**	0.547	-0.372
# words	0.469	0.383	0.588	-0.070	**0.750**	**0.631**	-0.286
# phonemes	0.365	0.375	0.559	-0.112	**0.770**	0.580	-0.365
# V phonemes	0.466	0.366	**0.600**	-0.075	**0.753**	0.641	-0.306
# U phonemes	-0.161	0.058	0.187	-0.239	0.522	0.122	-0.432
# plosives	-0.050	0.182	0.182	-0.430	**0.611**	0.413	-0.397
# fricatives	0.092	0.353	0.364	-0.071	**0.625**	0.256	-0.397
# nasals	0.459	0.068	0.513	0.068	0.513	0.513	-0.014
# laterals	0.426	0.158	0.328	-0.134	0.304	0.353	-0.012
# vowels	0.399	0.421	0.581	-0.094	**0.786**	0.609	-0.393

9.1.2 Automatic Evaluation of the Dysarthria Severity of Patients

The automatic classification of PD patients and HC subjects is important because it represents a step forward in the development of computer aided tools to support medical experts in the diagnosis process. However, once patients are already diagnosed, it is necessary to evaluate their disease severity. Particularly, the speech signals are suitable to evaluate the dysarthria severity of the patients. An accurate evaluation of the dysarthria severity helps to make timely decisions regarding the medication and the therapy for the patients. Additionally, if such a screening is performed from speech recordings, the treatment can be followed remotely, then the costs of the treatment would decrease dramatically.

The evaluation of the dysarthria severity of patients is based on the m-FDA scale, which was administered by phoniatricians using the collected speech recordings. As it was mentioned in Section 3.1.2, the m-FDA scale can be administered remotely using only speech recording of the patients. The evaluation of the dysarthria severity is performed in three scenarios: (1) the prediction of the value of the m-FDA scale using the methods described in Chapter 4 and regression algorithms; (2) the longitudinal evaluation of the patients from the Longitudinal and the At-Home corpus using both speaker models based on GMM-UBM systems and regression algorithms; and (3) the

classification of patients in different levels of dysarthria severity (mild, intermediate, severe) using multi-class classification methods.

Dysarthria Level Evaluation based on Regression Algorithms

Similar to the classification experiments previously addressed, the evaluation of the dysarthria severity of the participants is performed considering the different speech dimensions, speech tasks, and their combinations. The speech dimensions include the feature sets to model phonation, articulation, prosody, phonological, and the RAE-based models, in addition to the end-to-end modeling using deep learning techniques. The regression is performed using an SVR algorithm. The results obtained estimating the m-FDA score from the Multimodal corpus are presented in Table 9.8. The results are presented in terms of the Spearman's correlation coefficient. The OpenSMILE features are also included only as a baseline, and they are not included in the fusion strategies. Results of the best feature set per task are highlighted in bold.

Table 9.8: Results estimating the m-FDA scale of the subjects from the Multimodal corpus using different speech eature sets and SVR regressors. Results in terms of the Spearman's correlation coefficient.

	Feature sets						F. F.	F. F.	F. F.
Task	**Phonation**	**Articulation**	**Prosody**	**OpenSMILE**	**Phonological**	**RAE**	**Early**	**Late 1**	**Late 2**
DDK1	0.340	0.416	0.309	0.442	0.509	0.353	0.379	0.535	**0.600**
DDK2	0.333	0.512	0.446	0.474	0.409	0.389	0.429	0.531	**0.605**
DDK3	0.302	0.466	0.331	0.383	0.495	0.344	0.433	0.497	**0.566**
DDK4	0.343	0.413	0.353	0.357	0.390	0.383	0.428	0.519	**0.566**
DDK5	0.394	0.481	0.403	0.371	0.371	0.284	0.409	0.450	**0.543**
DDK6	0.402	0.454	0.355	0.498	0.396	0.289	0.470	0.478	**0.578**
Sentence 1	0.293	0.455	0.315	0.436	0.389	0.286	0.490	0.463	**0.548**
Sentence 2	0.280	0.405	0.438	0.375	0.344	0.390	0.391	0.457	**0.513**
Sentence 3	0.230	0.393	0.262	0.397	0.399	0.347	0.471	0.471	**0.520**
Sentence 4	-0.007	0.364	0.224	0.386	0.271	0.415	0.446	0.493	**0.540**
Sentence 5	0.236	0.356	0.345	0.434	0.341	0.257	0.473	0.487	**0.536**
Sentence 6	0.353	0.484	0.459	0.347	0.367	0.412	0.416	0.483	**0.513**
Sentence 7	0.258	0.334	0.253	0.351	0.328	0.315	0.446	0.442	**0.519**
Sentence 8	0.186	0.354	0.321	0.444	0.294	0.353	0.416	0.501	**0.561**
Sentence 9	0.368	0.436	0.292	0.399	0.387	0.294	0.321	0.529	**0.599**
Sentence 10	0.197	0.400	0.303	0.413	0.407	0.249	0.441	0.492	**0.514**
Read text	0.279	0.353	0.302	0.344	0.407	0.234	0.387	0.478	**0.537**
Monologue	0.321	0.415	0.391	0.437	0.379	0.382	0.381	0.476	**0.508**
F. T. Early	0.419	0.515	0.377	0.506	0.451	0.432	**0.591**		
F. T. Late 1	0.549	0.587	0.420	0.476	0.493	0.423	0.521	**0.567**	
F. T. Late 2	0.528	0.582	0.488	0.588	**0.611**	0.552	0.581		0.605

Results of the best feature set per task are highlighted in bold.
RAE: recurrent autoencoder. **F. T.**: Fusion of tasks. **F. F.**: fusion of features.

The highest correlation is obtained by combining information from all speech tasks and using the phonological features (0.611). The best results for individual speech tasks are obtained by the fusion of all features using the *late 2* fusion strategy. When considering the different features individually, moderate correlations (ρ >0.4) are observed mainly with the articulation and phonological features. This gives an idea about the suitability of these feature sets to accurately model the dysarthria severity of patients, and about which speech dimensions are more affected when the severity increases. The results obtained with phonation and prosody are around 0.35. This is likely because although phonation and prosody are impaired with the dysarthria severity, these are not the most sensitive symptoms. Finally, the most sensitive tasks

to evaluate the dysarthria severity are the DDK exercises, among which particularly DDK1 (/pa-ta-ka/) and DDK2 (/pa-ka-ta/) exhibit strong correlations with the total m-FDA scale (0.600, and 0.605, respectively).

Additional to the analysis based on feature extraction and the later SVR regression, the estimation of the m-FDA scale is performed using the ResNet models. The results are observed in Table 9.9. Unfortunately, the results in this case are not that accurate as the ones obtained in the classification experiments. The main reason would be the lack of labeled data about the m-FDA score of the patients to train the deep regression models. The results can improve if there is access to a largest group of patients labeled according of the m-FDA scale. The results could improve as well if a transfer learning strategy is considered by using an already existing pre-trained model for speech severity assessment, in a similar way as the previously addressed for pathological speech classification [Vasq 21c].

Table 9.9: Results estimating the m-FDA scale of the subjects from the Multimodal corpus using deep learning methods to model speech signals. Results in terms of the Spearman's correlation.

Task	ResNet Transitions	ResNet Full
DDK1	0.421	0.392
DDK2	0.423	0.334
DDK3	0.423	0.346
DDK4	0.405	0.326
DDK5	0.390	0.331
DDK6	0.412	0.327
Sentence 1	0.348	0.231
Sentence 2	0.308	0.249
Sentence 3	0.326	0.286
Sentence 4	0.332	0.280
Sentence 5	0.284	0.300
Sentence 6	0.386	0.340
Sentence 7	0.360	0.262
Sentence 8	0.374	0.269
Sentence 9	0.327	0.256
Sentence 10	0.305	0.262
Read Text	0.265	0.240
Monologue	0.219	0.304
F. T. Late 1	0.493	0.392
F. T. Late 2	0.486	0.391

F. T.: Fusion of tasks.

A summary of the best results obtained with each method is shown in Table 9.10, including additional metrics like the Pearson's correlation coefficient and the mean average error (MAE). The results include also the p-values of the correlations to test whether the results are significant or not. The best result is obtained using the phonological features, and combining all speech tasks using the *late 2* fusion at task-level. Figure 9.6 shows the errors in the evaluation of the m-FDA scores. The displayed result corresponds to the ones obtained with the phonological features combining all speech tasks. Although the result is satisfactory, other regression strategies can be considered to improve the correlation values.

Figure 9.7 shows the ranking of the estimation of the m-FDA score for the 10 sentences of the Multimodal corpus. The sentences are sorted according to their average correlation. Sentence 6 is the most accurate on average to estimate the

Table 9.10: Best results obtained for each method to evaluate the m-FDA scale of the subjects in the Multimodal corpus using speech signals.

Feature set	Task	r	p-val r	ρ	p-val ρ	MAE
Phonation	F. T. Late 1	0.217	0.002	0.549	$\ll$0.005	10.4
Articulation	F. T. Late 1	0.197	0.006	0.587	$\ll$0.005	10.5
Prosody	F. T. Late 2	0.565	$\ll$0.005	0.488	$\ll$0.005	6.9
OpenSMILE	F. T. Late 2	0.652	$\ll$0.005	0.588	$\ll$0.005	5.9
Phonological	**F. T. Late 2**	**0.689**	**$\ll$0.005**	**0.611**	**$\ll$0.005**	**5.8**
RAE	F. T. Late 2	0.587	$\ll$0.005	0.552	$\ll$0.005	5.8
F. F. Early	F. F. Early	0.660	$\ll$0.005	0.591	$\ll$0.005	6.3
F. F. Late 1	F. T. Late 1	0.595	$\ll$0.005	0.567	$\ll$0.005	9.6
F. F. Late 2	F. T. Late 2	0.672	$\ll$0.005	0.605	$\ll$0.005	6.4
ResNet Transitions	F. T. Late 1	0.363	$\ll$0.005	0.493	$\ll$0.005	10.3

r: Pearson's correlation, ρ: Spearman's correlation, **MAE**: mean average error, **RAE**: recurrent autoencoder. **F. T.**: Fusion of tasks. **F. F.**: fusion of feature sets.

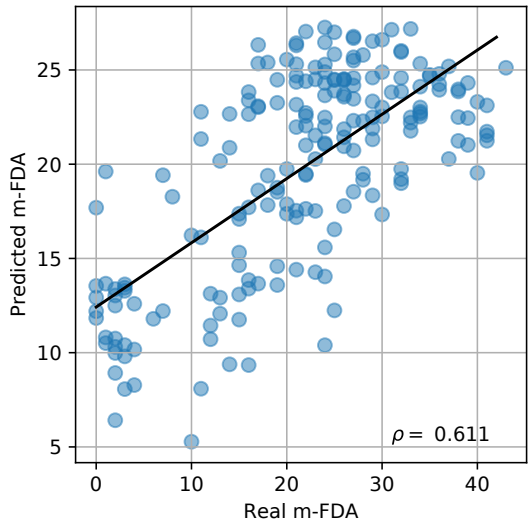

Figure 9.6: Details of the best result obtained estimating the m-FDA scale of the subjects from the Multimodal corpus using the different features and SVR regressors.

m-FDA score. Conversely, sentences 7 and 5 show to be the less accurate for the addressed problem.

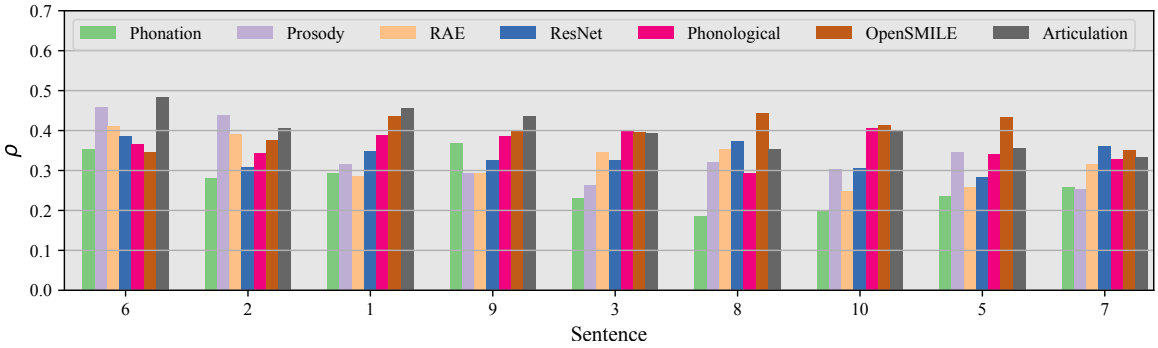

Figure 9.7: Ranking of the sentences of the Multimodal corpus estimating the m-FDA score of the subjects. The sentences are sorted according to their average Spearman's correlation.

The correlation between the Spearman's correlation obtained with each method and different phonetic aspects of the sentences is shown in Table 9.11. Those values with *strong* correlations ($|\rho| > 0.6$) are highlighted in bold. A negative correlation is observed between the number of nasal phonemes in the sentence and the performance of the regression algorithms for articulation and phonological features. There are also correlations between the number of laterals and the outputs of the openSMILE and RAE-based features. For phonation, prosody, and the ResNet models there are no correlation between the performance of the regressors and the phonetic content present in the sentences, which indicate that these methods are text-independent and they can be used for non-intrusive evaluation of the patients.

Table 9.11: Spearman's correlation between the performance obtained with each speech feature set to evaluate the m-FDA score and different phonetic aspects of the sentences.

Aspect	**Phonation**	**Articulation**	**Prosody**	**OpenSMILE**	**Phonological**	**RAE**	**ResNet**
Avg. Duration	-0.027	-0.538	-0.260	-0.349	-0.332	-0.138	0.166
Std. Duration	-0.058	-0.540	-0.204	-0.386	-0.423	-0.029	0.219
# words	-0.189	-0.588	-0.394	-0.329	-0.502	0.070	0.340
# phonemes	-0.133	-0.544	-0.328	-0.312	-0.423	-0.057	0.257
# V phonemes	-0.171	-0.590	-0.325	-0.302	-0.501	0.027	0.277
# U phonemes	-0.135	-0.445	-0.213	-0.071	-0.213	-0.355	0.135
# plosives	0.099	-0.347	0.149	-0.099	-0.132	-0.347	0.231
# fricatives	-0.147	-0.549	-0.517	-0.321	-0.299	-0.179	0.082
# nasals	-0.243	**-0.703**	-0.311	-0.095	**-0.743**	0.135	0.432
# laterals	-0.255	0.012	-0.182	**-0.693**	-0.328	**0.766**	0.474
# vowels	-0.149	-0.554	-0.354	-0.288	-0.393	-0.083	0.210

Longitudinal Assessment of Patients

The longitudinal evaluation of the dysarthria severity of patients is performed in two scenarios to cover both the long- and short-term progression of the disease. The long-term evaluation is performed with the Longitudinal corpus (see Section 3.3.2). The short-term evaluation is performed with the At-Home corpus (see Section 3.3.3). At the same time, two different methods are considered to model each corpus. The first one comprises the use of an SVR trained with the data from the Multimodal corpus (excluding those subjects from the longitudinal and At-Home data), and using both corpora as independent test sets for the regression problems. The second approach consists of the use of unsupervised speaker models based on the GMM-UBM approach introduced in [Aria 18a], and adapted here with more data and additional feature sets, in a similar way as in [Vasq 20b]. In this case UBM models are trained using information from the HC subjects from the Multimodal corpus. Then specific GMMs are adapted for each speaker in each session from the Longitudinal and the At-Home corpus. At the same time, the modeling is performed with the different feature sets considered in the previous experiments, and the different speech tasks.

The results predicting the dysarthria severity of the patients in the Longitudinal corpus in all sessions using the SVR regression strategy are shown in Table 9.12. The best result is obtained with the phonological features and the DDK exercises (ρ =0.506). This result gives insights about the generalization capacity of the trained

model to a new hidden test set. In this case, the fusion of the feature sets and tasks did not yield the most accurate results, contrary to the previous experiments evaluating the m-FDA score of the patients.

Table 9.12: Results predicting the m-FDA scale of the subjects from the Longitudinal corpus using different speech feature sets and SVR regressors. Results in terms of the Spearman's correlation coefficient.

	Feature sets					F. F.	F. F.	F. F.
Speech task	**Phonation**	**Articulation**	**Prosody**	**Phonological**	**RAE**	**Early**	**Late 1**	**Late 2**
DDK1	0.288	0.038	0.040	**0.506**	0.154	0.281	0.215	0.393
DDK2	0.027	0.085	0.014	0.312	0.034	0.322	0.211	0.314
DDK3	0.166	0.092	0.161	0.436	0.170	0.114	0.211	0.387
DDK4	-0.090	0.292	0.310	0.231	0.091	0.054	0.209	0.379
DDK5	0.049	0.095	-0.238	0.020	-0.080	0.253	0.019	0.210
DDK6	0.314	0.277	-0.148	0.209	0.057	0.155	0.122	0.342
Sentence 1	-0.117	0.082	-0.330	0.223	0.165	0.175	0.106	0.416
Sentence 2	0.168	-0.022	-0.074	0.272	-0.040	0.401	0.176	0.365
Sentence 3	0.264	0.057	0.100	-0.066	0.088	0.063	0.078	0.250
Sentence 4	0.049	0.135	-0.019	0.176	0.162	0.180	0.153	0.246
Sentence 5	0.128	0.052	-0.051	0.125	0.178	0.169	0.082	0.200
Sentence 6	0.153	0.029	-0.223	0.115	0.057	0.214	0.211	0.312
Sentence 7	0.178	0.120	-0.107	0.339	-0.023	0.180	0.138	0.306
Sentence 8	0.149	-0.017	0.029	0.185	0.227	0.178	0.158	0.183
Sentence 9	0.057	0.130	0.213	0.355	0.209	0.165	0.290	0.384
Sentence 10	0.098	0.032	-0.034	0.224	0.205	0.273	0.168	0.307
Read text	0.058	0.060	0.015	0.036	-0.121	-0.042	-0.065	0.152
Monologue	0.296	0.011	-0.253	0.217	0.126	0.081	0.032	0.439
F. T. Early	0.052	0.226	-0.382	0.081	0.073		0.217	0.280
F. T. Late 1	0.305	0.056	-0.055	0.278	0.103			
F. T. Late 2	0.361	0.240	-0.601	0.396	0.235			

RAE: recurrent autoencoder. **F. T.**: Fusion of tasks. **F. F.**: fusion of features.
Best result is highlighted in bold.

The results predicting the m-FDA score of the participants in the longitudinal corpus using the unsupervised GMM-UBM systems are shown in Table 9.13. In this case, the fusion of features and the fusion of tasks are obtained based on the median of the predictions using each method separately. The highest correlation (ρ =0.350) is observed when all features are considered together in the DDK5 task (/ta/). In general, the results are not satisfactory, and they are lower than the observed with the SVR approach. This is contrary to the results obtained in [Aria 18a], where the GMM-UBM was better than the SVR. The main reason is because here there are more data to train the regression models, thus obtaining more accurate results using a supervised model like the SVR.

Figure 9.8 displays curves with the comparison of the estimated m-FDA scores (cyan lines) and the real labels assigned by the phoniatrician (black lines) for each of the nine speakers of the Longitudinal corpus. The horizontal axis represents the recording session. The lines for each speaker represent the progression of the dysarthria level among recording sessions. The predicted scores follow the trend of the dysarthria level for most of the cases. The largest differences are observed in patients PD05 and PD07, specially in the first two recording sessions. The dysarthria progression is predicted with strong correlation (ρ >0.6) for four of the nine PD patients, and with moderate correlations (ρ >0.4) for six of the nine patients. The

Table 9.13: Results predicting the m-FDA scale of the subjects from the Longitudinal corpus using different speech feature sets and GMM-UBM models. Results in terms of the Spearman's correlation coefficient.

	Feature sets					
Speech task	**Phonation**	**Articulation**	**Prosody**	**Phonological**	**RAE**	**F. F.**
DDK1	-0.105	-0.094	-0.013	0.191	-0.026	0.156
DDK2	-0.100	-0.128	0.055	0.187	0.026	0.072
DDK3	-0.034	-0.195	0.196	0.117	0.026	0.044
DDK4	-0.037	-0.102	0.109	0.013	0.007	-0.172
DDK5	-0.162	-0.055	0.243	0.334	0.027	**0.350**
DDK6	-0.100	0.019	0.032	0.260	0.013	0.158
Sentence 1	0.069	0.008	0.052	0.009	0.100	-0.020
Sentence 2	0.137	0.213	0.167	0.053	0.186	0.170
Sentence 3	0.250	0.190	0.014	0.064	0.338	0.254
Sentence 4	0.200	0.131	0.119	0.073	0.235	0.267
Sentence 5	0.312	0.091	0.064	0.150	0.300	0.331
Sentence 6	0.112	0.039	0.054	0.159	0.126	0.098
Sentence 7	0.055	0.054	0.032	0.140	0.107	0.230
Sentence 8	0.185	0.058	0.075	-0.143	0.197	0.216
Sentence 9	0.058	0.077	-0.145	-0.061	0.287	0.180
Sentence 10	0.107	-0.028	-0.006	0.237	0.212	0.207
Read text	0.087	0.018	-0.034	0.087	0.150	0.222
Monologue	0.117	0.000	0.036	0.250	-0.024	0.112
F. T.	-0.004	0.056	-0.024	0.116	0.111	0.051

RAE: recurrent autoencoder. **F. T.**: Fusion of tasks. **F. F.**: fusion of features.
Best result is highlighted in bold.

results from Figure 9.8 suggests that the proposed approach is suitable to monitor the progression of the dysarthria level in PD patients.

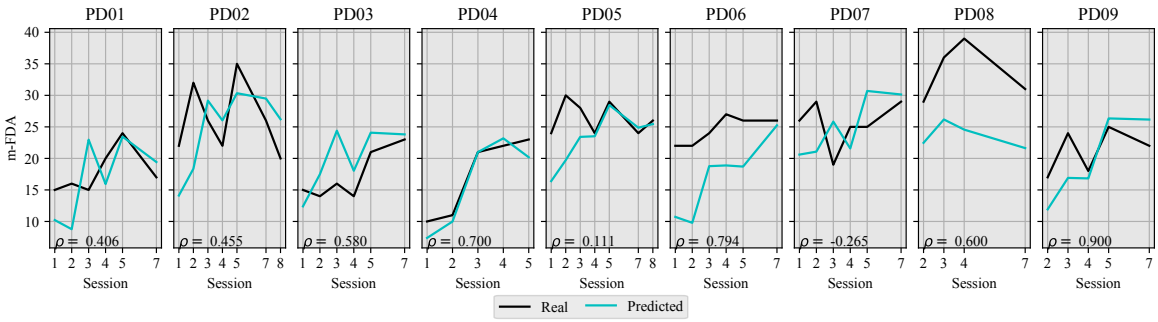

Figure 9.8: Predictions of the m-FDA score of each patient in the Longitudinal corpus with the phonological features and the SVR regression.

The results predicting the dysarthria severity of patients in the At-Home corpus using the SVR regression are shown in Table 9.14. The aim here is not only to predict the dysarthria severity in short-term time intervals but also the influence of the medication intake. The SVR is trained with the data from the Multimodal corpus, excluding those subjects that appear as well in the At-Home data, thus the results shown here corresponds to the ones obtained with an independent test set. The best result is observed as well with the phonological features (ρ =0.491), similar to the results obtained in the longitudinal corpus for long-term progression assessment. In this case, the best results is obtained when all speech tasks are combined using the *Late 2* fusion strategy.

Table 9.14: Results predicting the m-FDA scale of the subjects from the At-Home corpus using different speech feature sets and SVR regressors. Results in terms of the Spearman's correlation coefficient.

	Feature sets							
Speech task	**Phonation**	**Articulation**	**Prosody**	**Phonological**	**RAE**	**F. F. Early**	**F. F. Late 1**	**F. F. Late 2**
DDK1	-0.065	-0.288	0.129	0.245	-0.041	-0.114	-0.038	-0.365
DDK2	-0.032	-0.054	-0.045	0.411	0.001	0.264	0.125	0.462
DDK3	-0.201	-0.232	-0.092	-0.283	0.025	0.237	-0.171	-0.356
Read text	-0.314	0.145	0.308	0.219	-0.017	0.214	0.203	0.331
Monologue	0.000	-0.147	-0.135	-0.160	-0.243	-0.151	0.032	0.439
F. T. Early	0.000	-0.006	-0.180	-0.100	0.017		0.030	0.432
F. T. Late 1	0.000	-0.060	-0.043	-0.107	0.000			
F. T. Late 2	0.000	-0.406	0.334	**0.491**	-0.208			

RAE: recurrent autoencoder). **F. T.**: Fusion of tasks. **F. F.**: fusion of features.
Best result is highlighted in bold.

The results predicting the m-FDA score of the participants in the At-Home corpus using the unsupervised GMM-UBM systems are shown in Table 9.15. The highest correlation (ρ =0.496) is observed as well with the phonological features. In general, the results are similar to the ones obtained with the SVR. The results with the GMM-UBM system are better here than for the case of the Longitudinal corpus probably because in this case there is less variability in the m-FDA scores of the patients i.e., in short-term intervals there is less variation of the m-FDA score than the observed in long-term intervals for each speaker. Hence the GMM-UBM is more able to track these small changes of the m-FDA rather than the big changes that appear in long-term intervals.

Table 9.15: Results predicting the m-FDA scale of the subjects from the At-Home corpus using different speech feature sets and the GMM-UBM models. Results in terms of the Spearman's correlation coefficient.

	Feature sets					
Speech task	**Phonation**	**Articulation**	**Prosody**	**Phonological**	**RAE**	**F. F.**
DDK1	-0.141	0.146	0.002	-0.006	-0.227	0.062
DDK2	-0.131	-0.231	0.074	0.050	-0.182	-0.200
DDK3	-0.052	-0.310	0.123	-0.010	-0.145	-0.355
Read text	0.409	0.043	0.001	0.250	0.257	0.480
Monologue	0.248	0.365	0.303	**0.496**	0.017	-0.024
F. T.	-0.132	-0.092	0.172	0.181	-0.090	-0.199

RAE: recurrent autoencoder). **F. T.**: Fusion of tasks. **F. F.**: fusion of features.
Best result is highlighted in bold.

Figure 9.9 displays the curves comparing the estimated m-FDA scores (cyan lines) and the real labels assigned by the phoniatrician (black lines) for each speaker of the At-Home corpus. The predicted scores follows the trend of the dysarthria level for many of the patients. The largest differences are observed in patients PD02 and PD03. Particularly PD02 is the one with the highest m-FDA score among all participants. In general, the dysarthria progression is predicted with moderate correlation (ρ >0.4) for four of the seven patients. These results suggests that the proposed model is suitable to monitor the progression of the dysarthria level in PD patients in short-term intervals, as well as the influence of the medication intake.

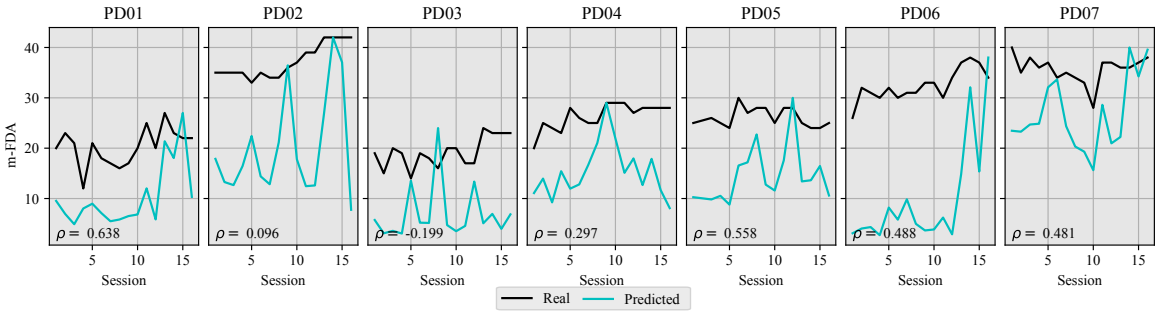

Figure 9.9: Predictions of the m-FDA score of each patient in the At-Home corpus with the phonological features and the GMM-UBM model.

Classification of Patients in Different Levels of the Dysarthria Severity

Although there is a correlation between the predicted and real m-FDA scores, from the clinical point of view it is easier to understand for the patients to know in which stage of the disease they are, rather than to have the prediction of a continuous scale. In addition, for medical applications it is difficult to have a great amount of data to train suitable regression algorithms like an SVR or a CNN, as it was observed in the previous section. For these reasons it should be better to divide the patients into different groups according to their disease severity. In order to perform these experiments the subjects from the Multimodal corpus were grouped into three classes according to their dysarthria severity based on the m-FDA scale. These classes are defined based on the 33rd and 66th percentiles of the total scale in order to discriminate between mild, intermediate, and severe levels of speech impairments associated with hypokinetic dysarthria. For the total m-FDA score the ranges per class are defined as follows: 0 to 16 (mild), 17 to 24 (intermediate), and higher than 25 (severe). The distribution and limits of the scores were chosen in order to have three equal priors for the classes. The distribution is shown in Figure 9.10.

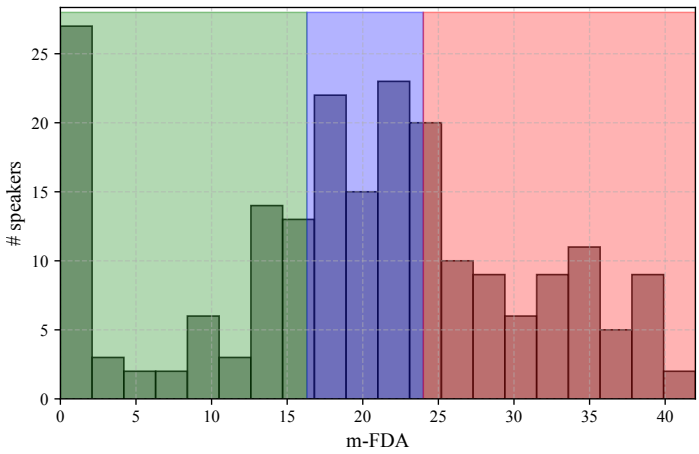

Figure 9.10: Histogram of the m-FDA score of the subjects from the Multimodal corpus and the three groups defined to classify subjects with mild (green), intermediate (blue) and severe (red) dysarthria severity.

With these new labels, a multiclass SVM is trained in a one vs. all strategy to classify the subjects in the three levels of the disease, using all speech features, all speech tasks, and their combinations both at feature and task-level. The results are presented in Table 9.16. As in the bi-class problems, the OpenSMILE features are included only as a baseline, and they are not included in the fusion strategies. The multi-class classification is a more challenging problem due to the distribution of classes. Results of the best feature set per task are highlighted in bold. The highest UAR is observed with the OpenSMILE features in the monologue (55.5%). These results indicate that non of the proposed feature sets neither their combinations were as good as the baseline for this particular and more challenging classification problem. In addition, the fusion of the speech tasks and feature sets does not improve the results.

Table 9.16: Results classifying subjects from the Multimodal corpus in different dysarthria levels using speech features sets and SVM classifiers. Results in terms of UAR [%].

	Feature sets						F. F.	F. F.
Task	**Phonation**	**Articulation**	**Prosody**	**OpenSMILE**	**Phonological**	**RAE**	**Early**	**Late 1**
DDK1	45.2 (9.2)	**48.8 (13.9)**	46.7 (8.5)	47.4 (5.1)	48.1 (10.1)	45.8 (6.3)	46.4 (6.3)	49.7
DDK2	46.4 (8.9)	**50.5 (9.7)**	46.5 (9.1)	50.2 (8.2)	46.8 (5.1)	48.0 (8.0)	48.8 (8.9)	49.7
DDK3	46.3 (7.3)	48.8 (7.7)	43.2 (9.9)	47.2 (7.7)	**53.9 (7.6)**	42.2 (5.3)	46.7 (10.8)	46.9
DDK4	43.5 (12.6)	42.8 (6.9)	45.8 (10.2)	**52.0 (5.9)**	49.1 (9.8)	45.2 (7.2)	48.4 (11.0)	47.4
DDK5	47.3 (8.7)	44.7 (8.8)	46.7 (9.6)	48.8 (3.4)	**49.8 (10.7)**	44.3 (4.3)	44.3 (7.2)	45.4
DDK6	42.6 (9.1)	45.5 (7.3)	43.4 (5.8)	46.3 (6.8)	50.4 (3.3)	46.8 (12.5)	**52.7 (8.2)**	47.4
Sentence 1	39.1 (11.4)	45.6 (11.2)	47.7 (8.4)	**48.6 (7.6)**	43.3 (8.1)	44.1 (6.9)	46.3 (9.0)	46.9
Sentence 2	39.8 (7.1)	47.1 (9.6)	**51.7 (12.2)**	42.0 (7.9)	40.0 (12.1)	44.4 (11.2)	46.0 (8.2)	47.9
Sentence 3	44.1 (8.0)	48.1 (7.1)	42.7 (14.4)	44.1 (6.4)	46.3 (10.3)	38.6 (12.6)	**48.7 (9.3)**	44.8
Sentence 4	42.0 (9.7)	45.4 (10.4)	43.9 (12.49	44.2 (6.6)	44.3 (8.6)	47.2 (10.3)	46.0 (7.7)	**47.3**
Sentence 5	36.5 (8.5)	48.4 (8.2)	46.3 (10.0)	**49.6 (9.6)**	42.4 (6.0)	41.2 (10.5)	45.9 (8.1)	40.7
Sentence 6	41.5 (9.3)	46.7 (11.8)	45.7 (5.5)	45.1 (9.5)	44.0 (7.6)	39.1 (9.6)	43.2 (7.2)	**47.8**
Sentence 7	45.2 (9.0)	46.1 (5.4)	47.0 (8.1)	48.2 (9.0)	**50.9 (15.8)**	48.7 (8.6)	46.5 (9.7)	46.3
Sentence 8	44.0 (9.0)	46.5 (8.5)	49.1 (13.6)	**54.4 (6.0)**	48.4 (9.5)	41.2 (3.1)	35.2 (3.8)	44.3
Sentence 9	43.6 (10.7)	44.6 (8.8)	42.5 (11.0)	46.6 (5.4)	45.2 (9.2)	38.5 (7.7)	43.6 (5.4)	**49.4**
Sentence 10	42.4 (4.8)	33.3 (0.0)	42.9 (11.9)	44.5 (11.0)	48.6 (12.4)	42.5 (7.7)	46.7 (9.9)	**52.2**
Read text	41.2 (11.6)	42.5 (10.3)	44.7 (14.0)	**52.7 (11.2)**	47.8 (10.4)	45.4 (8.3)	48.6 (9.1)	47.7
Monologue	43.1 (4.6)	50.4 (10.0)	45.7 (9.5)	**55.5 (9.2)**	50.6 (11.6)	44.7 (9.0)	46.9 (12.2)	45.3
F.T. Early	**49.8 (6.5)**	47.9 (7.1)	40.8 (6.8)	45.7 (8.4)	42.7 (8.7)	46.5		
F.T. Late 1	47.6	**52.6**	42.5	48.2	43.1	48.8	47.7	45.3

RAE: recurrent autoencoder. **F.T.**: Fusion of tasks. **F.F.**: fusion of feature sets.
Results of the best feature set per task are highlighted in bold.

The multi-class experiments are also performed with the ResNet-based models with the aim to improve the results. The results are shown in Table 9.17. An UAR up to 57.3% was obtained with this approach, which improves in 1.8% the results reported previously with the OpenSMILE features and the SVM classifier. The best result is obtained here with the combination of the speech tasks using the *Late 1* fusion strategy. Additionally, higher UARs are obtained using the ResNet models trained with the full spectrograms, rather than the ones computed only for onset transitions, which is contrary to the results observed for the bi-class problems (see Table 9.3). This is explained because the multi-class classification is a more challenging problem that needs more data to achieve better results.

The summary of the best results obtained using the different methods is shown in Table 9.18. The results include additional performance metrics like the weighted

Table 9.17: Results classifying subjects from the Multimodal corpus in different dysarthria levels using the deep learning models to model speech signals. Results in terms of UAR [%].

Task	ResNet Trasitions	ResNet Full
DDK1	53.4	52.5
DDK2	51.1	54.5
DDK3	46.5	55.6
DDK4	48.0	51.6
DDK5	49.5	55.4
DDK6	45.2	50.4
Sentence 1	47.9	57.2
Sentence 2	48.0	53.4
Sentence 3	51.4	51.9
Sentence 4	41.2	46.7
Sentence 5	40.8	51.8
Sentence 6	50.7	55.0
Sentence 7	47.4	49.5
Sentence 8	47.8	52.9
Sentence 9	45.4	50.4
Sentence 10	41.7	50.5
Read Text	48.7	52.0
Monologue	43.0	55.7
F. T. Late 1	50.4	**57.3**

accuracy, the F-score, and the accuracies obtained for each of the three classes. As it was stated previously, the best result is obtained with the ResNet-based models.

Table 9.18: Best results obtained for each method classifying subjects from the Multimodal corpus in different dysarthria levels according to the m-FDA score.

Feature set	Task	ACC	F-score	UAR	ACC Mild	ACC Intermediate	ACC Severe
Phonation	F. T. Early	50.7	0.492	49.7	61.0	40.0	49.0
Articulation	DDK2	52.3	0.511	52.6	78.0	34.0	45.0
Prosody	Sentence 2	51.6	0.510	51.7	64.0	38.0	54.0
OpenSMILE	Monologue	55.3	0.538	55.5	65.0	39.0	62.0
Phonological	DDK3	54.2	0.537	53.9	61.0	53.0	48.0
RAE	F. T. Late 1	48.2	0.442	48.7	88.0	16.0	42.0
F. F. Early	DDK6	52.7	0.506	52.7	67.0	29.0	61.0
F. F. Late 1	Sentence 10	51.8	0.507	52.2	76.0	35.0	45.0
ResNet Full	**F. T. Late 1**	**57.3**	**0.573**	**57.3**	**67.0**	**52.0**	**53.0**

ACC [%]: accuracy, **UAR [%]**: unweighted average recall,
RAE: recurrent autoencoder. **F. T.**: Fusion of tasks. **F. F.**: fusion of features.

Additional insights can be observed in Table 9.18 about why such a model produces the highest global accuracy. The ResNet is the model that provides the best accuracy to classify patients in intermediate stage of the disease, which is the most misclassified class for the other methods, except for the phonological features. Methods like the late fusion of features are the most accurate to classify the extreme classes (patients in initial and severe stages of the disease); however they are very inaccurate to classify patients in intermediate stage of the disease. This make sense because such particular methods were very accurate in the bi-class problems (see Table 9.6).

The confusion matrices from Figure 9.11 show the top-3 obtained results for the classification of patients in three levels of speech impairments. The ResNet model in

Figure 9.11a) shows the best results because it is the most balanced to classify the three classes. Most of the classification errors in the ResNet model correspond to patients in mild and severe stages that are classified in intermediate stage. This is very positive since they are not mainly misclassified in the other extreme class.

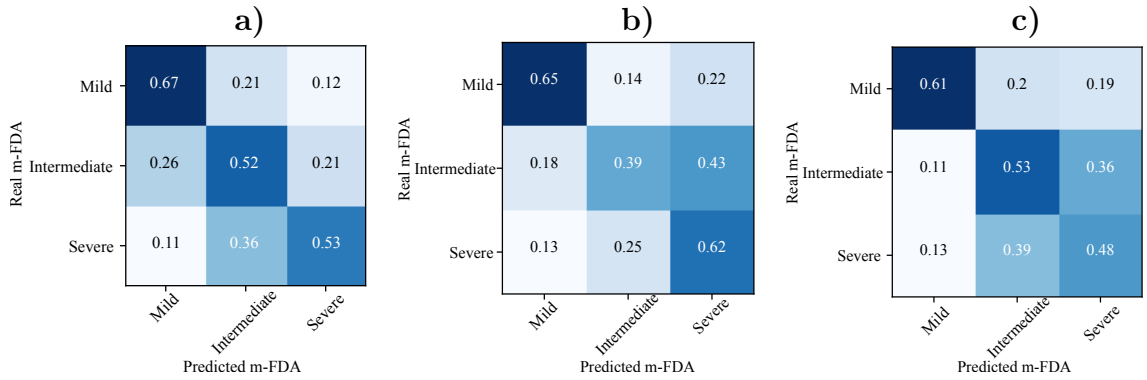

Figure 9.11: Details of the best result obtained classifying subjects from the Multimodal corpus in different dysarthria levels according to the m-FDA score. **a)** Fusion of speech tasks in the ResNet Full model. **b)** OpenSMILE features computed in the monologue. **c)** Phonological features from the DDK3 exercise.

9.1.3 Automatic Evaluation of the Motor State of Patients

The evaluation of the motor state of patients is based on the MDS-UPDRS-III scale, and it is performed in two scenarios: (1) the prediction of the value of the MDS-UPDRS-III scale using regression strategies and (2) the classification of patients in different levels of the disease severity (mild, intermediate, severe) using multi-class classification methods. The results obtained for the regression approach using the different feature sets are shown in Table 9.19. The results are presented in terms of the Spearman's correlation coefficient between the predicted and real scores assigned to each patient. The highest correlations are observed with the phonological features, although none of the models is accurate enough to evaluate the motor -state severity of the patients with a strong (> 0.6) or even moderate (> 0.4) correlations. This is expected since the MDS-UPDRS-III is a complete motor scale in which only one of the items is related to speech symptoms. Hence it is not suitable nor fair to try to evaluate the full motor-state severity of patients using only speech signals. However, these speech features could provide complementary information when they are combined with the handwriting and gait signals. These aspects are addressed in Section 9.4.

Similar to the experiments performed to classify patients in different dysarthria levels, three classes are defined based on the 33rd and 66th percentiles of the total MDS-UPDRS-III to discriminate between mild, intermediate, and severe levels of motor-state severity. For the total MDS-UPDRS-III score the ranges per class are defined as follows: 0 to 25 (mild), 26 to 40 (intermediate), and higher than 40 (severe). The distribution and limits of the scores were chosen in order to have three equal priors for the classes. The distribution is shown in Figure 9.12.

Table 9.19: Results estimating the MDS-UPDRS-III scale of the patients from the Multimodal corpus using different speech feature sets and SVR regressors. Results in terms of the Spearman's correlation coefficient.

	Feature sets						F. F.	F. F.	F. F.
Task	**Phonation**	**Articulation**	**Prosody**	**OpenSMILE**	**Phonological**	**RAE**	**Early**	**Late 1**	**Late 2**
DDK1	-0.210	0.159	0.090	0.176	0.252	0.192	0.214	0.227	0.274
DDK2	-0.287	-0.146	0.073	0.196	0.120	0.059	0.206	0.079	0.231
DDK3	-0.017	-0.266	0.046	0.102	-0.243	-0.289	0.070	0.075	-0.362
DDK4	-0.262	0.124	0.051	0.069	0.149	-0.086	0.095	0.074	0.176
DDK5	-0.239	-0.120	0.153	-0.253	0.174	-0.259	-0.272	0.169	-0.327
DDK6	-0.299	-0.317	-0.221	0.202	0.099	-0.335	0.200	0.162	-0.340
Sentence 1	-0.130	-0.340	-0.137	0.172	0.138	-0.252	0.078	-0.012	-0.276
Sentence 2	-0.249	-0.333	-0.253	0.170	0.100	-0.089	0.046	0.050	-0.383
Sentence 3	0.162	-0.005	-0.404	0.009	0.276	-0.317	-0.016	-0.065	-0.361
Sentence 4	-0.248	0.075	-0.391	0.159	0.067	0.135	0.152	-0.050	-0.328
Sentence 5	-0.244	-0.010	-0.353	0.138	0.189	-0.310	0.101	0.072	-0.381
Sentence 6	-0.167	-0.009	0.125	0.018	0.165	-0.379	0.004	-0.012	-0.309
Sentence 7	-0.281	-0.132	-0.345	0.033	0.206	-0.037	-0.085	0.068	-0.320
Sentence 8	-0.230	-0.330	-0.326	0.004	0.048	0.217	-0.015	0.151	-0.327
Sentence 9	-0.324	0.000	-0.341	0.285	0.367	-0.119	0.318	-0.204	-0.405
Sentence 10	-0.303	-0.182	0.019	0.014	0.129	-0.302	0.021	-0.100	-0.394
Read text	-0.045	-0.063	-0.300	0.099	0.174	0.065	0.086	-0.058	-0.255
Monologue	-0.070	-0.227	-0.428	0.229	0.102	-0.246	0.200	0.142	-0.396
F. T. Early	-0.060	-0.269	-0.176	0.126	0.149	0.081			
F. T. Late 1	-0.324	-0.064	0.017	0.213	0.255	0.106	0.159	0.105	
F. T. Late 2	-0.501	-0.466	-0.531	0.253	0.334	-0.495	0.280		-0.491

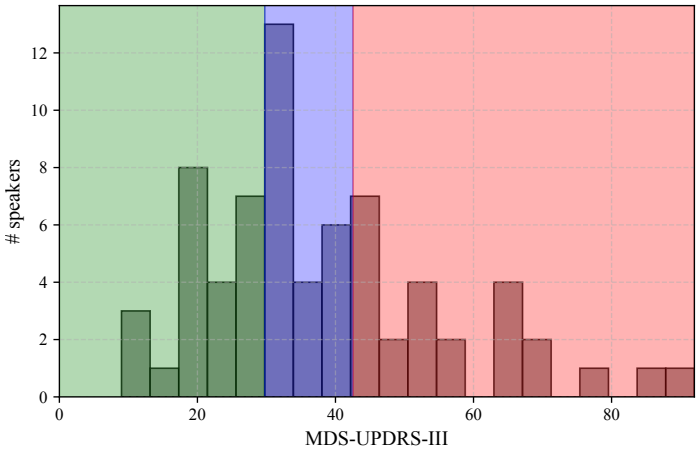

Figure 9.12: Histogram of the MDS-UPDRS-III score of the subjects from the Multimodal corpus and the three groups defined to classify subjects with mild (green), intermediate (blue) and severe (red) motor-state severity.

The results of the multi-classification are shown in Table 9.20. The highest UAR is observed with the fusion of the speech tasks and the phonation features (49.1%). In general, none of the models is accurate enough to discriminate the patients in three levels of the motor-state severity, similar to the observed in the regression experiment. However, these speech features could provide complementary information when they are combined with the handwriting and gait signals to model the general motor-state severity of the subjects. This multi-class classification problem was also addressed with the ResNet-based models. However, the results were also not satisfactory.

Table 9.20: Results classifying subjects from the Multimodal corpus in different motor state levels using speech features and SVM classifiers. Results in terms of UAR [%].

	Feature sets						F. F.	F. F.
Task	**Phonation**	**Articulation**	**Prosody**	**OpenSMILE**	**Phonological**	**RAE**	**Early**	**Late 1**
DDK1	29.3 (9.8)	33.7 (11.9)	33.9 (13.9)	38.7 (15.5)	36.8 (14.7)	31.3 (11.5)	39.4 (17.3)	36.6
DDK2	33.0 (14.9)	38.3 (12.8)	42.4 (9.6)	42.0 (9.2)	36.2 (12.4)	31.1 (6.7)	33.7 (7.9)	42.1
DDK3	39.0 (12.9)	31.8 (10.1)	33.6 (8.9)	31.9 (14.2)	36.0 (12.4)	33.9 (1.7)	36.1 (12.3)	33.9
DDK4	41.8 (13.7)	38.1 (15.2)	31.1 (11.6)	31.7 (8.7)	39.7 (7.6)	41.2 (13.4)	32.5 (9.4)	42.4
DDK5	34.7 (6.8)	33.5 (6.7)	35.1 (8.0)	33.3 (0.0)	39.2 (6.0)	31.8 (11.8)	33.3 (0.0)	37.7
DDK6	33.3 (0.0)	33.9 (8.1)	33.3 (0.0)	34.2 (8.4)	33.1 (8.8)	33.3 (0.0)	35.4 (13.0)	32.5
Sentence 1	32.9 (7.5)	39.1 (11.2)	39.6 (11.2)	37.3 (8.2)	35.7 (9.0)	31.3 (7.4)	39.0 (6.2)	34.0
Sentence 2	33.7 (13.9)	35.2 (6.9)	36.9 (8.4)	40.4 (10.4)	36.6 (12.1)	38.6 (10.7)	32.3 (11.9)	37.1
Sentence 3	40.1 (12.9)	33.6 (66.7)	31.7 (12.9)	30.9 (3.9)	39.8 (8.8)	36.2 (10.0)	39.4 (8.9)	39.6
Sentence 4	44.2 (10.1)	36.3 (9.9)	33.9 (1.7)	43.7 (17.9)	32.9 (1.3)	34.2 (2.7)	39.8 (9.5)	43.3
Sentence 5	32.7 (6.9)	31.7 (12.9)	32.9 (10.2)	41.8 (10.4)	39.2 (11.2)	31.9 (11.1)	44.2 (7.6)	37.1
Sentence 6	34.2 (7.2)	44.0 (11.2)	44.3 (13.0)	41.3 (9.0)	37.8 (16.0)	30.8 (10.6)	42.6 (9.3)	43.3
Sentence 7	37.4 (13.3)	39.0 (11.7)	38.3 (9.5)	33.2 (0.3)	37.1 (8.9)	33.6 (12.1)	29.4 (9.9)	42.5
Sentence 8	38.8 (13.4)	42.6 (8.4)	37.4 (9.4)	36.8 (5.7)	41.0 (9.1)	42.7 (13.1)	40.7 (14.5)	46.5
Sentence 9	42.3 (17.6)	36.0 (4.4)	32.2 (9.4)	32.2 (8.2)	40.1 (6.9)	31.3 (7.3)	41.3 (7.2)	42.0
Sentence 10	32.1 (10.4)	32.7 (9.4)	35.4 (11.1)	33.3 (0.0)	33.2 (0.3)	33.9 (11.8)	33.9 (1.6)	32.5
Read text	36.7 (11.6)	37.6 (9.1)	38.7 (9.0)	40.1 (8.1)	32.4 (8.9)	34.2 (2.7)	41.2 (8.6)	38.9
Monologue	39.2 (12.1)	36.7 (12.3)	40.7 (10.9)	33.6 (11.9)	42.3 (9.1)	32.8 (1.7)	36.0 (12.8)	38.4
F. T. Early	38.6 (5.3)	38.5 (9.9)	38,238	33.3 (0.0)	40.0 (11.5)	34.0 (5.5)		
F. T. Late 1	49.1	37.9	42.4	42.0	42.1	39.5	46.5	45.2

9.2 Handwriting Assessment

The handwriting assessment of PD patients is performed to cover the discrimination between PD patients and HC subjects and the evaluation of the motor-state severity of the patients based on the MDS-UPDRS-III scale. The estimation of the motor-state severity includes as well the longitudinal evaluation of PD patients from the Longitudinal corpus. The handwriting assessment of the patients is always performed considering both the kinematic, geometric, and in-air features, explained in Chapter 5 as well the deep learning models to process the offline and online handwriting data. The analysis includes also the different writing and drawing tasks like sentences written by the patients and Archimedean spirals, among others. The fusion at feature and task-levels is performed with three methods described in Section 7.2.

9.2.1 Automatic Classification of Parkinson's Disease Patients

The results obtained classifying PD patients vs. HC subjects from the Multimodal corpus (Section 3.3.1) are presented in Table 9.21 using the different feature sets and handwriting tasks. The feature sets include kinematic, geometric, in-air, and their combinations. The results are presented in terms of the UAR to avoid bias due to the unbalance in the groups. Results of the best feature set per task are highlighted in bold.

The highest UAR is obtained in the fusion of all handwriting tasks and all feature sets, using the early fusion for the features and the *Late 2* fusion for the tasks (97.8%). This was the same strategy that produced the highest accuracy for the case of speech signals (see Table 9.1). This confirms that early fusion is better to model the different and complementary information produced by each feature set, and also that late fusion methods deal better with the redundant information that appears when using the same features computed from different tasks. The analysis of the

Table 9.21: Results classifying PD patients vs. HC subjects from the Multimodal corpus using different handwriting feature sets and SVM classifiers. Results in terms of UAR [%].

	Feature sets					
Task	**Kinematic**	**Geometric**	**In-air**	**F.F Early**	**F.F Late 1**	**F.F Late 2**
Alphabet	78.6 (11.9)		**85.1 (7.1)**	77.3 (12.2)	80.2	80.8
Circle	75.1 (10.5)		72.5 (8.9)	74.8 (9.6)	75.1	**78.0**
Circle template	67.4 (14.0)		72.8 (6.8)	68.5 (13.5)	67.3	**76.0**
Cube	72.7 (12.1)		79.9 (9.1)	78.1 (10.9)	75.7	**81.3**
Free writing	67.9 (13.4)		**79.4 (8.3)**	70.6 (11.7)	74.0	78.4
House	74.2 (12.2)		**83.5 (9.3)**	77.2 (8.6)	78.6	80.0
ID	69.7 (14.0)		74.4 (11.9)	73.8 (9.4)	69.1	**77.6**
Name	73.1 (13.7)		75.4 (7.2)	74.4 (12.8)	74.0	**78.9**
Numbers	75.6 (5.2)		71.4 (7.7)	75.5 (8.3)	**77.7**	76.6
Rectangles	70.5 (8.3)		74.3 (12.1)	65.8 (8.7)	68.8	**75.0**
Rey Figure	70.9 (9.3)		**86.0 (12.4)**	75.1 (10.5)	80.3	82.3
Signature	73.0 (11.2)		**74.8 (11.2)**	74.7 (11.7)	72.7	72.7
Spiral	74.3 (13.1)	61.5 (8.2)	66.3 (9.4)	73.9 (12.6)	72.5	**74.6**
Spiral template	68.2 (9.6)	56.4 (9.0)	69.1 (10.4)	72.0 (6.1)	67.4	**76.8**
F.T Early	85.4 (11.6)		**96.2 (5.3)**			
F.T Late 1	85.9		86.4	**90.2**	85.9	
F.T Late 2	86.1		86.8	**97.8**		87.8

results for the individual tasks indicates that for many of the cases the best result is obtained combining the kinematic and in-air features, specially for the simple drawing tasks like Circle, Circle template, Cube, Rectangles, Spiral, and Spiral template. For more complex tasks like the Alphabet, the Free writing, and the Rey figure the best results are obtained with the in-air features. These more complex tasks have a lot of pen-up and pen-down transitions, which are particularly well modeled with the in-air features. Conversely, the simple drawing tasks like the spiral and the circle do not contain that high amount of transition movements, making the fusion of kinematic and in-air features more appropriate for the analysis. Unfortunately, the geometric features do not provide the expected accuracy. Additional analyses could be performed in such direction to model the trajectory and accuracy of the strokes performed by the patients when drawing different geometrical shapes.

The classification is also performed using the different deep learning methods explained in Section 5.5 to model both offline and online handwriting. The results are depicted in Table 9.22 and indicate that it is possible to classify PD patients and HC subjects with accuracies of up to 99.2% using the CNN model to process the online handwriting data from the pen-up and pen-down transitions. The best result is always obtained by combining the outputs of the different handwriting tasks, similar to the previous case with the feature extraction and classification. The results obtained for the offline handwriting model are not that accurate compared to the ones observed with the online models, indicating that there is important information in the temporal variability and structure of the handwriting process that is not available when considering only the spatial attributes of the figures and the characters written by the patients.

The comparison between the results observed for both online models (transitions and full segments) indicates that there are important differences depending on the

Table 9.22: Results classifying PD patients vs. HC subjects from the Multimodal corpus using deep learning strategies to model handwriting signals. Results in terms of UAR [%].

Task	Offline SqueezeNet	Online CNN Transition segments	Online CNN Full segments
Alphabet	61.4	**96.2**	92.9
Circle	55.5	72.9	**85.9**
Circle template	56.2	80.3	**93.7**
Cube	62.8	92.7	**95.2**
Free writing	61.6	84.4	**86.3**
House	66.7	92.5	**93.5**
ID	64.2	**95.3**	94.7
Name	61.3	**94.1**	93.7
Numbers	61.9	92.4	**93.0**
Rectangles	60.0	86.5	**92.1**
Rey Figure	64.1	**97.0**	96.6
Signature	71.2	71.5	**80.4**
Spiral	59.0	**61.0**	57.6
Spiral template	59.9	**64.3**	52.3
F.T Late 1	71.7	97.6	**98.8**
F.T Late 2	76.2	**99.2**	98.8

addressed handwriting task. Simple drawing shapes such as circles, cubes, and rectangles are better modeled with the Full segments. Conversely, more complex tasks like the alphabet are better modeled with the network that only consider the pen-up and pen-down transitions. This is similar to the results previously observed with the in-air features, since they were more accurate to model the most complex writing tasks. Finally, the results observed with the Archimedean spirals are not as good as expected, specially considering that this is one of the most used handwriting tasks for the assessment of PD patients in the literature. This aspect has to be carefully considered when designing evaluation protocols for the assessment of the disease.

The summary of the best results obtained for the classification is shown in Table 9.23, and include additional performance metrics like sensitivity, specificity, F-score, and the area under the ROC curve. As it was stated previously, the best results are always obtained when the different handwriting tasks are combined, specially using late fusion strategies. The best results are obtained with the deep learning model to process the online handwriting data from the pen-up and pen-down transitions, followed by the early fusion of in-air and kinematic features.

The confusion matrices, ROC curves, and histograms of the predictions for these top-2 models are observed in Figures 9.13 and 9.14. Note the high separability of the scores of the histograms both for the deep learning model in Figure 9.13c) and for the fusion of kinematic and in-air features in Figure 9.14c).

The sensitivity of 98% observed for the deep learning model in Figure 9.13a) shows that only two of the PD patients were misclassified. The first one is a 67 years old male PD patient with MDS-UPDRS=31. The second one corresponds to a 34 years old PD patient with MDS-UPDRS=17. This subject is the youngest patient in the corpus, and probably a patient with a similar age was never observed in the training set. In addition, this second patient has a low MDS-UPDRS-III score, compared to the rest

Table 9.23: Best results obtained for each method classifying PD patients and HC subjects in the Multimodal corpus using handwriting signals.

Feature set	Task	ACC	Fscore	UAR	SENS	SPEC	AUC
Kinematic	F.T Late 2	87.2	0.848	86.1	88.6	83.7	0.948
In-air	F.T Early	96.1	0.955	96.2	95.7	96.7	0.996
F.F Early	F.T Late 2	97.1	0.967	97.8	95.7	100.0	0.996
F.F Late 1	F.T Late 1	87.8	0.853	85.9	90.2	81.6	0.962
F.F Late 2	F.T Late 2	86.0	0.843	87.8	83.7	91.8	0.961
Online CNN-GRU	**F.T Late 2**	**98.8**	**0.985**	**99.2**	**98.4**	**100.0**	**0.999**

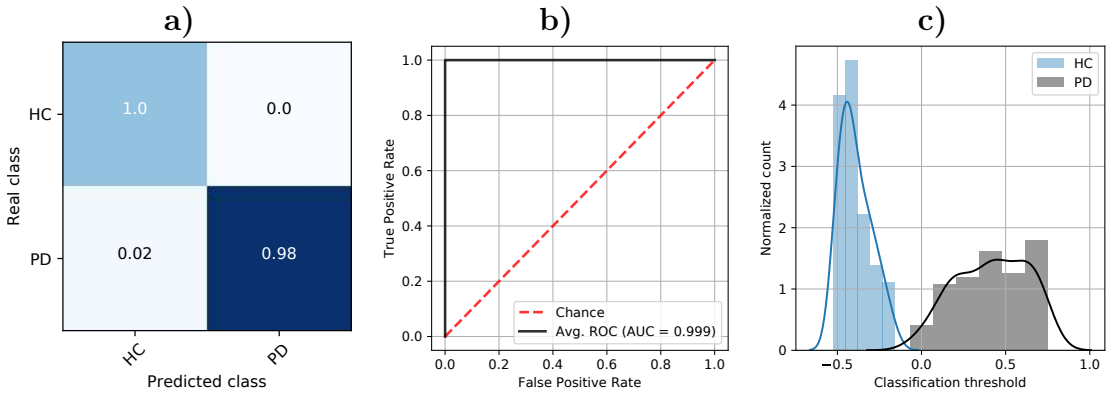

Figure 9.13: Details of the best result obtained classifying PD patients and HC subjects from the Multimodal corpus using the deep learning methods. **a)** Normalized confusion matrix. **b)** ROC curve. **c)** Distribution of the classification scores.

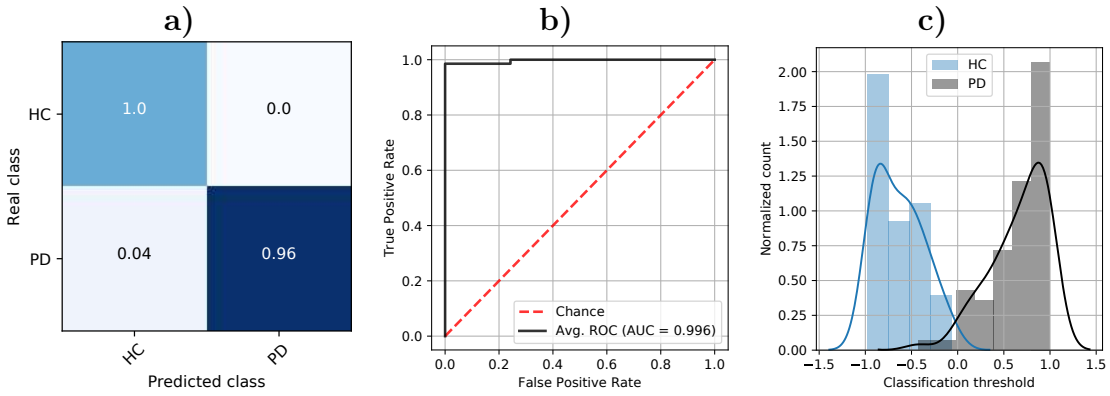

Figure 9.14: Details of the best result obtained classifying PD patients and HC subjects from the Multimodal corpus using the fusion of kinematic and in-air features. **a)** Normalized confusion matrix. **b)** ROC curve. **c)** Distribution of the classification scores.

of the patients in the corpus. Examples of the alphabet written by both misclassified patients are seen in Figure 9.15. The alphabet was chosen for the example because it was one of the most accurate tasks for the classification. Note that neither the writing of both subjects exhibit the specific aspects associated to PD dysagraphia, like tremor or micrographia. In addition, it is possible to understand and to read the

characters written by the patients, making these samples more similar to the ones observed for HC subjects.

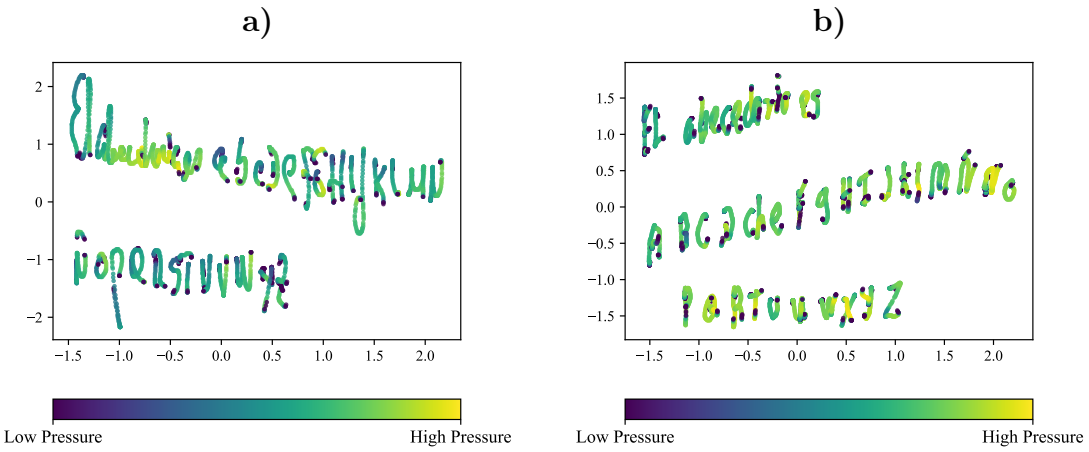

Figure 9.15: Alphabet written by the two misclassified PD patients from the Multimodal corpus. **a)** 67 years old male PD patient with MDS-UPDRS-III: 31. **b)** 34 years old Female PD patient with MDS-UPDRS-III: 17.

9.2.2 Automatic Evaluation of the Motor State Severity of Patients

The evaluation of the motor state severity of patients is based on the MDS-UPDRS-III scale. The evaluation of the motor state severity is performed in three scenarios with the handwriting signals: (1) the prediction of the value of the MDS-UPDRS-III scale using the handwriting models described in Chapter 5 and regression algorithms; (2) the progression evaluation of the patients from the Longitudinal corpus using both user models based on GMM-UBM systems and regression algorithms; and (3) the classification of patients in different levels of the disease severity (mild, intermediate, severe) using multi-class classification methods.

Motor State Evaluation based on Regression Algorithms

The evaluation of the motor-state severity is addressed considering the different handwriting features and tasks, and their combinations. The handwriting features include the kinematic, geometric, and in-air models, in addition to the end-to-end modeling using deep learning techniques. The regression is performed using an SVR algorithm. The results obtained estimating the MDS-UPDRS-III score from the Multimodal corpus are presented in Table 9.24. The results are presented in terms of the Spearman's correlation coefficient.

The highest correlation is obtained with the early fusion of the kinematic features for all handwriting tasks (ρ =0.616). Unfortunately, the in-air features are not that accurate to evaluate the global motor-severity as they were for the classification experiments. This is likely because of although these features model well known

Table 9.24: Results estimating the MDS-UPDRS-III scale of the subjects from the Multimodal corpus using handwriting features and SVR regressors. Results in terms of the Spearman's correlation coefficient.

	Feature set					
Task	**Kinematic**	**Geometric**	**In-air**	**F.F Early**	**F.F Late 1**	**F.F Late 2**
Alphabet	0.101		-0.247	0.015	0.060	-0.160
Circle	0.140		0.071	0.208	0.125	0.165
Circle template	0.105		0.163	0.059	0.125	0.018
Cube	0.070		-0.121	0.107	-0.120	-0.127
Free writing	0.080		-0.015	0.041	-0.011	0.066
House	0.206		0.045	0.211	0.174	0.100
ID	-0.042		-0.379	-0.104	-0.148	-0.095
Name	-0.201		-0.409	-0.356	-0.341	-0.368
Numbers	0.340		-0.380	0.300	0.095	0.201
Rectangles	0.182		0.040	0.158	0.137	0.092
Rey Figure	-0.475		-0.443	-0.446	-0.461	-0.471
Signature	0.052		-0.472	-0.052	-0.249	-0.242
Spiral	0.154	0.184	-0.145	0.133	0.001	0.050
Spiral template	0.453	0.107	-0.090	0.431	0.301	0.320
F.T Early	**0.616**		0.253			
F.T Late 1	0.328		-0.087	0.524	0.113	
F.T Late 2	0.532		-0.535	0.447		0.248

symptoms to characterize the presence of the disease like tremor and rigidity, these are not the most sensitive aspects included in the MDS-UPDRS-III scale to map the global motor performance of the patients.

The estimation of the MDS-UPDRS-III scale is also performed using the deep learning models. Only the models based on online handwriting were considered here because they were the most robust in the classification experiments. The results are observed in Table 9.25. Unfortunately, the results are not that accurate as the ones obtained in the classification experiments, similar to what occurs for the speech signals predicting the dysarthria level of the participants. The main reason would be the lack of labeled data about the MDS-UPDRS-III score of the patients to train the regression models.

A summary of the best results obtained with each method is shown in Table 9.26, including additional metrics like the Pearson's correlation coefficient and the MAE. The results include also the p-values of the correlations to test whether the correlations are significant or not. The best result was obtained with the kinematic features, as it was mentioned previously. In addition, the best results are achieved when considering all handwriting tasks together.

Figure 9.16 shows the errors in the evaluation of the MDS-UPDRS-III scores. The displayed result corresponds to the ones obtained with the kinematic features combining all handwriting task using the early fusion. Although the result is satisfactory (strong correlation), other regression strategies and feature sets can be considered to improve the correlation values.

Table 9.25: Results estimating the MDS-UPDRS-III scale of the subjects from the Multimodal corpus using handwriting signals and deep learning methods. Results in terms of the Spearman's correlation coefficient.

Task	**Online CNN Transition segments**	**Online CNN Full segments**
Alphabet	0.065	0.249
Circle	0.191	0.050
Circle template	-0.121	0.372
Cube	0.021	0.082
Free writing	0.011	0.125
House	-0.008	0.031
ID	0.066	0.135
Name	-0.010	0.158
Numbers	0.037	-0.042
Rectangles	-0.037	0.229
Rey Figure	0.018	0.083
Signature	-0.099	0.100
Spiral	-0.036	0.070
Spiral template	0.051	0.112
F.T Late 1	0.054	0.087
F.T Late 2	0.239	0.254

Table 9.26: Best results obtained for each method to evaluate the MDS-UPDRS-III scale of the subjects in the Multimodal corpus using handwriting signals.

Feature set	**Task**	r	**p-val** r	ρ	**p-val** ρ	**MAE**
Kinematic	**F.T. Early**	**0.586**	**$\ll$0.005**	**0.616**	**$\ll$0.005**	**8.5**
In-air	F.T. Early	0.238	0.103	0.269	0.065	8.7
F.F Early	F.T Late 1	0.502	$\ll$0.005	0.524	$\ll$0.005	8.7
F.F Late 1	Spiral template	0.310	0.003	0.301	0.005	8.5
F.F Late 2	Spiral template	0.362	0.001	0.320	0.003	10.2
Online CNN-GRU	F.T Late 2	0.284	0.007	0.254	0.017	10.0

Longitudinal Assessment of Patients

Two methods are considered to model the disease progression of the patients from the Longitudinal corpus based on the MDS-UPDRS-III scale. The first one comprises the use of an SVR trained with the data from the Multimodal corpus. The patients from the longitudinal data will act as an independent hidden test set for the regression method. The second approach consists of applying user models based on the GMM-UBM approach introduced in [Aria 18a] and extended in [Vasq 20b] to model as well handwriting signals. The UBMs are trained using information from the HC subjects from the Multimodal corpus. Then specific GMMs are adapted for each patient in each session from the Longitudinal corpus. The results are shown in Table 9.27. Only the kinematic features were considered for the analysis because they were the most accurate for the regression problem in the previous subsection. The best results are generally obtained with the SVR approach, similar to the results obtained with speech signals to evaluate the disease progression based on the m-FDA score (see

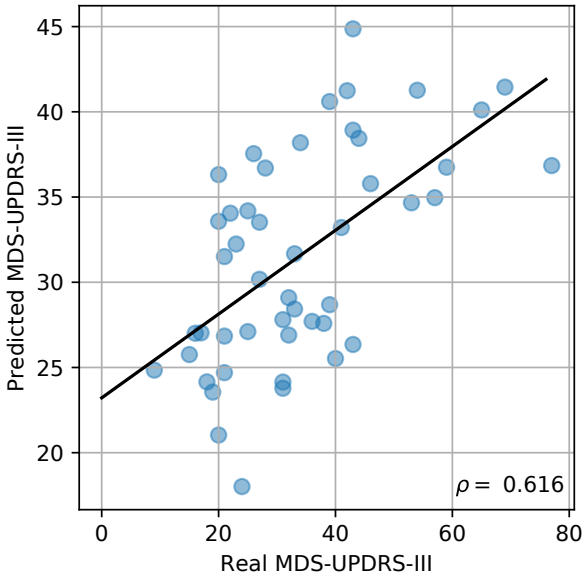

Figure 9.16: Details of the best result obtained estimating the MDS-UPDRS-III scale of the subjects in the Multimodal corpus using kinematic features and SVR regressors.

Tables 9.12 and 9.13). The main reason is because here there are more data to train the regression models, thus obtaining more accurate results using a supervised model like the SVR instead of the unsupervised model based on GMM-UBM. A strong correlation is obtained (ρ =0.704) with the SVR approach and combining all handwriting tasks using the *Late 2* fusion strategy. High correlations are also obtained with tasks related to writing exercises like Numbers (ρ =0.627), Name (ρ =0.562), ID (ρ =0.449), and Free writing (ρ =0.497). These tasks are the longest to perform by the patients and contain a lot of strokes that are better to model the disease severity of the patient.

Figure 9.17 displays curves with the comparison of the predicted MDS-UPDRS-III scores (cyan lines) and the real labels assigned by the neurologist (black lines) for each of the nine speakers of the Longitudinal corpus, using both the SVR and the GMM-UBM systems. The horizontal axis represents the recording session. The lines for each speaker represent the progression of the motor state severity level due to the disease progression. The predicted scores follows the trend of the MDS-UPDRS-III level for most of the cases, specially for the SVR in Figure 9.17a). The results suggest that the proposed approach is suitable to monitor the progression of the motor state severity in PD patients using the handwriting signals; however the results have to be validated with more data from additional patients collected in more sessions.

Classification of Patients in Different Levels of the Disease Severity

Although strong correlation were found between the predicted and the real MDS-UPDRS-III scores, it is more informative for the patients to know in which stage of the disease they are. In addition, for medical applications it is difficult to have a great amount of data to train suitable regression algorithms like an SVR or a CNN. For these reasons it should be better to divide the patients into different groups according to their disease severity, based on their MDS-UPDRS-III score. Three

Table 9.27: Results predicting the MDS-UPDRS-III scale of the subjects from the Longitudinal corpus using kinematic features. Results in terms of the Spearman's correlation coefficient.

Task	SVR	GMM-UBM
Alphabet	0.330	0.157
Circle	-0.169	0.080
Cube	0.159	0.181
Free writing	0.497	0.121
House	0.141	0.155
ID	0.449	-0.208
Name	0.562	0.105
Numbers	0.627	0.267
Rectangles	-0.008	0.215
Rey Figure	-0.118	**0.237**
Signature	-0.158	0.009
Spiral	0.009	0.015
Spiral template	0.251	0.028
F.T Early	0.152	
F.T Late 1	0.363	0.115
F.T Late 2	**0.704**	

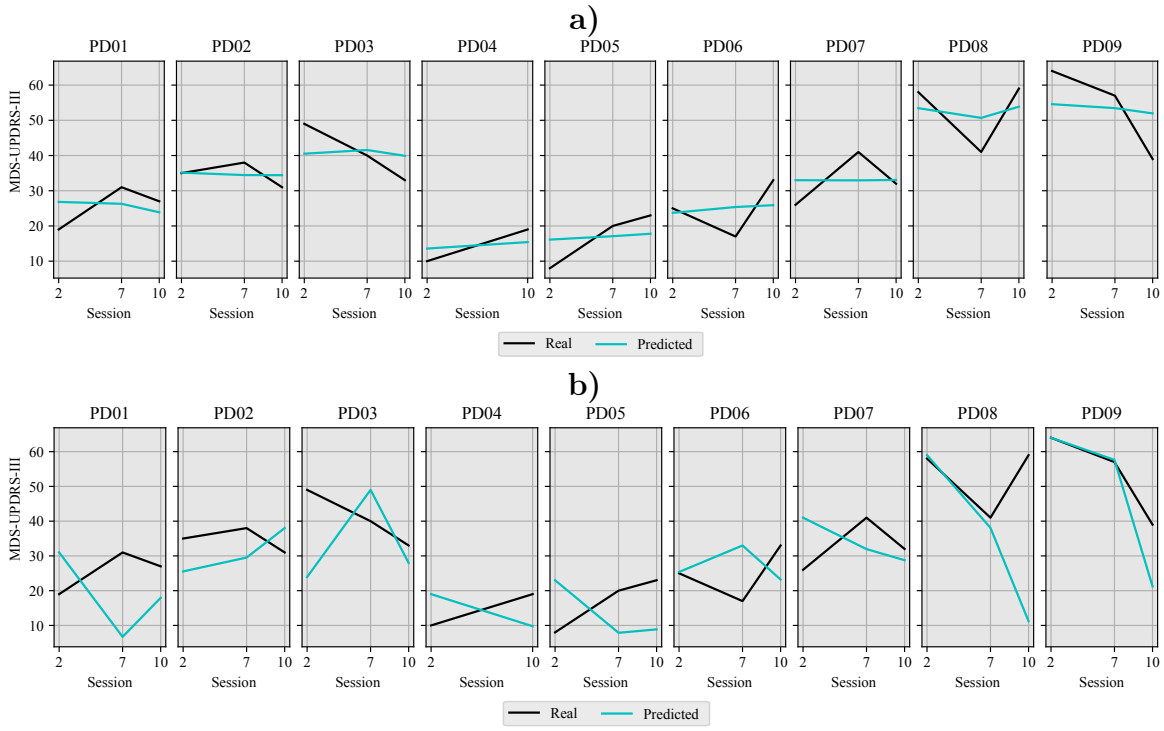

Figure 9.17: Predictions of the MDS-UPDRS-III scale of the patients from the Longitudinal corpus using handwriting signals. **a)** SVR regression. **b)** GMM-UBM.

classes are defined based on the 33rd and 66th percentiles of the total scale in order to discriminate between mild, intermediate, and severe levels of motor-state severities. For the total MDS-UPDRS-III score the ranges per class are defined as follows: 0 to

25 (mild), 26 to 40 (intermediate), and higher than 40 (severe). See Figure 9.12 for more details.

With these new labels, a multiclass SVM is trained in a one vs. all strategy to classify the subjects in the three levels of the disease. The results are presented in Table 9.28. The most accurate models per task are highlighted in bold. The best result is obtained with early fusion at feature-level and the *Late 1* fusion at task-level (UAR=56.5%). Relatively similar results are obtained with the kinematic and in-air feature sets. Regarding the handwriting tasks, there is no particular pattern of which tasks are better for the addressed problem.

Table 9.28: Results classifying subjects from the Multimodal corpus in different motor state levels using handwriting features sets and SVM classifiers. Results in terms of UAR [%].

	Feature set				
Task	**Kinematic**	**Geometric**	**In-air**	**F.F Early**	**F.F Late 1**
Alphabet	33.9 (23.6)		36.7 (16.7)	35.6 (18.3)	**43.0**
Circle	33.3 (0.0)		35.6 (18.3)	**37.8 (13.3)**	33.5
Circle template	38.3 (13.0)		**48.3 (22.9)**	28.3 (22.4)	34.4
Cube	41.7 (14.8)		32.8 (19.0)	40.6 (13.2)	**42.1**
Free writing	35.0 (12.7)		35.0 (11.1)	**42.2 (16.9)**	38.8
House	42.8 (16.5)		30.0 (5.1)	42.8 (16.5)	**44.1**
ID	**41.7 (13.9)**		33.3 (8.6)	34.4 (18.1)	36.6
Name	35.0 (11.4)		36.1 (13.0)	**40.6 (13.4)**	37.9
Numbers	41.7 (15.6)		43.3 (20.6)	**46.7 (13.9)**	40.4
Rectangles	**52.2 (12.0)**		40.0 (17.4)	51.7 (16.5)	52.0
Rey Figure	**45.0 (18.5)**		42.8 (22.2)	42.2 (12.2)	38.9
Signature	27.2 (13.9)		33.3 (19.7)	31.7 (5.0)	**37.1**
Spiral	38.9 (15.1)	37.8 (16.6)	**42.8 (14.9)**	42.2 (16.3)	41.5
Spiral template	49.4 (18.0)	41.7 (13.0)	33.9 (16.6)	**49.4 (18.7)**	48.8
F.T Early	**51.7 (13.8)**		46.7 (22.1)		
F.T Late 1	53.9		41.7	**56.5**	53.9

The multi-class experiments are also performed with the deep learning models with the aim to improve the results. The results are shown in Table 9.29. An UAR up to 56.4% was obtained, which is similar to the reported one combining the kinematic and in-air features for all tasks. The best result is obtained here with the Alphabet task, using the deep learning model to process the pen-up and pen-down transitions.

The summary of the best results obtained using the different methods is shown in Table 9.30. The highest UARs are obtained with the deep learning models and with the fusion of all tasks and feature sets. These results are highlighted in bold. Taking a look to the other performance metrics, it is observed that the deep learning model has higher accuracy and F-score values than the fusion of all feature sets and tasks. In addition, all models are very accurate to detect patients in severe stages of the disease. The most difficult class corresponds to patients in intermediate state.

The confusion matrices from Figure 9.18 show the top-3 results for the classification of patients in three levels of the disease. The three models are very accurate to

Table 9.29: Results classifying subjects from the Multimodal corpus in different motor state levels using the deep learning models. Results in terms of UAR [%].

Task	Online CNN-GRU Transition segments	Online CNN-GRU Full segments
Alphabet	**56.5**	47.5
Circle	36.2	37.5
Circle template	31.9	36.8
Cube	39.4	40.0
Free writing	43.7	44.0
House	43.0	43.1
ID	55.2	48.0
Name	49.7	49.1
Numbers	49.3	52.4
Rectangles	38.6	42.8
Rey Figure	35.4	39.9
Signature	54.3	50.6
Spiral	41.2	35.8
Spiral template	37.1	40.4
F.T Late 1	49.5	44.4

Table 9.30: Best results obtained for each method classifying subjects from the Multimodal corpus in different motor-state levels according to the MDS-UPDRS-III score.

Feature set	Task	ACC	Fscore	UAR	ACC Mild	ACC Intermediate	ACC Severe
Kinematic	F.T Late 1	53.4	0.525	53.9	58.0	33.0	70.0
In-air	Circle template	49.3	0.452	48.3	32.0	50.0	65.0
F.F Early	**F.T Late 1**	**54.2**	**0.537**	**56.5**	**56.0**	**29.0**	**85.0**
F.F Late 1	F.T Late 1	53.4	0.525	53.9	58.0	33.0	70.0
Online CNN-GRU	**Alphabet**	**60.1**	**0.555**	**56.5**	**59.0**	**34.0**	**76.0**

detect patients in severe stages of the disease, followed by patients in initial stage. Patients in intermediate state of the disease are very difficult to classify, for the three methods. These results suggest that would be better and easier to classify patients in two levels of the disease (divided by the median) rather than the classification in the three levels addressed here.

9.3 Gait Assessment

The gait analysis of PD patients covers the discrimination between PD patients and HC subjects and the evaluation of the disease severity of the patients based on the MDS-UPDRS-III scale. The experiments include the evaluation of the gait signals collected with the eGait sensors for the Multimodal corpus and the signals collected with the Apkinson app. The gait assessment is always performed considering both the kinematic, spectral, and NLD features, explained in Chapter 6 as well the deep learning models to process the raw gait signals. The analysis includes the different gait exercises for the Multimodal corpus, and the gait and hand movement tasks for the Apkinson corpus.

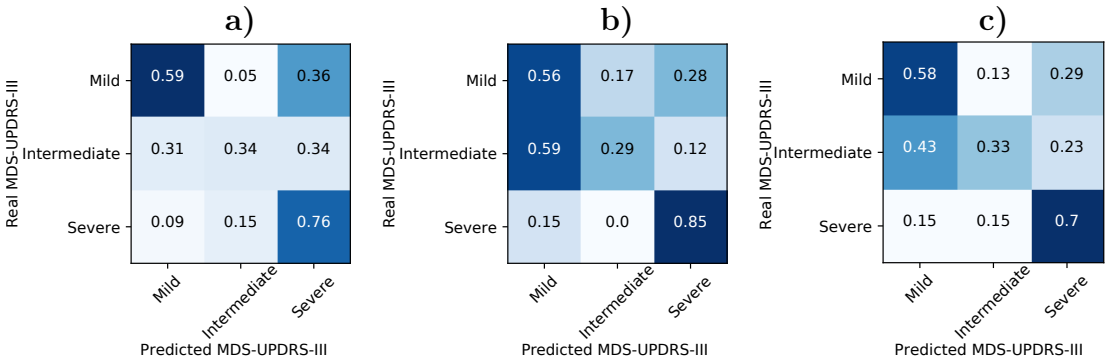

Figure 9.18: Details of the best result obtained classifying subjects from the Multimodal corpus in different motor-state severity levels according to the MDS-UPDRS-III score. **a)** Online CNN-GRU from the alphabet task. **b)** Early fusion of kinematic and in-air features, combining all task with the late fusion strategy. **c)** Late fusion of kinematic and in-air features, combining all task with the late fusion.

9.3.1 Automatic Classification of Parkinson's Disease Patients

The results obtained classifying PD patients vs. HC subjects from the Multimodal corpus (Section 3.3.1) are presented in Table 9.31 using the different feature sets and gait tasks. The feature sets include kinematic, spectral and NLD features, and their combinations. The last column shows the results using the deep learning approach based on CNN-GRU networks. The results are presented in terms of the UAR to avoid bias due to the unbalance in the groups. Results of the best model per task are highlighted in bold.

Table 9.31: Results classifying PD patients vs. HC subjects from the Multimodal corpus using gait signals. Results in terms of UAR [%].

	Feature sets						
Task	**Kinematic**	**Spectral**	**NLD**	**F. F. Early**	**F. F. Late 1**	**F. F. Late 2**	**CNN-GRU**
2x10	64.2 (15.9)	76.8 (9.9)	79.0 (12.4)	79.3 (8.1)	79.9	78.8	**96.6 (2.4)**
4x10	72.4 (9.4)	70.7 (8.7)	68.0 (8.5)	73.7 (4.6)	68.2	73.1	**96.5 (2.4)**
Stop & Go	74.3 (23.5)	74.2 (8.5)	81.5 (14.2)	78.3 (8.0)	76.3	81.8	**98.7 (2.4)**
Heel toe Left		74.9 (9.1)	71.5 (6.3)	75.9 (8.3)	71.7	77.2	**90.6 (2.4)**
Heel toe Right		70.8 (8.6)	**91.9 (6.2)**	83.5 (15.9)	78.3	76.9	90.6 (2.4)
TUG		69.1 (12.3)	71.6 (6.8)	72.9 (10.9)	68.7	68.1	**96.5 (2.4)**
F. T. Early	64.1 (9.0)	78.8 (10.7)	**94.3 (6.5)**				
F. T. Late 1	70.2	81.5	86.2	75.8	85.7		**98.7**
F. T. Late 2	72.6	81.6	86.3	77.1		82.1	**98.3**

The best results are obtained in general with the deep learning approach, which achieved an UAR up to 98.7% when all gait tasks are combined with a late fusion strategy. The NLD features are also very accurate for some of the gait tasks like the heel toe tapping. The results obtained with the kinematic and spectral features are similar, the later one being slightly higher. Kinematic features could not be computed for the heel toe and the TUG tasks because a template to properly segment the strides was not available. The analysis for the individual tasks indicates that for all methods the best result is obtained with the Stop & go task, which is the one when the patients have to perform more start/stop movements of the lower limbs, causing FoG

episodes in the patients. These results confirm the importance of such exercises for the assessment of the gait impairments of PD patients, and it should be carefully considered when designing evaluation protocols.

The results in the Multimodal corpus suggest that the proposed methods are valid and accurate to model the gait of PD patients in a clinical setting. Now the aim is to evaluate whether those methods are also accurate to model the gait of PD patients in at-home environments using the smartphone data, which can be used to monitor the state of the patients at-home. The results classifying PD patients vs. HC subjects from the Apkinson corpus are presented in Table 9.32. Note that for the case of the Apkinson corpus additional hand movement exercises like Finger to nose, Circles, Postural tremor, and Pronation / Supination are included in the analysis. The best results per task are also highlighted in bold. The kinematic features were not included in this analysis because the template to properly segment the individual strides was not available.

Table 9.32: Results classifying PD patients vs. HC subjects from the Apkinson corpus using gait signals. Results in terms of UAR [%].

	Feature sets					
Task	**Spectral**	**Non-linear**	**F. F. Early**	**F. F. Late 1**	**F. F. Late 2**	**CNN-GRU**
4x10	51.3 (15.5)	65.0 (14.5)	60.8 (15.3)	60.5	61.2	**83.0 (12.2)**
Free Gait	64.2 (10.7)	71.3 (20.5)	70.4 (17.1)	75.5	68.7	**84.4 (12.2)**
Circles Left	50.4 (18.4)	69.2 (7.0)	64.6 (11.7)	64.1	71.1	**72.9 (12.2)**
Circles Right	59.2 (18.0)	65.0 (9.7)	69.2 (12.4)	63.3	62.2	**78.2 (12.2)**
Finger to nose Left	59.7 (17.9)	70.7 (16.4)	67.3 (9.8)	60.7	71.2	**76.1 (12.2)**
Finger to nose Right	62.3 (12.8)	**67.7 (13.9)**	64.8 (13.4)	70.3	67.5	66.7 (12.2)
Postural Tremor Left	50.0 (0.0)	57.9 (11.9)	**70.0 (12.6)**	70.0	59.9	55.3 (12.2)
Postural Tremor Right	55.8 (17.5)	**78.3 (17.2)**	72.5 (21.4)	67.9	74.8	53.7 (12.2)
Posture	55.0 (20.4)	55.0 (16.9)	**62.1 (16.1)**	58.5	56.4	59.8 (12.2)
Pronation / Supination Left	51.8 (23.6)	**71.2 (15.8)**	67.2 (12.9)	58.7	71.1	56.3 (12.2)
Pronation / Supination Right	58.8 (8.9)	**70.6 (19.1)**	59.3 (10.8)	68.3	63.7	54.7.0 (12.2)
F. T. Early	67.0 (12.5)	**79.5(15.7)**				
F. T. Late 1	66.3	78.3	81.7	82.1		**83.3**
F. T. Late 2	62.5	78.8	80.4		75.0	**83.8**

The comparison between the results observed for the Multimodal and Apkinson corpus, using each method indicates that the accuracy is reduced in about 14.6% for the spectral features, 14.8% for the NLD features, and 14.9% for the CNN-GRU model. This result is explained because in the smartphone data only one 3-axial accelerometer is available, compared to the eGait system where both a 3-axial accelerometer and gyroscope is attached to each foot. The highest UARs are obtained with the CNN-GRU model (84.4%). The deep learning model is specially very accurate to model the gait tasks like 4x10 and the Free gait. Hand movement tasks like Finger to nose and Circles produce moderate accuracies. Conversely, tasks such as Postural tremor, Posture, or Pronation / Supination are not accurate for the classification using the proposed CNN-GRU model. This is probably because these tasks do not have high temporal variability, thus the information provided by these exercises is not properly exploited by the neural network. These particular hand movement exercises are better modeled with the NLD features.

A summary of the best results obtained using the different methods is shown in Table 9.33. The results include additional performance metrics like sensitivity, specificity, F-score, and the AUC. The best result obtained for the Multimodal and

Apkinson corpus are highlighted in bold, and they were obtained with the deep CNN-GRU models. In addition, for almost all cases, the best results are obtained combining the different tasks with the early or late fusion strategies. After the results using the CNN-GRU models, the best results for the Multimodal corpus are obtained with the early fusion at task-level using the NLD features(UAR=94.3%) and for the Apkinson corpus using the *Late 1* fusion both at feature- and task-levels (UAR=82.1%).

Table 9.33: Best results obtained for each method classifying PD patients and HC subjects in the Multimodal and Apkinson corpus using gait signals.

Feature set	Task	ACC	Fscore	UAR	SENS	SPEC	AUC
Multimodal corpus							
Kinematic	Stop & Go	77.1	0.724	74.3	83.5	65.0	0.757
Spectral	F. T. Late 2	82.0	0.799	81.6	82.8	80.4	0.890
NLD	F. T. Early	90.3	0.874	94.3	88.6	100.0	0.954
F. F. Early	Heel toe Right	78.9	0.720	83.5	76.9	90.0	0.905
F. F. Late 1	2x10	82.0	0.793	79.9	85.3	74.5	0.853
F. F. Late 2	F. T. Late 2	82.0	0.800	82.1	81.9	82.4	0.898
CNN-GRU	**F. T. Late 1**	**98.2**	**0.979**	**98.7**	**97.4**	**100.0**	**0.999**
Apkinson corpus							
Spectral	F. T. Early	77.5	0.670	67.0	40.0	94.0	0.665
NLD	F. T. Early	83.6	0.796	79.5	68.3	90.7	0.848
F. F. Early	F. T. Late 1	85.0	0.832	81.7	65.0	98.3	0.898
F. F. Late 1	F. T. Late 1	82.0	0.815	82.1	82.5	81.7	0.924
F. F. Late 2	F. T. Late 2	80.0	0.762	75.0	50.0	100.0	0.924
CNN-GRU	**Free Gait**	**85.6**	**0.845**	**84.4**	**79.4**	**89.3**	**0.938**

The confusion matrices, ROC curves, and the histograms obtained with the scores of the predictions of the CNN-GRU models are shown in Figures 9.19 and 9.20 for the Multimodal and Apkinson corpus, respectively. For the case of the Multimodal corpus there are no HCs that are misclassified and there are three misclassified patients. Note that at least two of them are very close to the decision boundary of the classifier, which is set at 0 (see the histogram in Figure 9.19c). For the case of the Apkinson corpus there are six misclassified HC subjects and seven PD patients. Note in the histogram in Figure 9.20c) that most of the misclassified subjects in the Apkinson corpus are just in the decision boundary and not in the other extreme class.

The results are compared to the ones obtained with the classical feature extraction and classification using the SVM classifier. The confusion matrices, ROC curves, and the respective histograms obtained using the traditional techniques are shown in Figures 9.21 and 9.22 for the Multimodal and Apkinson corpus, respectively. The results for the Multimodal corpus correspond to the ones obtained with the NLD features and the early fusion at task-level. The results for the Apkinson corpus correspond to the *Late 1* fusion at feature- and task-levels.

9.3.2 Automatic Evaluation of the Motor State Severity of Patients

Similar to the handwriting analysis, the evaluation of the motor state severity of patients is based on the MDS-UPDRS-III scale. The evaluation is performed in three

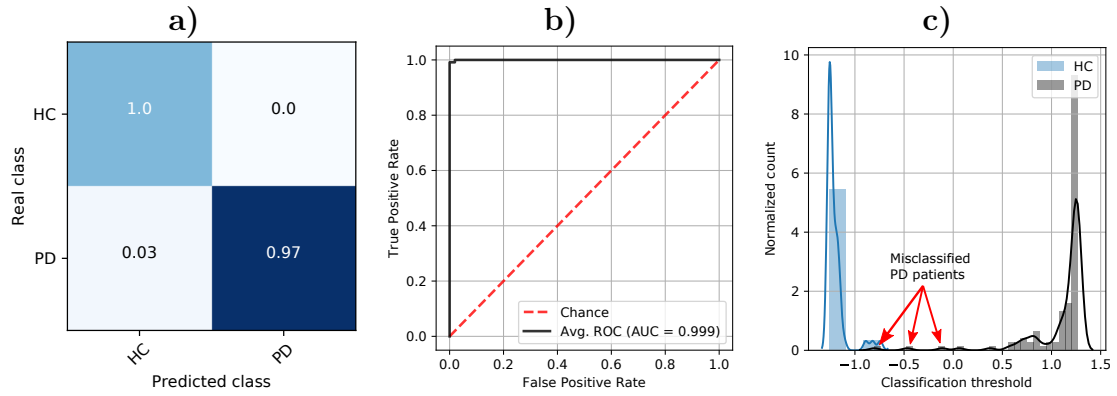

Figure 9.19: Details of the best result obtained classifying PD patients and HC subjects from the Multimodal corpus using the CNN-GRU model. **a)** Normalized confusion matrix. **b)** ROC curve. **c)** Distribution of the classification scores.

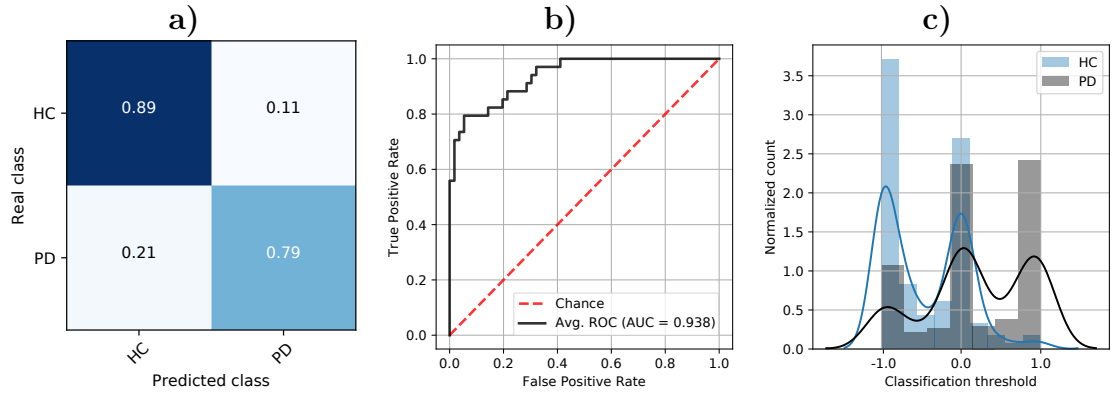

Figure 9.20: Details of the best result obtained classifying PD patients and HC subjects from the Apkinson corpus using the CNN-GRU model. **a)** Normalized confusion matrix. **b)** ROC curve. **c)** Distribution of the classification scores.

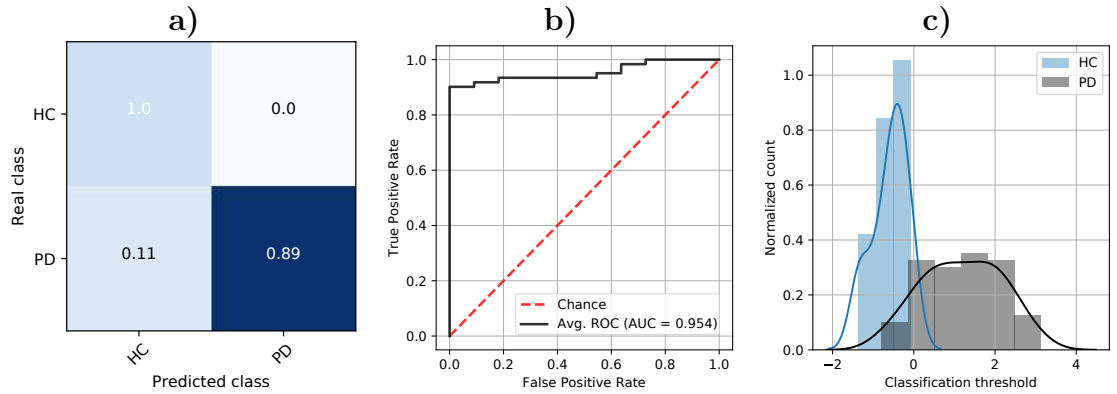

Figure 9.21: Details of the best result obtained classifying PD patients and HC subjects from the Multimodal corpus using NLD features and SVM classifiers. **a)** Normalized confusion matrix. **b)** ROC curve. **c)** Distribution of the classification scores.

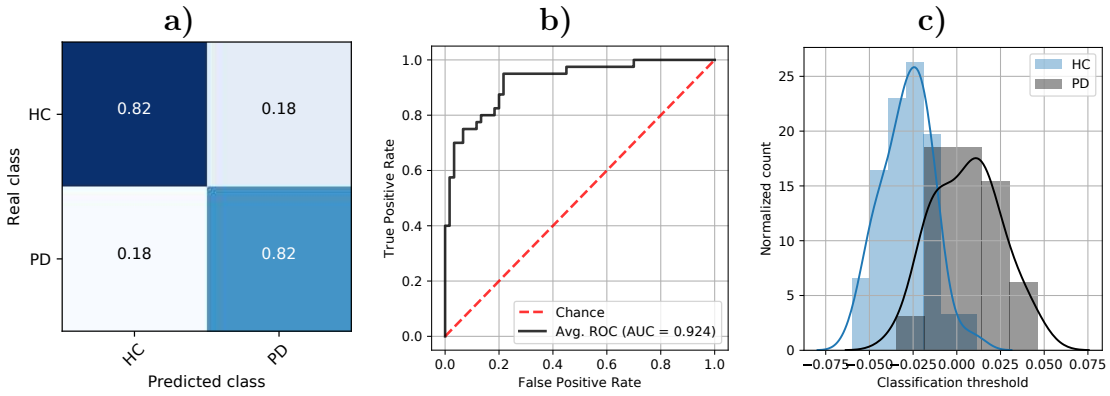

Figure 9.22: Details of the best result obtained classifying PD patients and HC subjects from the Apkinson corpus using different feature sets and SVM classifiers. **a)** Normalized confusion matrix. **b)** ROC curve. **c)** Distribution of the classification scores.

scenarios with the gait signals: (1) the prediction of the value of the MDS-UPDRS-III scale using the gait models described in Chapter 6 and regression algorithms; (2) the progression evaluation of the patients from the Longitudinal corpus using both user models based on GMM-UBM systems and regression algorithms; and (3) the classification of patients in different levels of the disease severity (mild, intermediate, severe) using multi-class classification methods.

Motor State Evaluation based on Regression Algorithms

The aim of this experiment is to evaluate the motor-state severity of the patients by estimating the value of the MDS-UPDRS-III using the different gait features and regression algorithms. The SVR regression is trained with the kinematic, spectral, and NLD features. The early and late fusion strategies are also considered. The CNN-GRU model is also trained for this particular task. The results obtained for the Multimodal corpus are observed in Table 9.34. Unfortunately, so far there is not enough labeled data to train the regression algorithms for the Apkinson corpus. The most correlated result is observed when all gait tasks and features are combined together with the *Late 2* fusion method (ρ=0.646). For many of the gait tasks the best result is obtained when only the NLD features were considered, which validates the importance of these features to model the walking process of PD patients. There is also a high correlation obtained with the spectral features and the 4x10 task (ρ=0.620), which is very close to the best result achieved. The results using the deep learning models are not accurate to evaluate the motor-state severity of the patients, similar to the results obtained with the handwriting signals. The main reason is because the lack of labeled data to train the regression network.

The summary of the best results obtained with each method is shown in Table 9.35. The best result was obtained with the fusion of all gait tasks and feature sets, as it was mentioned previously. For all methods, excluding the CNN-GRU models the correlation between the estimated and real MDS-UPDRS-III scores is statistically significant. Figure 9.23 shows the errors in the evaluation of the MDS-

Table 9.34: Results estimating the MDS-UPDRS-III scale of the subjects from the Multimodal corpus using gait signals. Results in terms of the Spearman's correlation coefficient.

	Feature sets						
Task	**Kinematic**	**Spectral**	**NLD**	**F. F. Early**	**F. F. Late 1**	**F. F. Late 2**	**CNN-GRU**
2x10	0.435	0.526	0.481	0.490	0.448	**0.542**	0.036
4x10	0.312	**0.620**	0.532	0.626	0.608	0.617	0.095
Stop & Go	0.261	0.315	0.477	0.474	0.182	**0.467**	-0.012
Heel toe Left		-0.218	**0.196**	0.130	-0.026	0.013	0.009
Heel toe Right		-0.208	0.028	**0.109**	0.044	-0.085	0.009
TUG		0.340	**0.424**	0.256	0.413	0.378	0.090
F. T. Early	0.372	0.506	**0.635**				
F. T. Late 1	0.317	0.391	**0.504**	0.452	0.459		0.045
F. T. Late 2	0.447	0.607	0.572	0.550		**0.646**	0.098

UPDRS-III scores for the best model. Although the result is satisfactory (strong correlation), other regression strategies and feature sets can be considered to improve the correlation values.

Table 9.35: Best results obtained for each method to evaluate the MDS-UPDRS-III scale of the subjects in the Multimodal corpus using gait signals.

Feature set	**Task**	r	**p-val** r	ρ	**p-val** ρ	**MAE**
Kinematic	F. T. Late 2	0.439	$\ll$0.005	0.447	$\ll$0.005	9.2
Spectral	4x10	0.567	$\ll$0.005	0.620	$\ll$0.005	6.9
NLD	F. T. Early	0.560	$\ll$0.005	0.635	$\ll$0.005	8.3
F. F. Early	4x10	0.580	$\ll$0.005	0.626	$\ll$0.005	7.5
F. F. Late 1	4x10	0.580	$\ll$0.005	0.608	$\ll$0.005	9.2
F. F. Late 2	**F. T. Late 2**	**0.554**	**$\ll$0.005**	**0.646**	**$\ll$0.005**	**9.1**
CNN-GRU	F. T. Late 2	0.148	0.187	0.098	0.382	9.7

Longitudinal Assessment of Patients

The SVR and the GMM-UBM models are considered as well to model the disease progression of the patients from the Longitudinal corpus based on the MDS-UPDRS-III scale, similar to the experiments addressed with speech and handwriting signals. The SVR is trained with the data from the Multimodal corpus, excluding those subjects from the Longitudinal corpus. Then, patients from the Longitudinal data form an independent hidden test set for the regression method. For the GMM-UBM system the UBMs are trained using information from the HC subjects from the Multimodal corpus. Then specific GMMs are adapted for each patient in each session from the Longitudinal corpus, similar to the experiments addressed in [Vasq 20b]. The prediction is performed using both spectral and NLD features. The results obtained using both methods are shown in Table 9.36. The best results are obtained with the SVR approach, similar to the observed for the cases of handwriting and speech signals. The main reason again is because here there are more data to train the regression models, thus obtaining more accurate results using a supervised model like the SVR instead of the unsupervised model based on GMM-UBM. The strongest

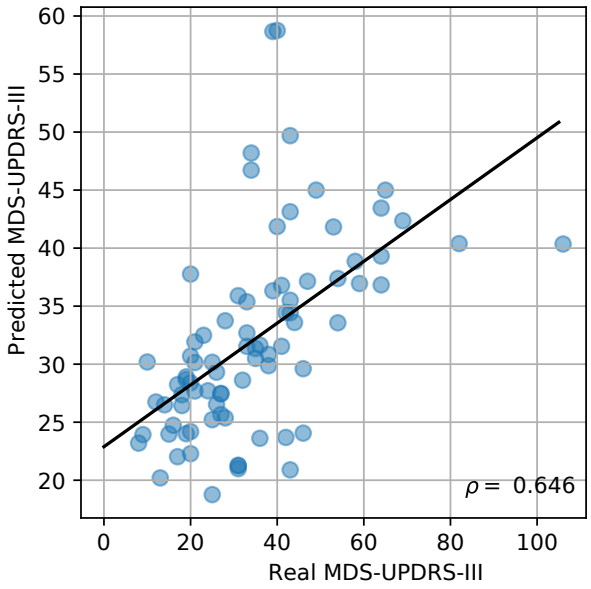

Figure 9.23: Details of the best result obtained estimating the MDS-UPDRS-III scale of the subjects in the Multimodal corpus using all gait features features and SVR regressors.

correlation (ρ =0.770) is obtained using the 4x10 task and the fusion of spectral and NLD features. Strong correlations are also obtained considering separately the spectral features in the 2x10 (ρ =0.697), 4x10 (ρ =0.710), and Stop & Go (ρ =0.644) tasks. Similar results are obtained with the NLD features in the 2x10 (ρ =0.652), 4x10 (ρ =0.628), and TUG (ρ =0.603) tasks. This is very positive considering the fact that these results are obtained as an independent test set that was never seen during a cross-validation strategy.

Table 9.36: Results predicting the MDS-UPDRS-III scale of the subjects from the Longitudinal corpus using different gait features. Results in terms of the Spearman's correlation coefficient.

	SVR					GMM-UBM		
Task	**Spectral**	**NLD**	**F. F. Early**	**F. F. Late 1**	**F. F. Late 2**	**Spectral**	**NLD**	**F. F**
2x10	**0.697**	0.652	0.702	0.206	0.666	0.322	0.097	0.120
4x10	0.710	0.628	0.734	-0.253	**0.770**	0.051	0.080	-0.374
Stop & Go	0.644	0.481	**0.648**	0.617	0.646	0.423	0.241	0.038
Heel toe Left	0.283	0.320	0.214	**0.332**	0.314	-0.190	0.036	0.038
Heel toe Right	0.321	0.387	**0.461**	0.387	0.248	-0.190	0.036	-0.004
TUG	0.266	**0.667**	0.636	0.457	0.503	0.564	0.268	0.184
F. T. Early	0.395	**0.603**						
F. T. Late 1	-0.016	**0.704**						
F. T. Late 2	0.733	**0.760**						

Figure 9.24 shows the prediction of the MDS-UPDRS-III scores for the nine patients of the Longitudinal corpus for the best model i.e., the late fusion of spectral and NLD features for the 4x10 task. The lines for each speaker represent the progression of the motor state severity level. The predicted scores follows the trend of the MDS-UPDRS-III level in many cases. Note specially that for patients PD04, PD05, and

PD07 the prediction is very similar to the real score. The results suggests that the proposed approach is suitable to monitor the progression of the motor state severity in PD patients using the gait signals; however the results have to be validated with more data from additional patients collected in more sessions.

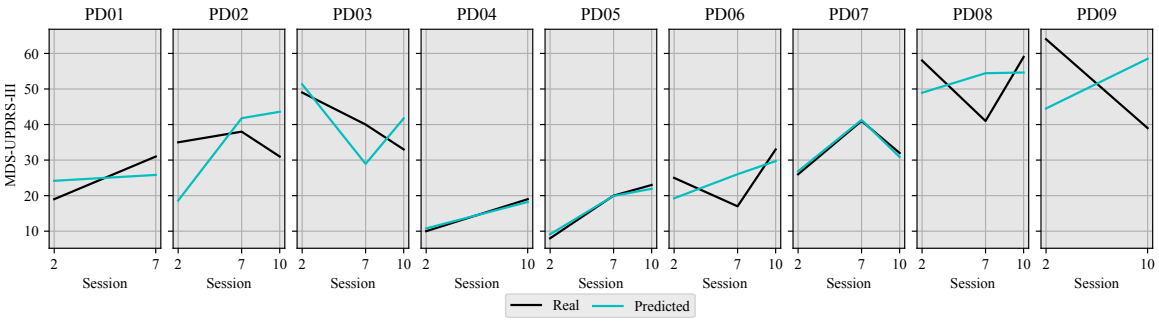

Figure 9.24: Predictions of the MDS-UPDRS-III score of each patient in the Longitudinal corpus with the gait features and the SVR regression.

Classification of Patients in Different Levels of the Disease Severity

The classification of patients in different levels of the disease severity is also performed with gait signals in the same way as the addressed for the case of speech and handwriting data. Three classes are defined based on the 33rd and 66th percentiles of the total MDS-UPDRS-III scale in order to discriminate between mild, intermediate, and severe levels of motor-state severities. The ranges of the total MDS-UPDRS-III score are defined as follows: 0 to 25 (mild), 26 to 40 (intermediate), and higher than 40 (severe). The distribution and limits of the scores were chosen in order to have three equal priors for the classes. The distribution was shown in Figure 9.12. With these new labels, a multiclass SVM is trained in a one vs. all strategy to classify the subjects in the three levels of the disease. The CNN-GRU models were also trained for such a purpose. The results obtained for the patients from the Multimodal corpus are observed in Table 9.37. The best result is obtained with early fusion at task-level and the NLD features (UAR=70.6%). This result is 11.6% higher than the best one observed with handwriting signals, which was up to 59.0% (see Table 9.30). There is no particular trend about which is the best model for the classification, however the CNN-GRU achieved the highest accuracy in four of the six tasks, specially for the TUG exercise (64.9%). In addition, the 4x10 and the TUG exhibit on average higher accuracies than the ones observed in other tasks like heel toe tapping and 2x10. Unfortunately, the late fusion strategy did not produce accurate results for this problem.

The multi-class classification was also performed with the movement signals collected using the Apkinson app. The results are observed in Table 9.38. For this case there are differences of up to 10% with respect to the results observed for the Multimodal corpus. The best result is observed here with the spectral features computed over the posture exercise (60.0%). Additional data should be collected and labeled in order to improve the results using the different proposed methods.

Table 9.37: Results classifying subjects from the Multimodal corpus in motor state levels using gait signals. Results in terms of UAR [%].

	Feature sets					
Task	**Kinematic**	**Spectral**	**NLD**	**F.F. Early**	**F.F. Late 1**	**CNN-GRU**
2x10	35.0 (5.0)	53.9 (8.3)	57.8 (18.5)	51.1 (10.5)	55.1	**60.8 (9.6)**
4x10	33.3 (0.0)	**59.4 (18.1)**	57.8 (7.9)	57.8 (13.9)	60.0	58.1 (9.6)
Stop & Go	43.3 (24.9)	46.1 (16.7)	47.8 (14.3)	45.6 (15.1)	50.4	**62.9 (9.6)**
Heel toe Left		38.3 (13.5)	42.8 (19.9)	38.3 (21.1)	42.4	**46.4 (9.6)**
Heel toe Right		**47.2 (18.6)**	44.4 (18.3)	39.4 (18.3)	44.5	46.4 (9.6)
TUG		55.0 (15.8)	55.0 (17.5)	51.7 (11.7)	54.4	**64.9 (9.6)**
F. T. Early	52.2 (9.0)	58.3 (16.5)	**70.6 (13.1)**			
F. T. Late 1	41.0	51.5	56.1	57.9	**59.0**	54.9 (9.6)

Table 9.38: Results classifying subjects from the Apkinson corpus in motor state levels using gait signals. Results in terms of UAR [%].

	Feature sets				
Task	**Spectral**	**NLD**	**F. F. Early**	**F. F. Late 1**	**CNN-GRU**
4x10	33.3 (21.1)	43.3 (24.9)	43.3 (13.3)	25.6	**44.4 (10.9)**
Free Gait	33.3 (10.5)	30.0 (22.1)	**43.3 (20.0)**	28.1	37.2 (10.9)
Circles Left	43.3 (22.6)	43.3 (17.0)	50.0 (18.3)	**52.4**	37.5 (10.9)
Circles Right	33.3 (0.0)	30.0 (6.7)	33.3 (0.0)	**48.0**	20.0 (10.9)
Finger to nose Left	33.3 (0.0)	30.0 (6.7)	33.3 (0.0)	**35.3**	21.0 (10.9)
Finger to nose Right	33.3 (10.5)	26.7 (13.3)	36.7 (6.7)	23.4	**37.5 (10.9)**
Postural Tremor Left	36.7 (19.4)	**50.0 (10.5)**	30.0 (6.7)	13.9	35.6 (10.9)
Postural Tremor Right	26.7 (13.3)	33.3 (0.0)	33.3 (0.0)	30.6	**59.0 (10.9)**
Posture	**60.0 (22.6)**	50.0 (10.5)	46.7 (16.3)	35.3	33.3 (10.9)
Pronation / Supination Left	53.3 (16.3)	**56.7 (13.3)**	46.7 (16.3)	53.2	19.0 (10.9)
Pronation / Supination Right	30.0 (12.5)	40.0 (13.3)	30.0 (19.4)	**49.6**	33.3 (10.9)
F. T. Early	37.9 (13.5)	**39.4 (12.6)**			
F. T. Late 1	29.8	27.8	**38.8**	35.9	34.4

The summary of the best results obtained using the different methods is shown in Table 9.39 both for the Multimodal and Apkinson corpus. The results include additional performance metrics like the weighted accuracy, the F-score, and the accuracies obtained for each of the three classes. The model based on NLD features in the Multimodal corpus is the method that provides the best accuracy to classify patients in intermediate stage of the disease, which is the most misclassified class for the other methods. It also shows to be the most accurate method to classify patients in severe stages of the disease (89%).

The confusion matrices from Figure 9.25 show the top-3 results for the classification of patients in the three disease levels for the Multimodal corpus. The results observed for the case of NLD features in Figure 9.11a) show to be the most accurate because this particular model is the most balanced to classify the three classes. Most of the classification errors with the NLD features correspond to misclassifications between patients in mild and intermediate stages of the disease. Patients in severe stages are also very accurately classified. These results indicate that there is a clear separation between the patients in severe stages with respect to the other two classes, suggesting additional experiments classifying not three but only two levels of the disease severity.

Table 9.39: Best results obtained for each method classifying subjects from the Multimodal and Apkinson corpus in different motor-state levels according to the MDS-UPDRS-III score.

Features	Task	ACC	F-score	UAR	ACC MILD	ACC INTERMEDIATE	ACC SEVERE
Multimodal corpus							
Kinematic	F. T. Early	52.0	0.483	52.2	45.0	39.0	70.0
Spectral	4x10	60.9	0.579	59.4	71.0	40.0	68.0
NLD	**F. T. Early**	**69.9**	**0.647**	**70.6**	**61.0**	**67.0**	**89.0**
F. F. Early	4x10	59.5	0.556	57.8	75.0	32.0	68.0
F. F. Late 1	4x10	60.5	0.586	60.0	80.0	33.0	67.0
CNN-GRU	TUG	62.9	0.632	64.9	66.0	53.0	78.0
Apkinson corpus							
Spectral	**Posture**	**56.7**	**0.556**	**60.0**	**56.0**	**60.0**	**56.0**
NLD	Pron. / Sup. Left	54.7	0.438	56.7	50.0	60.0	50.0
F. F. Early	Circles Left	49.3	0.436	50.0	33.0	30.0	89.0
F. F. Late 1	Pron. / Sup. Left	52.9	0.520	53.2	67.0	50.0	43.0
CNN-GRU	Postural Tremor Right	46.2	0.411	59.0	57.0	20.0	100.0

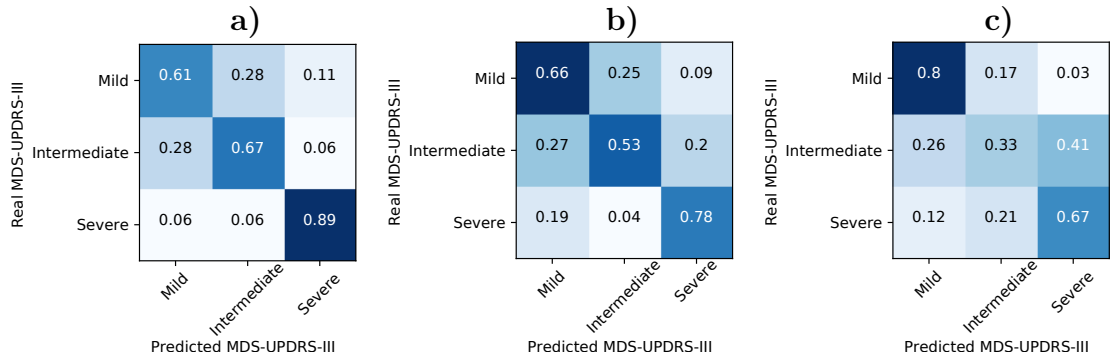

Figure 9.25: Details of the best result obtained classifying subjects from the Multimodal corpus in different motor-state severity levels according to the MDS-UPDRS-III score. **a)** NLD features combining all tasks. **b)** CNN-GRU model from the TUG task. **c)** Late fusion of kinematic, spectral, and NLD features for the 4x10 task.

The confusion matrices from Figure 9.26 show as well the top-3 results for the Apkinson corpus. The results for the spectral features in Figure 9.26a) show to be the most balanced among the three classes. The CNN-GRU model in Figure 9.26b) is very accurate to classify patients in severe stages of the disease, despite the misclassifications of patients in mild stage classified as severe. A more problematic error would be the contrary i.e., patients in severe stages classified as mild because they wont be prescribed with the proper treatment according to their disease severity.

9.4 Asynchronous Multimodal Assessment

This section presents the results combining speech, handwriting, and gait signals both to classify PD vs. HC subjects and to evaluate the motor-state severity of the patients. Different strategies based on early and late fusion approaches were

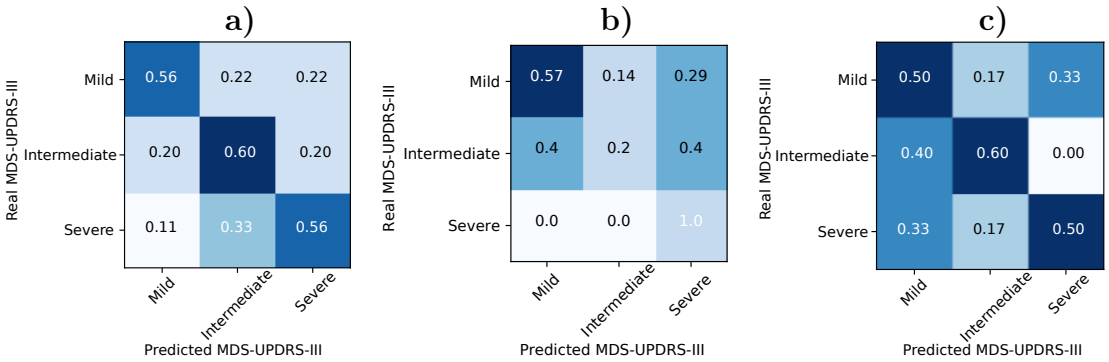

Figure 9.26: Details of the best result obtained classifying subjects from the Apkinson corpus in different motor-state severity levels according to the MDS-UPDRS-III score. **a)** Spectral features computed upon the posture exercise. **b)** CNN-GRU model from the Postural tremor task. **c)** NLD features from the Pronation/Supination task.

considered for each case. The results of the fusion to classify PD patients and HC subjects from the Multimodal corpus are shown in Table 9.40. The results include experiments performed using the extracted features for each modality and the later SVM classifiers, and the models based on deep learning methods. The fusion is always performed considering both the top-5 methods that produced the best results with each modality, and using all methods applied to each bio-signal. For the case of speech signals the models considered in the top-5 fusion includes the phonological features computed over the DDK tasks, the monologue, the read text, and the 6th and 7th sentences. The top-5 methods for handwriting signals include the in-air features computed over the most complex handwriting exercises such as the alphabet, the cube, the free writing, the house and the Rey-Osterrieth figure. The top-5 models for gait signals include NLD features computed over the 2x10, Stop & Go, and heel toe tapping tasks, and spectral features computed over the 2x10 and heel toe tapping tasks. For the case of the *all* fusion, only the late fusion approaches were considered. The top-5 fusion for the deep learning models include those tasks that achieve the most accurate results for each modality, and include the DDKs and the sentences 2, 4, 8, and 9 for speech signals, the alphabet, cube, ID, name, and the Rey-Osterrieth figure for handwriting, and the five tasks from the gait signals.

The best result using the feature extraction and the SVM classifiers is obtained for the early fusion of the top-5 features from the handwriting and gait signals (UAR=99.2%). This results improves in up to 12.2% the ones obtained using only speech signals (see Table 9.1), in up to 3.3% the ones obtained with handwriting signals (see Table 9.21), and in up to 4.9% the ones obtained using only gait signals (see Table 9.31). No high differences are observed when using the early or late fusion strategies, neither using the top-5 or the *all* fusions. However, the results using the top-5 models are slightly more accurate. The fusion using the deep learning techniques show that the best results is also obtained when the handwriting and gait signals are combined (UAR=99.4%) using the late fusion of the top-5 model for each bio-signal. The are not big differences observed when using the *Late 1* or *Late 2* fusion approaches.

Table 9.40: Classification of PD patients vs. HC subjects from the Multimodal corpus combining the different speech, handwriting, and gait models.

Fusion	Modalities	# Feature sets / Classifiers	ACC	Fscore	UAR	SENS	SPEC	AUC
Feature extraction and SVMs								
Early fusion Top-5	Speech+Handwriting	10	88.8	0.856	89.6	88.3	90.8	0.953
Late fusion 1 Top 5	Speech+Handwriting	10	88.9	0.876	89.5	87.9	91.0	0.953
Late fusion 2 Top 5	Speech+Handwriting	10	88.3	0.869	88.8	87.5	90.0	0.953
Late fusion 1 all	Speech+Handwriting	120	86.7	0.849	86.0	87.9	84.0	0.954
Late fusion 2 all	Speech+Handwriting	120	87.3	0.862	89.2	84.4	94.0	0.959
Early fusion Top-5	Speech+Gait	10	95.7	0.936	97.5	95.0	100.0	0.991
Late fusion 1 Top 5	Speech+Gait	10	88.5	0.872	89.2	87.5	90.9	0.947
Late fusion 2 Top 5	Speech+Gait	10	89.5	0.881	89.6	89.3	89.9	0.949
Late fusion 1 all	Speech+Gait	108	85.4	0.833	84.2	87.5	80.8	0.943
Late fusion 2 all	Speech+Gait	108	87.9	0.868	89.9	84.8	94.9	0.956
Early fusion Top-5	**Handwriting+Gait**	**10**	**98.8**	**0.985**	**99.2**	**98.3**	**100.0**	**0.998**
Late fusion 1 Top 5	Handwriting+Gait	10	91.4	0.902	92.2	90.2	94.2	0.983
Late fusion 2 Top 5	Handwriting+Gait	10	91.4	0.902	92.2	90.2	94.2	0.976
Late fusion 1 all	Handwriting+Gait	48	88.0	0.846	82.6	95.9	69.2	0.948
Late fusion 2 all	Handwriting+Gait	48	87.4	0.857	87.7	87.0	88.5	0.956
Early fusion Top-5	Speech+Handwriting+Gait	15	95.4	0.941	97.2	94.3	100.0	0.995
Late fusion 1 Top 5	Speech+Handwriting+Gait	15	90.4	0.893	91.4	88.8	94.0	0.956
Late fusion 2 Top 5	Speech+Handwriting+Gait	15	89.5	0.883	90.5	87.9	93.0	0.956
Late fusion 1 all	Speech+Handwriting+Gait	138	87.0	0.855	87.6	86.2	89.0	0.951
Late fusion 2 all	Speech+Handwriting+Gait	138	89.2	0.881	91.1	86.2	96.0	0.961
Deep learning models								
Late fusion 1 Top 5	Speech+Handwriting	10	95.7	0.951	96.9	93.8	100.0	0.981
Late fusion 2 Top 5	Speech+Handwriting	10	93.5	0.928	95.3	90.6	100.0	0.996
Late fusion 1 all	Speech+Handwriting	33	96.9	0.965	97.8	95.5	100.0	0.976
Late fusion 2 all	Speech+Handwriting	33	96.0	0.955	97.1	94.2	100.0	0.997
Late fusion 1 Top 5	Speech+Gait	10	93.4	0.924	95.3	90.7	100.0	0.998
Late fusion 2 Top 5	Speech+Gait	10	93.7	0.927	95.6	91.2	100.0	0.998
Late fusion 1 all	Speech+Gait	24	97.5	0.971	98.2	96.4	100.0	0.998
Late fusion 2 all	Speech+Gait	24	96.3	0.958	97.3	94.6	100.0	0.999
Late fusion 1 Top 5	**Handwriting+Gait**	**10**	**99.4**	**0.993**	**99.6**	**99.2**	**100.0**	**0.998**
Late fusion 2 Top 5	Handwriting+Gait	10	98.8	0.986	99.2	98.4	100.0	0.998
Late fusion 1 all	Handwriting+Gait	21	99.4	0.993	99.6	99.2	100.0	0.998
Late fusion 2 all	Handwriting+Gait	21	98.9	0.986	99.2	98.4	100.0	0.998
Late fusion 1 Top 5	Speech+Handwriting+Gait	15	96.6	0.961	97.5	95.1	100.0	0.981
Late fusion 2 Top 5	Speech+Handwriting+Gait	15	95.4	0.948	96.7	93.3	100.0	0.996
Late fusion 1 all	Speech+Handwriting+Gait	39	96.6	0.961	97.5	95.1	100.0	0.995
Late fusion 2 all	Speech+Handwriting+Gait	39	96.3	0.958	97.3	94.6	100.0	0.998

The scatter plots observed in Figure 9.27 show the distribution of the classification scores for the classifiers when the models for speech, handwriting and gait are combined. The decision function of the multimodal system is also depicted in the figures, and computed based on Equation 9.1 for the *Late 1* fusion, being M the number of modalities, y_i the scores for the i-th modality, and $\widehat{\alpha}_i$ the weight associated to each classifier (see Equation 7.2). There are overlaps of the samples when the traditional features and SVM classifiers are considered (Figures 9.27a) to d)). The results using the deep learning models show that the scores for the HC subjects are mainly concentrated on the left bottom part of the figures, indicating the confidence of the classifiers to discriminate this group. The scores for the PD patients are distributed over the complete space, showing in some cases that they are misclassified in one of the modalities, but when the two or three bio-signals are considered together, the classification is performed correctly.

$$Y = \sum_{i=1}^{M} y_i \widehat{\alpha}_i \tag{9.1}$$

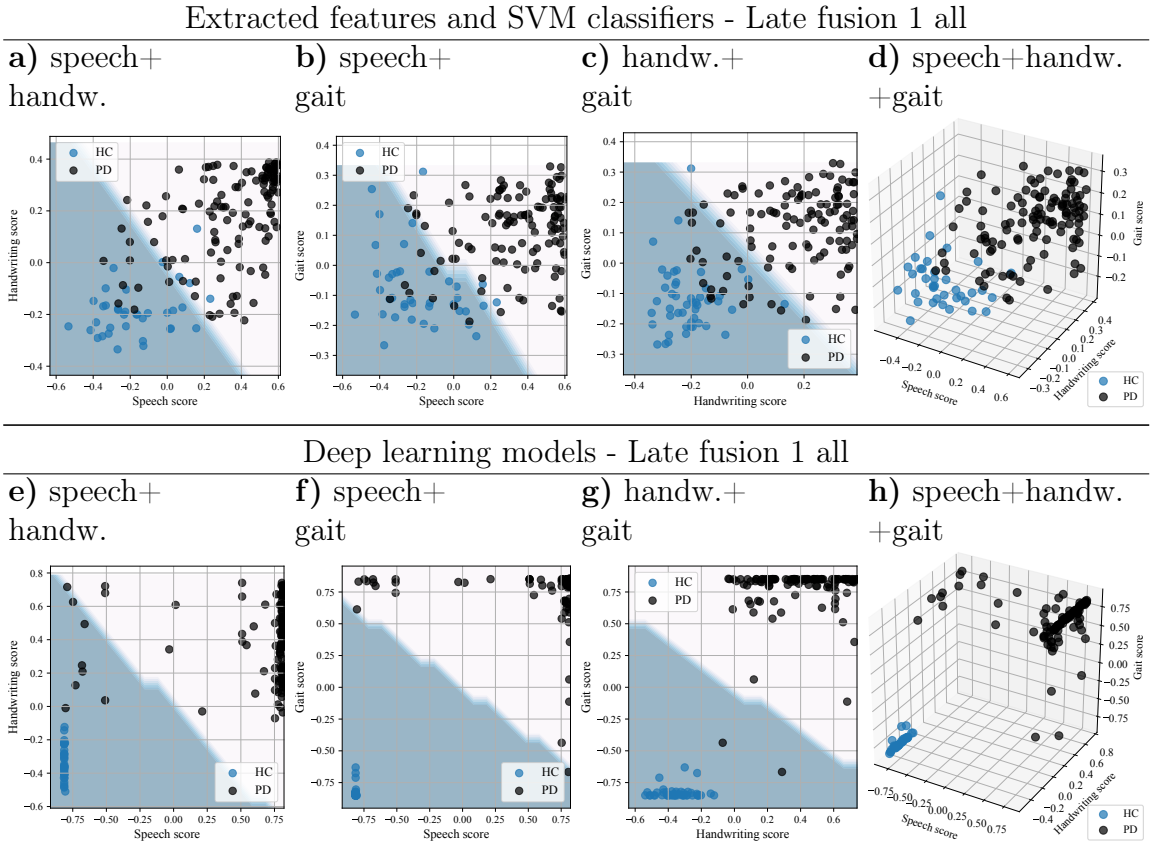

Figure 9.27: Distribution of the classification scores from the subjects of the Multimodal corpus for each modality.

The multimodal analysis is also performed in the Apkinson corpus, combining the outputs from the speech and movement models. The results are observed in Table 9.41. The top-5 speech feature sets used in the fusion include the RAE-based features computed upon the DDK1 (/pa-ta-ka/), monologue, and sentence 8 tasks, the articulation features computed upon the DDK1 task, and the phonological features computed upon the monologue. The top-5 features from the movement analysis include the NLD features computed upon the finger-to-nose, postural tremor, free gait, pronation-supination, and 4x10 tasks. The top speech tasks for the deep learning fusion include the three DDK exercises and sentences 8 and 9. The top movement tasks used in the deep learning fusion include the 4x10, free gait, circles, and finger-to-nose tasks.

The best results combining speech and movement features and using the SVM classifier were obtained with the *Late 2* fusion of the top-5 feature sets (UAR=92.2%), which improves in up to 2.7% the results obtained using only speech features (see Table 9.2), and in up to 10.1% the results obtained using only the movement features (see Table 9.32). The highest UAR obtained combining the deep learning models

is 99.2%, which is 2% higher than the ones observed using only speech signals, and 14.9% than the obtained one using only the movement signals. This result should be taken carefully because the small size of the Apkinson corpus, and need to be validated with additional data collected from more patients and HC subjects.

Table 9.41: Classification of PD patients vs. HC subjects from the Apkinson corpus combining the speech and movement models.

Fusion	**Modalities**	**# Feature sets / Classifiers**	**ACC**	**Fscore**	**UAR**	**SENS**	**SPEC**	**AUC**
Feature extraction and SVMs								
Early fusion Top-5	Speech+Movement	10	86.7	0.830	83.5	75.0	92.0	0.944
Late fusion 1 Top 5	Speech+Movement	10	91.3	0.907	89.5	79.1	100.0	0.960
Late fusion 2 Top 5	Speech+Movement	10	93.2	0.929	92.2	86.0	98.3	0.978
Late fusion 1 all	Speech+Movement	102	86.4	0.863	87.3	93.0	81.7	0.953
Late fusion 2 all	Speech+Movement	102	87.4	0.862	84.9	69.8	100.0	0.987
Deep learning models								
Late fusion 1 Top 5	Speech+Movement	10	95.1	0.951	95.8	100.0	91.7	0.998
Late fusion 2 Top 5	Speech+Movement	10	99.0	0.990	99.2	100.0	98.3	0.998
Late fusion 1 all	Speech+Movement	25	96.9	0.973	96.5	94.7	98.3	0.998
Late fusion 2 all	Speech+Movement	25	99.0	0.990	98.8	97.7	100.0	0.998

The scatter plots for the *Late 1 all fusion* shown in Figure 9.28 show the distribution of the speech and movement scores for the fusion system. For the case of the fusion of traditional features and SVM classifiers in Figure 9.28a) note the overlap that appears close to the decision function. Conversely, for the case of the fusion of the deep learning models in Figure 9.28b) note that there is not such an overlap close to the boundary. The misclassified subjects (1 HC subject and 2 PD patients) are far from the decision boundary of the classifier. Particularly, the misclassified HC subject (71 years old female) in the bottom right part of the figure was misclassified based on her speech model but she was correctly classified based on her movement model. For the case of the two misclassified patients they were as well misclassified based on their speech, but correctly classified based on their movement. These type of visualizations help to better understand the decisions made by each classifier and how each one contributes to the global decision.

The previous experiments show that the fusion of the different bio-signals improves the classification accuracy to discriminate between PD patients and HC subjects. Now the aim is to evaluate whether the fusion of the different modalities also helps to improve the assessment of the motor-state severity of the patients based on the estimation of the MDS-UPDRS-III. The analysis is not performed with the Apkinson corpus because the lack of enough labeled data to train the regression approaches. The results obtained combining the outputs of the different bio-signals to estimate the MDS-UPDRS-III of the patients are shown in Table 9.42. The analysis did not include the results of the deep learning models because they were not accurate for the individual modalities, thus only the models based on feature extraction and SVR regression were considered. The fusion based on the top modalities only considers the feature sets that individually achieved fair Spearman's correlations ($|\rho| > 0.3$). The top results for speech signals include phonological features computed from the DDK1 (/pa-ta-ka/) and the sentences 3, 8, and 9 tasks. The top features for handwriting

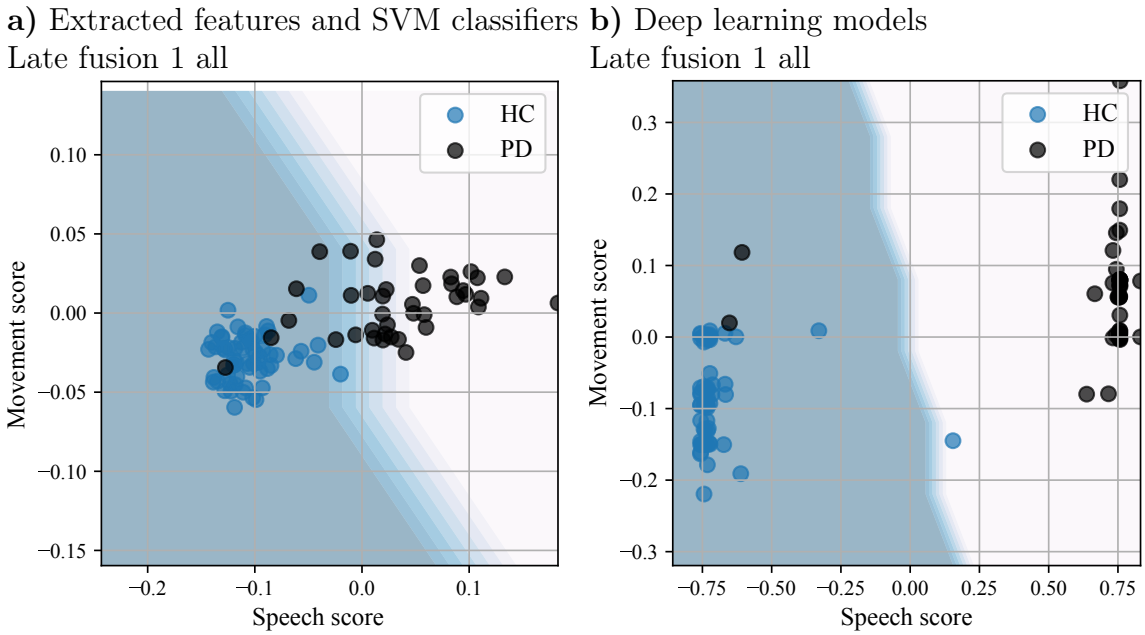

Figure 9.28: Distribution of the classification scores from the subjects of the Apkinson corpus for each modality.

only includes kinematic features computed from the spiral template, numbers, and house tasks. Finally, the top features for the gait signals include spectral features computed from the 2x10 and 4x10 tasks, and NLD features computed from the 2x10, 4x10, and Stop & Go tasks.

Table 9.42: Evaluation of the motor-state severity of patients from the Multimodal corpus based on the MDS-UPDRS-III using the speech, handwriting, and gait features.

Fusion	**Modalities**	**# of Feature sets / Classifiers**	r	r **p-value**	ρ	ρ **p-value**	**MAE**
Feature extraction and SVMs							
Early fusion Top	Speech+Handwriting	7	0.329	0.002	0.329	0.002	10.0
Late fusion 1 Top	Speech+Handwriting	7	0.177	0.024	0.267	0.001	13.5
Late fusion 2 Top	Speech+Handwriting	7	0.535	0.000	0.533	0.000	9.8
Late fusion 1 all	Speech+Handwriting	120	0.042	0.594	0.114	0.147	14.1
Late fusion 2 all	Speech+Handwriting	120	-0.215	0.006	-0.364	0.000	11.1
Early fusion Top	Speech+Gait	9	0.581	0.000	0.633	0.000	8.3
Late fusion 1 Top	Speech+Gait	9	0.220	0.005	0.300	0.000	14.3
Late fusion 2 Top	Speech+Gait	9	0.578	0.000	0.598	0.000	9.1
Late fusion 1 all	Speech+Gait	108	0.116	0.143	0.243	0.002	14.3
Late fusion 2 all	Speech+Gait	108	-0.026	0.742	-0.243	0.002	12.6
Early fusion Top	Handwriting+Gait	8	0.614	0.000	0.594	0.000	8.7
Late fusion 1 Top	Handwriting+Gait	8	0.450	0.000	0.648	0.000	14.1
Late fusion 2 Top	Handwriting+Gait	8	0.594	0.000	0.663	0.000	7.9
Late fusion 1 all	Handwriting+Gait	48	0.525	0.000	0.471	0.000	16.0
Late fusion 2 all	Handwriting+Gait	48	0.350	0.001	0.283	0.007	15.5
Early fusion Top	Speech+Handwriting+Gait	12	0.612	0.000	0.619	0.000	7.4
Late fusion 1 Top	Speech+Handwriting+Gait	12	0.450	0.000	0.648	0.000	14.1
Late fusion 2 Top	Speech+Handwriting+Gait	12	0.594	0.000	0.663	0.000	7.9
Late fusion 1 all	Speech+Handwriting+Gait	138	0.056	0.482	0.129	0.101	14.4
Late fusion 2 all	Speech+Handwriting+Gait	138	0.107	0.174	0.228	0.004	13.8

The highest correlation between the real and estimated MDS-UPDRS-III scores is observed when gait and handwriting signals are combined using the *Late 2* fusion approach of the top features for each modality ($\rho = 0.663$). The same result is also obtained by combining the speech, handwriting, and gait signals. The results obtained here are 80.6% more correlated than the best results obtained using only speech signals ($\rho = 0.367$). The results also improve in 7.6% with respect to the ones reported using only handwriting signals ($\rho = 0.616$), and in 2.6% with respect to the ones obtained using only gait signals ($\rho = 0.646$). The best results are in general obtained with the fusion of the top features for each modality. The correlation is reduced when all features and tasks are considered together.

The last experiment consists of the fusion of the three bio-signals to classify PD patients in different motor-state severity levels based on the MDS-UPDRS-III score, in the same way as the addressed separately for each bio-signal. The results are shown in Table 9.43. The top-5 features considered for the speech signals includes prosody features computed from the DDK2 (/pa-ka-ka/) and the sentence 6, phonation features computed from the sentence 4, and articulation features computed from the sentences 6 and 8. The top-5 handwriting features includes in-air features computed from the circle drawing and the numbers tasks, and kinematic features extracted from the rectangles, the Rey-Osterrieth figure, and the spiral template. Finally, the top-5 gait features includes spectral features extracted from the 4x10 and TUG tasks, and NLD features computed from the 2x10, 4x10, and TUG tasks. These results give an idea about the importance of the different feature sets and tasks for the evaluation of the disease severity of the patients.

Table 9.43: Classification of patients from the Multimodal corpus in three different severity levels based on the MDS-UPDRS-III using the speech, handwriting, and gait models.

Fusion	**Modalities**	**# Feature sets / Classifiers**	**ACC**	**F-score**	**UAR**	**ACC Mild**	**ACC Intermediate**	**ACC Severe**
Feature extraction and SVMs								
Early fusion Top-5	Speech+Handwriting	10	45.8	0.421	46.7	47.0	44.0	56.0
Late fusion 1 Top 5	Speech+Handwriting	10	51.9	0.515	51.9	67.0	40.0	49.0
Late fusion 1 all	Speech+Handwriting	120	51.2	0.497	51.5	59.0	27.0	68.0
Early fusion Top-5	Speech+Gait	10	53.6	0.511	52.8	50.0	56.0	59.0
Late fusion 1 Top 5	Speech+Gait	10	56.8	0.564	56.9	69.0	44.0	58.0
Late fusion 1 all	Speech+Gait	108	53.7	0.522	53.9	69.0	29.0	64.0
Early fusion Top-5	Handwriting+Gait	10	54.2	0.436	53.3	62.0	50.0	40.0
Late fusion 1 Top 5	**Handwriting+Gait**	**10**	**62.5**	**0.612**	**62.4**	**81.0**	**40.0**	**67.0**
Late fusion 1 all	Handwriting+Gait	48	55.7	0.545	55.8	71.0	33.0	63.0
Early fusion Top-5	Speech+Handwriting+Gait	15	55.3	0.467	55.0	62.0	29.0	67.0
Late fusion 1 Top 5	Speech+Handwriting+Gait	15	58.0	0.577	58.1	70.0	45.0	68.0
Late fusion 1 all	Speech+Handwriting+Gait	138	53.1	0.505	53.4	70.0	22.0	68.0
Deep learning models								
Late fusion 1 Top 5	Speech+Handwriting	10	46.9	0.467	47.0	37.0	47.0	57.0
Late fusion 1 all	Speech+Handwriting	33	39.9	0.395	39.6	44.0	47.0	28.0
Late fusion 1 Top 5	Speech+Gait	10	42.4	0.411	42.1	57.0	46.0	24.0
Late fusion 1 all	Speech+Gait	24	41.1	0.394	40.5	55.0	45.0	25.0
Late fusion 1 Top 5	Handwriting+Gait	10	47.1	0.450	49.0	26.0	37.0	85.0
Late fusion 1 all	Handwriting+Gait	21	42.3	0.400	45.5	27.0	36.0	80.0
Late fusion 1 Top 5	Speech+Handwriting+Gait	15	46.9	0.467	47.0	37.0	47.0	57.0
Late fusion 1 all	Speech+Handwriting+Gait	39	48.2	0.475	46.0	38.0	46.0	57.0

The most accurate results are observed again when handwriting and gait signals are combined using the late fusion over the top-5 features for each bio-signal

(UAR=62.4%). This result is 15.9% higher than the reported using only speech signals (UAR=46.5%) and 5.9% higher than the reported only with handwriting signals (UAR=56.5%). The result is 8.2% lower than the observed one only with gait signals (UAR=70.6%, see Table 9.39).

9.5 Analysis of the Experimental Results

The automatic assessment of PD patients is evaluated in this chapter using information from speech, handwriting, and gait signals. Each modality includes a set of different exercises that the participants perform for the analysis.

Speech analysis is based on different dimensions to model the phonation, articulation, prosody, and phonological aspects of the speech. The analysis includes also the evaluation of representation learning strategies based on autoencoders and deep learning methods to characterize and evaluate the speech of the participants. The speech exercises performed by the subjects include DDK tasks like the rapid repetition of the syllables /pa-ta-ka/, read sentences, a phonetically balanced read text, and a monologue. Binary classification experiments i.e., PD vs. HC are addressed to test the suitability of each speech task and each method for the automatic detection of the disease. Additionally, regression experiments are performed to evaluate the dysarthria severity of the subjects according to the proposed m-FDA scale. Multiclass experiments are also performed to group patients in different levels of dysarthria severity (mild, intermediate, and severe). The results indicate that it is possible to classify PD patients and HC subjects with accuracies up to 87.7% when combining the different speech features and speech exercises (see Table 9.1), and up to 96.2% using the deep learning techniques (see Table 9.3). The results also indicate that phonological features are the most accurate to classify PD patients and HC subjects, and that the other feature sets are complementary to improve the accuracy. The most accurate speech exercises to classify PD vs. HC subjects are the DDKs and the read text, which allows to define a battery of exercises used in clinical practice for the assessment of the disease. The classification of PD vs. HC subjects is also performed on the data collected using the Apkinson app. The results indicate that there is not a visible impact in the classification when speech signals collected from smartphones are considered, and that only phonation and prosody analyses are affected because the use of the smartphones.

The accuracy obtained with prosody and phonological features depends to some extend on the phonetic content present in the sentences pronounced by the patients. Particularly, sentences with a higher number of voiced phonemes are the most accurate for the prosody analysis, and sentences with more presence of plosives, fricatives, and vowels are more accurate when considering the phonological analysis. There are other methods like the end-to-end deep learning strategy based on ResNet that do not depend on the phonetic content of the sentence, which indicates that such methods can be used for non-intrusive evaluation of the patients.

The experiments addressed to evaluate the dysarthria severity of the subjects are performed in three scenarios: (1) prediction of the value of the m-FDA scale using regression approaches; (2) longitudinal evaluation of the patients from the Longitudinal and the At-Home corpus using both speaker models and regression algorithms;

and (3) classification of patients in different levels of dysarthria severity (mild, intermediate, severe) using multi-class classification methods. The results evaluating the dysarthria severity using the regression approach show correlations up to 0.61, which are obtained when all speech tasks are considered, and using phonological features. This fact gives an idea about the importance of these features to model the dysarthria severity of the participants. In addition, the most sensitive tasks to evaluate the severity of the speech impairments of the patients are the DDK exercises. This is particularly useful specially considering that those tasks are very easy to produce and are potentially useful to evaluate the speech of patients in almost every language. Unfortunately, the results using the deep learning models are not accurate to evaluate the dysarthria severity of the patients. The main reason would be because the lack of enough labeled data to train the regression models.

The longitudinal evaluation of the dysarthria severity of patients is performed in two scenarios to cover both the long- and short-term progression of the disease, using the Longitudinal and the At-Home corpus, respectively. The disease progression of patients is modeled using two strategies: (1) an SVR regression approach, and (2) unsupervised speaker models based on GMM-UBM. The disease progression in long-term time intervals is predicted with a Spearman's correlation up to 0.51 using phonological features and DDK exercises, similar to previous experiments predicting the m-FDA score of the patients. This result gives insights about the generalization capacity of the trained model to the dysarthria severity of new patients. The results using the GMM-UBM system to predict the disease progression of the patients were not as good as the ones obtained using the supervised model based on the SVR. This result is contrary to the presented previously in [Aria 18a], where the GMM-UBM systems were better. However, for this thesis, nearly the double of training data was available, in comparison with the previously available, which makes the SVR model more accurate than the GMM-UBM system. The disease progression in short-term time intervals is predicted with a Spearman's correlation up to 0.49 using as well phonological features. These results allow to evaluate not only the dysarthria level but also to map the outcomes with the medication intake of the patients. For the At-Home corpus, similar results are obtained with both the SVR and the GMM-UBM systems. In summary, the progression of the dysarthria level both in short- and long-term intervals is predicted with moderate correlations for most of the patients. There are even some patients in the Longitudinal corpus whose dysarthria level progression is predicted with strong correlations. These results suggests that the proposed models are suitable to monitor the progression of the dysarthria level in PD patients both in long- and short-term intervals, as well as the possible influence of the medication intake on the dysarthria severity.

Although there is some correlation between the predicted and the real m-FDA scores, it is more suitable for the patients to know in which stage of the disease they are, rather than to have the prediction of a continuous scale. In addition, for medical applications it is difficult to have a great amount of data to train suitable regression algorithms like an SVR or a CNN, as it was observed in the regression results. For these reasons it is better to divide the patients into different groups according to their disease severity, thus, it is possible to classify patients in mild, intermediate, and severe stages of the disease. The subjects from the Multimodal

corpus were grouped into three classes according to their dysarthria severity based on the m-FDA scale. These classes are defined based on the 33rd and 66th percentiles of the total scale in order to have equal priors for each class (see Figure 9.10). The results indicate that it is possible to classify the patients in different levels of the disease with accuracies of up to 61% using the ResNet-based models. Most of the classification errors correspond to patients in mild and severe stages of the disease that are classified in intermediate stage. This is very positive since they are not mainly misclassified in the other extreme class. These results are also consistent with the disease progression.

The evaluation of the motor-state severity of the patients using the MDS-UPDRS-III scale was also performed using the speech signals. The highest correlations are observed with the phonological features, although none of the models is accurate enough to evaluate the motor-state severity of the patients. This is expected since the MDS-UPDRS-III is a complete motor scale in which only one of the items is related to speech symptoms. Hence it is not suitable nor fair to try to evaluate the full motor-state severity of patients using only speech signals. However, these speech features could provide complementary information when they are combined with handwriting and gait signals.

The handwriting analysis is based on different features to model the kinematic, in-air, and geometric aspects of the strokes. The assessment includes also the evaluation of deep learning methods to characterize and evaluate both the online and offline handwriting information of the participants. The handwriting exercises performed by the subjects include drawing of geometrical shapes like Archimedean spirals, cubes, circles, houses, among others. The protocol include as well writing tasks like the alphabet, a free sentence, the numbers, the name of the participants, among others. Finally, the analysis includes the drawing of the Rey Osterrieth figure, which is a standard neurophysiological test to evaluate cognitive aspects of the participants [Shin 06]. The experiments include binary classification to test the suitability of each handwriting task and each method for the automatic detection of the disease. Additionally, regression experiments are performed to evaluate the motor-state severity of the patients according to the MDS-UPDRS-III scale. Finally, multi-class experiments are also performed to classify patients in different levels of the disease severity (mild, intermediate, and severe), according to the MDS-UPDRS-III scale. The results indicate that it is possible to classify PD patients and HC subjects with accuracies up to 97.8% when combining the different handwriting features and tasks (see Table 9.21). The results also suggest that early fusion is better to model the different and complementary information produced by each feature set, and that late fusion methods deal better with the redundant information that appears when using the same features computed from different tasks. On the one hand the more complex exercises like the alphabet, the free writing or the Rey figure are better modeled with in-air features. These more complex tasks have a lot of pen-up and pen-down transitions, which are accurately characterized with the proposed in-air features. On the other hand, the simple drawing exercises like Archimedean spirals and circles are better modeled with the combination of kinematic and in-air features because they do not contain that high amount of transition movements. Unfortunately, the proposed geometric features do not provide the expected accuracy. Additional analyses

and features can be proposed to model the trajectory and accuracy of the strokes performed by the patients when drawing different geometric shapes.

The results using the deep learning methods to classify PD patients and HC subjects using handwriting signals show accuracies of up to 99.2% using online handwriting samples from the pen-up and pen-down transitions (see Table 9.22). The result obtained for the offline handwriting model are not that accurate compared to the ones obtained with the online models, which suggest that there is important information in the handwriting aspects of PD patients that is only available with the online analysis. The comparison between the results observed for both online models (transitions and full segments) shows that there are important differences depending on the addressed handwriting task. Simple drawing shapes are generally better modeled with the *Full segment* models. Conversely, the most complex tasks like the alphabet are better modeled with the network that only consider the pen-up and pen-down transitions. Finally, the results observed with the Archimedean spirals are not as good as expected, having in mind that this is one of the most used handwriting tasks for the assessment of PD patients in the literature (see Table 5.1). This aspect has to be carefully considered when designing evaluation protocols for the assessment of patients.

The experiments addressed to evaluate the motor-state severity of the subjects are performed in three scenarios using the handwriting signals: (1) the prediction of the value of the MDS-UPDRS-III scale using regression approaches; (2) the disease progression evaluation of the patients from the Longitudinal corpus using both user models and regression algorithms; and (3) the classification of patients in different levels of the motor state severity (mild, intermediate, severe) using multi-class classification methods. The results evaluating the MDS-UPDRS-III severity using the regression approach show correlations up to 0.61, which are obtained when all handwriting tasks are considered, and using the kinematic features. Unfortunately, the results using the deep learning models are not accurate to evaluate the motor-state severity of the patients. The main reason is the lack of labeled data to train the regression models.

The longitudinal evaluation of the motor-state severity of patients is performed using the handwriting signals with the same two strategies considered for the speech modeling: (1) an SVR regression approach, and (2) unsupervised user models based on GMM-UBM. These methods are used to predict the progression of the MDS-UPDRS-III of patients from the Longitudinal corpus. The disease progression is predicted with a Spearman's correlation up to 0.70 using the kinematic features and the fusion of all handwriting tasks. In addition, the most complex tasks like writing the name, the numbers, or writing a sentence are the most accurate ones to predict the disease progression of the patients. The results using the GMM-UBM system to predict the disease progression of the patients were not as accurate as the ones obtained using the supervised model based on the SVR, similar to the results observed for the case of speech signals.

Although the strong correlations obtained between the predicted and the real MDS-UPDRS-III scores, and similar to the experiments addressed with the speech signals, it is more suitable for the patients to know in which stage of the disease they are, rather than to have the prediction of a continuous scale. The patients were

divided into three groups according to their disease severity, thus it is possible to classify patients in mild, intermediate, and severe stages of the disease. The division for the three classes was performed using the 33rd and 66th percentiles of the total MDS-UPDRS-III scale, in order to have equal priors for each class (see Figure 9.12). The results indicate that it is possible to classify the patients in different levels of the disease with accuracies up to 56.5% using both the deep learning models and the fusion of all feature sets and handwriting tasks. The models are very accurate to detect patients in severe stages of the disease. Patients in intermediate state are very difficult to classify. These results suggest that would be better and easier to classify patients in two levels of the disease (divided by the median) rather than the classification in the three levels addressed here.

The gait analysis is based on features to model the kinematic, spectral, and non-linear aspects of the walking process of the patients. Kinematic features include different measurements to model different properties in the strides such as time, distance, and velocity of each step. Spectral features are designed to model the spectral wealth and the harmonic structure of the gait signals, and include features like the Freeze index [Zach 15], which is highly used to evaluate the presence of FoG episodes in PD patients. Finally, NLD features are designed to model stability, regularity, and long-range autocorrelations of the gait signals. The methods include as well a proposed CNN-GRU neural network to process the raw gait signals. The gait analysis is performed both with the data collected using the eGait inertial sensors and using the Apkinson app. Different gait exercises are collected and analyzed with the eGait system, including short walk exercises like 2x10 and 4x10 tasks, in addition to specifically designed tasks such as Stop & Go, heel toe tapping, and the TUG test. For the case of the Apkinson data additional hand movement exercises like Finger-to-nose, Circles, Postural tremor, and Pronation / Supination are included in the analysis. The performed experiments include binary classification to test the suitability of each model to discriminate between PD patients and HC subjects. Additionally, regression experiments are performed to evaluate the motor-state severity of the patients according to the MDS-UPDRS-III scale. Finally, multi-class experiments are performed to classify patients in different levels of the disease severity (mild, intermediate, and severe). The results indicate that it is possible to classify PD patients and HC subjects with accuracies up to 86.2% using NLD features, and up to 98.7% using the deep learning techniques. The most accurate gait exercise to classify PD vs. HC subjects is the Stop & Go task, which is the one when the patients have to perform more start/stop movements of the lower limbs.

The results with the Apkinson data indicate a reduction in the accuracy of about 15%. This result is explained because in the smartphone data only one 3-axial accelerometer is available, compared to the eGait system where both 3-axial accelerometers and gyroscopes are attached to each foot. However, the smartphone offers the cheapest solution to evaluate the gait of PD patients. With the observed accuracy (up to 84.4% using the CNN-GRU), it is possible to perform a preliminary evaluation of the patient at-home. Then if some change is detected in the gait and movement of the patient, (s)he can go to the clinic to be evaluated with more robust system like the eGait.

The evaluation of the MDS-UPDRS-III score of the patients using gait signals is performed with kinematic, spectral, and NLD features. The motor state-severity can be estimated with Spearman's correlation of up to 0.65 combining all feature sets and gait tasks. In addition, the best result is obtained for many of the gait tasks when only the NLD features were considered, which validates the importance of these features to model the walking process of PD patients. High correlations are obtained also with spectral features computed upon the 4x10 task (ρ=0.62), which is very close to the best result achieved.

The longitudinal evaluation of the motor-state severity is addressed with the gait signals with the same two strategies considered for the speech and handwriting modeling: (1) an SVR regression approach, and (2) user models based on GMM-UBM. These methods are used to predict the progression of the MDS-UPDRS-III of patients from the Longitudinal corpus. The disease progression is predicted with a Spearman's correlation up to 0.77 using the SVR and combining spectral and NLD features from the 4x10 task. Strong correlations are also obtained considering separately spectral and NLD features for some of the gait tasks. This is very positive considering the fact that these results are obtained using an independent test set that was never seen during a cross-validation strategy. In addition, the results using the GMM-UBM system to predict the disease progression of the patients were not as accurate as the ones obtained using the supervised model based on the SVR, similar to the observed for the case of speech and handwriting signals.

The classification of patients in different levels of the disease severity is also performed with gait signals in a similar way to the addressed with speech and handwriting signals. Three classes are defined based on the 33rd and 66th percentiles of the total MDS-UPDRS-III scale in order to discriminate between mild, intermediate, and severe levels of motor-state severity. The results indicate that it is possible to discriminate among the three severity levels with an accuracy up to 70.6%, which is 11.6% higher than the best result observed with handwriting signals. The best result is also obtained with the use of NLD features, similar to the bi-class problem and the estimation of the MDS-UPDRS-III using the regression approach. These results confirm the importance of the NLD analysis to model the different gait impairments that appear due to neurodegeneration [Dier 17, Chom 19, Pere 20b]. In addition, the 4x10 and the TUG tasks exhibit on average higher accuracies than the ones observed in other tasks like heel-toe tapping and 2x10. The results using the signals from the Apkinson app indicate that there are differences of up to 10% with respect to the results observed for the Multimodal corpus. These differences are explained again because in the smartphone data only one 3-axial accelerometer is available, compared to the eGait system where both a 3-axial accelerometer and a gyroscope are attached to each foot. The best result for the Apkinson data was obtained with spectral features computed over the posture exercise (60.0%). Additional data should be collected and labeled in order to improve the results using the different proposed methods.

The fusion of the speech, handwriting, and gait signals is performed both with early and late fusion strategies. The fusion was performed considering both the top-5 methods that produced the best results with each modality, and with the total of methods applied to each bio-signal. The best result using the traditional feature extraction techniques was obtained with the early fusion of the top-5 features from

the handwriting and gait signals (UAR=99.2%). This results improved in 12.2% the ones obtained using only speech signals, in up to 3.3% the ones obtained with handwriting signals, and in up to 4.9% the results obtained with gait signals. No high differences are observed when using the early or late fusion strategies, neither using the top-5 or the fusion of all features and tasks. However, the results using the top-5 models are slightly more accurate. The fusion using the deep learning techniques shows that the best result was also obtained combining the handwriting and gait signals (UAR=99.4%). These results were also better than the reported ones for the individual bio-signals.

The speech and movement signals from the Apkinson corpus were also combined. The best result using the SVM classifiers was obtained with the late fusion of the top-5 feature sets from each modality (UAR=92.2%), which improved in up to 2.7% the results obtained using only speech features, and in up to 10.1% the results obtained using only the movement features. The highest UAR combining the deep learning models was 99.2%, which is 2% higher than the observed one using only speech signals, and 14.9% than the obtained using only the movement signals. This result should be taken carefully because the small size of the Apkinson corpus, and they need to be validated with additional data collected from more patients and HC subjects.

The fusion of speech, handwriting, and gait features also improved the assessment of the motor-state severity of patients based on the MDS-UPDRS-III scale. The best result was observed when gait and handwriting signals are combined using the late fusion approach of the top features for each modality ($\rho = 0.663$). The same result was also obtained combining the speech, handwriting, and gait signals. The results improved in up to 80.6% the best results obtained using only speech signals ($\rho = 0.367$), in 7.6% the ones reported using only handwriting signals ($\rho = 0.616$), and in 2.6% the ones obtained using only gait signals ($\rho = 0.646$).

The last experiment consisted of the fusion of the three bio-signals to classify patients in three levels of the disease severity based on the MDS-UPDRS-III score. The results of the multimodal system outperformed those reported using only speech and handwriting signals, but not the ones obtained using gait signals, which are the most accurate for the addressed problem with an accuracy of up to 70.6%.

Chapter 10

Outlook

The methods and techniques proposed in this thesis can be extended to other applications. A potential scenario that can be considered is the automatic evaluation of patients affected by other neuro-degenerative disorders with similar symptoms such as Huntington's disease or essential tremor. It is particularly known that Huntington's disease is more invasive than PD, producing more aggressive motor and cognitive impairments. This is particularly important from demographic reasons because there is evidence about the high incidence and clusters of genetically affected Huntington's disease patients in Venezuela and in the north coast of Colombia [Para 08, Cast 16]. The automatic assessment of patients affected by Huntington's disease will help to get better treatment to slow-down the progression of the disease. At the same time, the proposed methodologies can be extended to problems related to automatic discrimination between patients affected by different neuro-degenerative diseases.

The proposed methods can also be potentially used to detect prodromal stages of the PD, which would benefit the development of future neuro-protective therapies [Post 15]. There is evidence showing that the detection of prodromal stages of PD is possible from speech [Hlav 17] and gait [Alib 16]. The main difficulty for these studies is to find patients in pre-clinical stages, i.e. before the disease appears or is at least observable by clinicians. Once the target group is found, it is required to start the monitoring of patients over time in order to understand which are the patterns that become abnormal when early signs of the disease appear. This is particularly important given the fact that the north of Antioquia (Colombia) is perhaps one of the areas with the highest prevalence of genetic PD [Pine 06]. Another population that can be considered to study prodromal stages of the disease includes patients suffering from rapid eye movement (REM) sleep behavior disorders (RBD). According to recent studies, there is a high probability for these people to develop PD within their next twelve years of life [Post 09, Hlav 17].

An additional scenario that can be potentially covered using the proposed methods includes the analysis and classification of patients affected by sporadic PD, and those affected by different genetic mutations. The aim is to evaluate how they differ in terms of disease progression and aggressiveness of the disease. Recent studies suggest that patients affected by mutations in the GBA gene are characterized by higher rates of dementia and a faster progression of the disease, while patients affected by

mutations in the LRRK2 and dardarin genes are characterized by an early onset and slow progression of the disease [Gan 10, Yaha 19].

Some of the results obtained in this thesis should be validated with data collected from additional patients in more recording sessions. For instance, the classification of PD patients and HC subjects using the data collected using Apkinson should be performed with additional smartphone data from several brands and with a higher number of participants, in order to have more conclusive results. At the same time, longitudinal data should be collected using Apkinson with the aim to track the disease progression per patient. Finally, longitudinal analyses of handwriting and gait should be performed as well with more recording sessions and subjects to validate the results obtained in this thesis. The longitudinal analysis will help not only to monitor the disease progression per patient, but also to evaluate the medication intake, and how the pharmacotherapy affects the response of the upper and lower limbs, and the speech production process.

The proposed methods can be complemented with additional modalities like facial images in order to evaluate more symptoms exhibited by PD patients like hyphomimia, which is the lost of facial expression [Gome 21]. At the same, this thesis only covers asynchronous multimodal analysis of the patients. Hence, it is important to consider as well synchronous multimodal analysis. This can be performed with the joint analysis of speech, facial images, and text; or by considering dual activities e.g, reading a text while while carrying an object [Sama 18]. Further research should be performed to develop robust deep learning strategies to combine these synchronous modalities.

The speech analysis of PD patients can be extended by a more in depth evaluation of which speech dimensions are more affected for each patient individually. This can be performed with a more detailed evaluation of phonological and phonetic features to model specific speech deficits of the patients. For instance, it is clinically useful to detect whether the patients have more problems in pronouncing plosives, fricatives, or nasal sounds. This would help the expert phoniatrician to select a more specific and focused therapy for the PD patients. This analysis can be complemented with the use of the sub-items of the m-FDA scale, and using specific phonological features for each item. For instance to consider only features based on labial phonemes to evaluate the items related to lip movements, or dental phonemes to evaluate the items related to tongue movements, among other analyses.

The analysis of specific symptoms of the disease can be extended to the domain of handwriting and gait signals. The analysis can be extended to evaluate specific impairments of the upper and lower limbs by considering only those items from the MDS-UPDRS-III score that are related to the features computed from each modality. Hence those items related to the gait process can be better modeled using specific gait features, like those designed to detect FoG episodes in the patients like the FI. Conversely, online handwriting analysis can be used to evaluate only those items of the MDS-UPDRS-III related to the movement of upper limbs.

Regarding handwriting analysis, one important aspect to be considered is that most of the studies are based mainly on kinematic and pressure features, which limits the scope of the models. This thesis extends previous studies by proposing a set of features to model the in-air movements of the handwriting process, which is

very accurate to classify PD patients vs. HC subjects, but it is not robust enough to evaluate the motor-state severity of the patients. Additional handwriting features should be proposed to assess other aspects of PD dysgraphia such as fluency or size. Preliminary results from a Bachelor thesis in our lab suggest that it is possible to obtain features to model handwriting tremor by modeling the strokes performed by the patients using a spectral-based approach [Kupf 20]. In addition, the use of small neural networks like the SqueezeNet used in this thesis makes it possible to use trained models to evaluate handwriting images in low-power devices like smartphones. These types of models can be included in further releases of Apkinson, where a patient can perform one or several handwriting exercises using normal pen and paper, and then take a picture with his/her smartphone, which will be processed locally to evaluate the upper motor skills of the patients.

The analysis of motor symptoms using Apkinson can be complemented using the exercises included in the app to evaluate fine-motor skills such as finger-taping. This type of exercises has not been extensively studied yet, and they can be a good complement to the models proposed in this thesis to evaluate the motor-state severity using smartphone data from speech, gait, and hand movements.

Finally, this thesis only covered the assessment of motor impairments of PD patients. However, the evaluation of non-motor symptoms such as depression, anxiety, or cognitive decline is also important and can be automated by using state-of-the-art technologies of speech and natural language processing. These techniques have been successfully applied in patients with other neurodegenerative disorders such as Alzheimer's disease [Pere 21a, Pere 21d], and have been recently considered to evaluate language aspects of PD patients [Pere 19, Garc 16, Nore 20, Garc 21].

Chapter 11

Summary

The aim of this thesis was to develop robust models for the accurate diagnosis of PD and to evaluate the disease severity of patients using different bio-signals such as speech, online handwriting, gait, and those signals collected from smartphones. Identifying accurate bio-markers for early and differential diagnosis, severity, and response to therapy is a primary goal of the research on PD today. A computerized approach for continuous monitoring of the state of the patients will help in slowing down the impact of PD, and to improve the quality of life of patients.

The proposed models based on speech, handwriting, gait, and smartphone data are evaluated in three main scenarios: (1) The automatic classification of healthy subjects and PD patients. (2) The evaluation of the disease severity of the patients based on a clinical scale, including both the motor-state severity and the dysarthria level of the subjects. (3) The classification of PD patients into different groups according to their disease severity e.g., mild, intermediate, and severe. For these applications, two different machine learning paradigms for automatic classification and regression are considered: (1) a traditional pattern recognition approach using SVMs and GMMs, and (2) a novel approach based on deep learning methods for an end-to-end analysis of the data collected from the patients.

The ground truth to label the motor-state severity of the patients is based on the MDS-UPDRS-III scale. This scale has only one item out of 32 to evaluate the speech of the patients. However, the speech production is highly affected in PD, thus it makes sense to consider a specific scale to evaluate the severity of the speech impairments. The recently introduced m-FDA scale is considered as a ground truth to evaluate the dysarthria severity of the patients. The scale can be administered considering only speech recordings, thus it can be applied remotely and represents a step towards the automatic administration of speech and language therapy for PD patients.

Four databases are considered for the validation of the methods proposed in this thesis. The first one is the Multimodal corpus, which comprises speech, online handwriting, and gait signals signals from 106 PD patients and 105 HC subjects. The speech protocol includes utterances of the vowel /ah/, DDK exercises, isolated sentences, a read text, and a monologue. Handwriting data include drawing geometrical shapes and writing tasks. Gait data are collected with the eGait system, which consists of inertial sensors attached to the lateral heel of the shoes. The gait protocol

includes different walking and heel-toe tapping exercises. The second database is the Longitudinal corpus, which is designed to evaluate the impact of the motor deficits of the patients in long-term. The corpus includes data from nine PD patients recorded in seven sessions from 2012 to 2019. The third database is the At-home corpus, which is considered to monitor the progress of the speech deficits of PD patients in short-term periods of time. This corpus comprises a group of seven PD patients recorded four times per day (every two hours), once per month during four months. The last database is the Apkinson corpus, which includes speech and movement signals collected with the Apkinson android application, and which includes at the moment data from 38 PD patients and 60 HC subjects.

The automatic assessment of the speech of PD patients has been classically modeled in terms of phonation, articulation, prosody, and intelligibility. In the last five years, deep learning methods started as well to be used to evaluate the speech of PD patients. The methods considered in this thesis to model speech signals include the ones classically addressed in the literature to model phonation, articulation, and prosody aspects, which are used as baselines for the proposed approaches. Three approaches are introduced to model the speech of PD patients. The first one comprises phonological analysis of speech signals using deep learning methods. Phonological features are more interpretable for clinicians since they encode information about the mode and manner of articulation, which is specifically related with the movements of the articulators in the vocal tract. A model to extract phonological features is proposed and released as a toolkit called *Phonet* to extract phonological posterior probabilities from speech. The model is based on a bidirectional RNN with GRU units trained to detect the presence of 18 phonological classes in speech frames using a multitask learning strategy. The phonological features are obtained at the output of the neural network as the conditional posterior probability of a speech frame to belong to one or more phonological classes. These phonological posteriors are transformed into phonological log-likelihood ratio features to model the capabilities of the speakers to pronounce different groups of phonemes. The transformed log-likelihood ratio features overcome the non-Gaussian nature of phonological posteriors, which is better to exploit different classification methods. The second proposed approach to model the speech of PD patients involves the use of representation learning strategies using recurrent autoencoders, which have the potential to extract more abstract and robust features than those traditionally computed. A recurrent autoencoder is trained to characterize the temporal structure of input spectrograms. Two different feature sets are computed using the trained autoencoder: the bottleneck features in the encoder's output, and the reconstruction error between the input and the decoded spectrograms in different frequency bands. Finally, the third proposed model is based on CNNs trained to process time-frequency representations of the speech of PD patients. Two different spectral representations are considered as input for the neural network: (1) Spectrograms of onset and offset transitions to evaluate the capabilities of the patients to start/stop the vibration of the vocal folds. (2) Continuous speech segments with the aim to model the full temporal and spectral information from the speech of PD patients.

Handwriting impairments in PD patients are traditionally addressed using kinematic and pressure features that only model some of the handwriting aspects of the

disease. The use of deep learning methods has increased as well within the last years to model handwriting impairments of PD patients. The analysis performed in this thesis involves the computation of kinematic, geometric, and in-air features; in addition to the use of deep learning methods. Kinematic features are based on the trajectory, velocity, and acceleration of the strokes, both in the horizontal, vertical, radial, and angular axes. These features also include measures based on the pressure of the pen and those based on the azimuth and altitude angles. The geometric analysis aims to model geometric shape and symmetry aspects in Archimedean spirals drawn by the patients. The trajectory of the spirals is modeled as an amplitude-modulated signal. The feature set is based on the parameters of the modeled trajectory and the error between the real and modeled trajectories. The in-air features are based on the transitions between in-air and on-surface segments in order to model the difficulties of patients to start/stop movements in the upper limbs. In-air features include the number of pen-ups and pen-downs per second to model hesitations to start or to stop writing, the slopes of the pen-up and pen-down transitions to model the stability of the hand when placing/lifting the pen from the tablet surface, and the percentage of time in-air, among others. The proposed deep learning models are designed to process both online and reconstructed offline handwriting data. The neural network to process the online handwriting is formed with a stack of three 1D-convolution layers to process the raw signals collected from the tablet. Two different inputs are considered for the CNN. The first one comprises the difference among consecutive samples in order to transform the signal from a point-level sequence, which depends on the position of the tablet, into a stroke-level sequence, which represents the direction of the pen movement. The second input sequence for the CNN corresponds to the pen-up and pen-down transitions to evaluate the difficulties observed in the patients when they start/stop the handwriting movements. Finally, the online handwriting data are processed to reconstruct the images drawn by the patients, thus it is possible to have similar images to what a patient would draw with a normal pen and paper. These reconstructed offline images are processed with a CNN based on the SqueezeNet architecture, using a transfer learning strategy from ImageNet.

Gait analysis of PD patients is commonly addressed with inertial sensors attached to the body of the subjects. Most studies consider kinematic features based on the duration and velocity of the steps, or spectral features to evaluate the harmonic structure of the gait signals. There are some studies that aim to model the non-linearities that appear during the walking process using NLD features. Deep learning methods have also gained attention from the research community to evaluate Parkinsonian gait. The analysis performed in this thesis involves the computation of kinematic, spectral, and NLD features; in addition to deep learning methods to model gait impairments of PD patients. Kinematic features include different properties in the strides such as time, distance, and velocity. These features are used as a baseline and they are based on the methods proposed in [Bart 17]. Spectral features are designed to model the spectral wealth and the harmonic structure of the gait signals. The features are based on the CWT extracted from the time series. The feature set is formed with the energy content in 8 frequency bands and three spectral centroids of the wavelet spectrum. The features also include the energy content in the locomotor (0.5–3 Hz) and freeze (3–8 Hz) bands, and the freeze index. NLD features aim

to model the local dynamic stability, recurrence, and complexity properties of the walking process. The features include the computation of the correlation dimension, the largest Lyapunov exponent, the Hurst exponent, the Lempel-Ziv complexity, the sample entropy, and the detrended fluctuation analysis. Finally, the proposed deep learning models are based on 1D-convolutions to learn a filter bank from the raw gait signals, followed by a stack of two bidirectional GRU layers to model the temporal structure of the sequence. The proposed network includes at the end a layer with an attention mechanism with the aim to learn and give more importance to specific parts of the gait sequence, e.g. pauses, swing phase, or stance phase. For the particular case of the data collected using Apkinson and due to the fact that the smartphone can be always placed in a different orientation when performing the gait tasks, a data augmentation strategy is proposed by randomly switching the axes of the inertial sensors in the input to the deep learning model.

Multimodal analysis of speech, handwriting, and gait of PD patients imposes challenges in terms of information perception and data fusion strategies. Different modalities can be complementary, redundant, or even conflicting. For instance, we can have PD patients with a healthy or "normal" handwriting, but a very impaired gait or speech. Multimodal analyses have not been extensively studied for PD analysis. The main reason is because the lack of available data. Fusion of different modalities is a relevant task, which can be executed at data-, feature- and decision-levels. An example of data-level fusion is the combination of accelerometer and gyroscope signals for the gait analysis, or the fusion of position and pressure for the handwriting analysis. Then, fusion at feature-level or early fusion is a general type of fusion when speech, handwriting, and gait features, or features from different tasks within the same modality are stacked together, before the classification. Finally, the decision-level or late fusion consists of training individual models with data from each modality, and then combine the local decisions using a set of rules to get a global decision. The experiments addressed in this thesis are carried out using both early and late fusion strategies.

The analysis of PD patients using smartphone technologies is carried out using the Apkinson app. The aim of Apkinson is to monitor the disease progression of PD patients and to make a motor evaluation that includes specifically the speech production. Apkinson records several signals using sensors embedded on the smartphone (microphone, accelerometer, and the touch screen). The App incorporates exercises and models for speech, movement, and finger tapping. The patient receives individual feedback with the results of the exercises with the aim to motivate them to continue using the App and trying to perform better every day. The speech analysis in Apkinson is focused on evaluating phonation, prosody, and intelligibility of PD patients. The movement evaluation in the upper and lower limbs is performed in Apkinson with measures of regularity of movements, FoG, hand tremor, postural stability, and gait dynamics. Finally, the evaluation of fine-motor skills of the patients is performed with tapping exercises based on the tapping accuracy, velocity, and precision.

The results using speech signals indicate that it is possible to classify PD patients and HC subjects with accuracies up to 87.7% combining the different speech features and tasks, and up to 96.2% using the proposed CNNs. The results evaluating the dysarthria severity of the patients according to the m-FDA scale show correlations

up to 0.61 by using all speech features and tasks. Unfortunately, the results using the deep learning models are not accurate to evaluate the dysarthria severity of the patients. The main reason is because the lack of enough labeled data to train the regression models. Phonological features are the most accurate to evaluate the speech of PD patients. The other feature sets are complementary to improve both the classification and the disease severity evaluation. The most accurate speech tasks are the DDKs and the read text, which allow to define a battery of exercises used in clinical practice for the assessment of the disease. In addition, the results indicate that there is not a visible impact in the classification when considering speech signals collected from smartphones. Only phonation and prosody features are negatively affected due to the use of smartphones, which implies changing the recording conditions. The progression of the dysarthria level both in short- and long-term intervals is predicted with moderate correlations for most patients. There are even some patients in the Longitudinal corpus whose dysarthria level progression is predicted with strong correlations. These results suggest that the proposed models are suitable to monitor the progression of the dysarthria level in PD patients both in long- and short-term intervals, as well as the possible influence of the medication intake in the dysarthria severity. The results also indicate that it is possible to classify the patients in three different speech severity levels with accuracies up to 61% using the proposed CNN models. Most classification errors correspond to patients in mild and severe stages of the disease that are classified in intermediate stage. Finally, none of the speech-based models is accurate enough to evaluate the motor-state severity of the patients based on the MDS-UPDRS-III score. This is expected since the MDS-UPDRS-III is a complete motor scale in which only one of the items is related to speech symptoms.

The results using handwriting signals indicate that it is possible to classify PD patients and HC subjects with accuracies up to 97.8% combining the different handwriting features and tasks, and up to 99.2% using the CNNs to process the online handwriting samples from the pen-up and pen-down transitions. The results suggest that early fusion is better to model the complementary information produced by each feature set, and late fusion methods deal better with the redundant information that appears when using the same features computed from different tasks. More complex exercises like free writing are better modeled with the in-air features because they have a lot of pen-up and pen-down transition movements. Conversely, simple drawing exercises like Archimedean spirals are better modeled with the combination of kinematic and in-air features because they do not contain that high amount of transition movements. The result obtained using the CNN to process the reconstructed offline handwriting model are not that accurate compared to the ones obtained with the online models. This fact suggests that there is important information in the handwriting aspects of PD patients that is only available with the online analysis. The results evaluating the MDS-UPDRS-III severity show correlations up to 0.61, which are obtained using kinematic features and combining all handwriting tasks. The results using the deep learning models are not accurate to evaluate the motor-state severity of the patients. The disease progression per patient in the Longitudinal corpus is also predicted with strong correlations ($\rho = 0.70$) using the kinematic features and the fusion of all handwriting tasks. Finally, the results indicate that it is possible to classify patients in different levels of the disease with accuracies up to 56.5% using

both the deep learning models and the fusion of all feature sets and handwriting tasks. The models are very accurate to detect patients in severe stages of the disease, while patients in intermediate state are very difficult to classify.

The results using gait signals include both the ones obtained using the high-quality sensors of the eGaIT system and the signals collected using Apkinson. The results indicate that it is possible to classify PD patients and HC subjects with accuracies up to 86.2% using NLD features, and up to 98.7% using the deep learning techniques. The most accurate gait exercise is the Stop & Go task, which is the one when the patients have to perform more start/stop movements of the lower limbs. The results with the Apkinson data indicate a reduction in the accuracy of about 15%. This is explained because in the smartphone data only one 3D accelerometer is available, compared to the eGait system where both 3D accelerometers and a gyroscopes are attached to each foot. With the observed accuracy in Apkinson (up to 84.4% using the deep learning models), it is possible to perform a preliminary evaluation of the patient in at-home environments. Then if there is some change detected in the gait and movement of the patient, (s)he can go to the clinic to be evaluated with a more robust system like the eGaIT. The results evaluating the MDS-UPDRS-III severity show correlations up to 0.65 combining all feature sets and gait tasks. The disease progression for the Longitudinal data is predicted with a Spearman's correlation up to 0.77, combining spectral and NLD features from the 4x10 task. These results are very positive because they were obtained as an independent test set that was never seen during a cross-validation strategy. The results also indicate that it is possible to discriminate among three disease severity levels with an accuracy up to 70.6%, which is 11.6% higher than the best result observed with handwriting signals.

The most accurate results combining the information from speech, handwriting, and gait are observed when handwriting and gait signals are combined (99.2%) using the traditional extracted features and SVM classifiers. This result improves in 12.2% the ones obtained using only speech, in up to 3.3% the ones obtained with handwriting, and in up to 4.9% the ones obtained with gait signals. The fusion using the deep learning models shows that the most accurate results were also obtained combining the handwriting and gait signals (99.4%). These results are also higher than the ones reported for the individual bio-signals. The speech and movement signals from the Apkinson corpus were also combined. The best result using the SVM classifiers was obtained with the late fusion of the feature sets from each modality (92.2%), which improved in up to 2.7% the results obtained using only speech, and in up to 10.1% the ones obtained using only movement features. The fusion of speech, handwriting, and gait features also improved the assessment of the motor-state severity of patients based on the MDS-UPDRS-III scale. The best result was observed when gait and handwriting signals are combined ($\rho = 0.66$). The results improved in up to 80.6% the results obtained using only speech ($\rho = 0.37$), in 7.6% the ones reported using only handwriting ($\rho = 0.62$), and in 2.6% the ones obtained using only gait signals ($\rho = 0.65$). The last experiment consisted in the fusion of the three bio-signals to classify patients in three severity levels. The results of the multimodal system outperform those reported using separately speech and handwriting signals, but not the ones obtained using gait signals, which are the most accurate for the addressed problem with an accuracy of up to 70.6%.

Appendix A

Publications emerging from the development of this work

This appendix comprises a list of the most relevant publications derived from my doctoral studies. Of course all these papers are the result of the work and effort made by many colleagues and students who helped me doing several things like collecting data, running experiments, writing and/or correcting parts of the manuscripts, and discussing results, among others.

All experiments addressed in this thesis and that were previously published were re-evaluated and repeated with the updated data considered for this thesis. In addition, note that all my first author publications were written mainly by myself. Thanks to all my co-authors for their great work.

Publications related with speech assessment of Parkinson's disease

Journal papers

- A. M. García, T. Arias-Vergara, **J. C. Vasquez-Correa**, E. Nöth, M. Schuster, A. E. Welch, ... & J. R. Orozco-Arroyave (2021). Cognitive Determinants of Dysarthria in Parkinson's Disease: An Automated Machine Learning Approach. Movement Disorders.

- **J. C. Vasquez-Correa**, C. D. Rios-Urrego, T. Arias-Vergara, M. Schuster, J. Rusz, E. Nöth, & J. R. Orozco-Arroyave, (2021). Transfer learning helps to improve the accuracy to classify patients with different speech disorders in different languages. Pattern Recognition Letters, 150, 272-279.

- V. Mendoza Ramos, **J. C. Vasquez-Correa**, R. Cremers, L. Van Den Steen, E. Nöth, M. De Bodt, & G. Van Nuffelen, (2021). Automatic boost articulation therapy in adults with dysarthria: Acceptability, usability and user interaction. International Journal of Language & Communication Disorders.

- P. A. Perez-Toro, P. A., **J. C. Vasquez-Correa**, T. Bocklet, E. Nöth, & J. R. Orozco-Arroyave (2021). User State Modeling Based on the Arousal-

Valence Plane: Applications in Customer Satisfaction and Health-Care. IEEE Transactions on Affective Computing (IN PRESS).

- **J. C. Vasquez-Correa**, T. Arias-Vergara, M. Schuster, J. R. Orozco-Arroyave, & E. Nöth, (2020). Parallel Representation Learning for the Classification of Pathological Speech: Studies on Parkinson's Disease and Cleft Lip and Palate. Speech Communication, 122, 56-67.

- T. Arias-Vergara, P. Arguello-Velez, **J. C. Vásquez-Correa**, E. Nöth, M. Schuster, M. C. González-Rátiva, & J. R. Orozco-Arroyave (2020). Automatic detection of Voice Onset Time in voiceless plosives using gated recurrent units. Digital Signal Processing, 104, 102779.

- J. R. Orozco-Arroyave, **J. C. Vásquez-Correa**, J. F. Vargas-Bonilla, et al. (2018). NeuroSpeech: An open-source software for Parkinson's speech analysis. Digital Signal Processing, 77, 207-221.

- **J. C. Vásquez-Correa**, J. R. Orozco-Arroyave, T. Bocklet, & E. Nöth (2018). Towards an automatic evaluation of the dysarthria level of patients with Parkinson's disease. Journal of communication disorders, 76, 21-36.

- T. Arias-Vergara, **J. C. Vásquez-Correa**, & J. R. Orozco-Arroyave, (2017). Parkinson's disease and aging: analysis of their effect in phonation and articulation of speech. Cognitive Computation, 9(6), 731-748.

- M. Cernak, J. R. Orozco-Arroyave, F. Rudzicz, H. Christensen, **J. C. Vásquez-Correa** & E. Nöth, (2017). Characterisation of voice quality of Parkinson's disease using differential phonological posterior features. Computer Speech & Language, 46, 196-208.

Conference proceedings

- **J. C. Vasquez-Correa**, J. Fritsch, J. R. Orozco-Arroyave, E. Nöth, & M. Magimai-Doss. (2021). On Modeling Glottal Source Information for Phonation Assessment in Parkinson's Disease. INTERSPEECH 2021, 26-30.

- P. A. Perez-Toro, P. A., **J. C. Vasquez-Correa**, T. Arias-Vergara, P. Klumpp, M. Schuster, E. Nöth, & J. R. Orozco-Arroyave, (2021). Emotional State Modeling for the Assessment of Depression in Parkinson's Disease. In International Conference on Text, Speech, and Dialogue (pp. 457-468).

- C. D. Rios-Urrego, **J. C. Vasquez-Correa**, J. R. Orozco-Arroyave, & E. Nöth (2021). Is There Any Additional Information in a Neural Network Trained for Pathological Speech Classification?. In International Conference on Text, Speech, and Dialogue (pp. 435-447).

- G. F. Miller, **J. C. Vásquez-Correa**, & E. Nöth (2020). Assessing the Dysarthria Level of Parkinson's Disease Patients with GMM-UBM Supervectors Using Phonological Posteriors and Diadochokinetic Exercises. In International Conference on Text, Speech, and Dialogue (pp. 356-365).

- C. D. Rios-Urrego, **J. C. Vásquez-Correa**, J. R. Orozco-Arroyave,& E. Nöth. (2020). Transfer Learning to Detect Parkinson's Disease from Speech In Different Languages Using Convolutional Neural Networks with Layer Freezing. In International Conference on Text, Speech, and Dialogue (pp. 331-339).

- **J. C. Vásquez-Correa**, P. Klumpp, J. R. Orozco-Arroyave, & E. Nöth (2019). Phonet: A Tool Based on Gated Recurrent Neural Networks to Extract Phonological Posteriors from Speech. In INTERSPEECH (pp. 549-553).

- P. Klumpp, **J. C. Vásquez-Correa**, T. Haderlein, & E. Nöth (2019). Feature Space Visualization with Spatial Similarity Maps for Pathological Speech Data. In INTERSPEECH (pp. 3068-3072).

- A. Rueda, **J. C. Vásquez-Correa**, C. D. Rios-Urrego, J. R. Orozco-Arroyave, S. Krishnan, & E. Nöth, (2019). Feature Representation of Pathophysiology of Parkinsonian Dysarthria. In INTERSPEECH (pp. 3048-3052).

- **J. C. Vásquez-Correa**, C. D. Rios-Urrego, A. Rueda, J. R. Orozco-Arroyave, S. Krishnan, & E. Nöth (2019). Articulation and Empirical Mode Decomposition Features in Diadochokinetic Exercises for the Speech Assessment of Parkinson's Disease Patients. In Iberoamerican Congress on Pattern Recognition (pp. 688-696).

- **J. C. Vásquez-Correa**, T. Arias-Vergara, T. C. D. Rios-Urrego, M. Schuster, J. Rusz, J. R. Orozco-Arroyave, & E. Nöth (2019). Convolutional Neural Networks and a Transfer Learning Strategy to Classify Parkinson's Disease from Speech in Three Different Languages. In Iberoamerican Congress on Pattern Recognition (pp. 697-706). *BEST PAPER AWARD*

- **J. C. Vásquez-Correa**, T. Arias-Vergara, J. R. Orozco-Arroyave, & E. Nöth (2018). A Multitask Learning Approach to Assess the Dysarthria Severity in Patients with Parkinson's Disease. In INTERSPEECH (pp. 456-460).

- **J. C. Vásquez-Correa**, N. Garcia-Ospina, J. R. Orozco-Arroyave, M. Cernak, & E. Nöth (2018). Phonological Posteriors and GRU Recurrent Units to Assess Speech Impairments of Patients with Parkinson's Disease. In International Conference on Text, Speech, and Dialogue (pp. 453-461).

- N. Garcia-Ospina, T. Arias-Vergara, **J. C. Vásquez-Correa**, J. R. Orozco-Arroyave, M. Cernak, & E. Nöth (2018). Phonological i-Vectors to Detect Parkinson's Disease. In International Conference on Text, Speech, and Dialogue (pp. 462-470).

- L. F. Parra-Gallego, T. Arias-Vergara, **J. C. Vásquez-Correa**, N. Garcia-Ospina, J. R. Orozco-Arroyave, & E. Nöth (2018). Automatic intelligibility assessment of Parkinson's disease with diadochokinetic exercises. In Workshop on Engineering Applications (pp. 223-230).

- **J. C. Vásquez-Correa**, J. R. Orozco-Arroyave, & E. Nöth, E. (2017). Convolutional Neural Network to Model Articulation Impairments in Patients with Parkinson's Disease. In INTERSPEECH (pp. 314-318).

- **J. C. Vásquez-Correa**, J. Serrá, J. R. Orozco-Arroyave, J. F. Vargas-Bonilla, & E. Nöth (2017). Effect of acoustic conditions on algorithms to detect Parkinson's disease from speech. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5065-5069).

- M. Cernak, E. Nöth, F. Rudzicz, H. Christensen, J. R. Orozco-Arroyave, R. Arora, T. Bocklet, H. Chinaei, J. Hannick, P. S. Nidadavolu, **J. C. vasquez-Correa**, M. Yancheva, A. Vann, & N. Vogler (2017). On the impact of non-modal phonation on phonological features. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5090-5094).

- J. C. Jiménez-Monsalve, **J. C. Vasquez-Correa**, J. R. Orozco-Arroyave & P. Gomez-Vilda. (2017). Phonation and articulation analyses in laryngeal pathologies, cleft lip and palate, and Parkinson's disease. In International Work-Conference on the Interplay Between Natural and Artificial Computation (IWINAC) (pp. 424-434).

- **J. C. Vásquez-Correa**, R. Castrillón, T. Arias-Vergara, J. R. Orozco-Arroyave, & E. Nöth. (2017). Speaker Model to Monitor the Neurological State and the Dysarthria Level of Patients with Parkinson's Disease. In International Conference on Text, Speech, and Dialogue (pp. 272-280).

- N. Garcia, N., **J. C Vásquez-Correa**, J. R. Orozco-Arroyave, N. Dehak, & E. Nöth (2017). Language independent assessment of motor impairments of patients with Parkinson's disease using i-vectors. In International Conference on Text, Speech, and Dialogue (pp. 147-155).

Publications related with handwriting assessment of Parkinson's disease

Journal papers

- C. D. Rios-Urrego, **J. C. Vásquez-Correa**, J. F. Vargas-Bonilla, E. Nöth, F. Lopera, & J. R. Orozco-Arroyave (2019). Analysis and evaluation of handwriting in patients with Parkinson's disease using kinematic, geometrical, and non-linear features. Computer Methods and Programs in Biomedicine, 173, 43-52.

Publications related with gait assessment of Parkinson's disease

Journal papers

- P. A. Pérez-Toro, **J. C. Vásquez-Correa**, T. Arias-Vergara, E. Nöth, & J. R. Orozco-Arroyave. Nonlinear dynamics and Poincaré sections to model gait impairments in different stages of Parkinson's disease. Nonlinear Dyn, 100, 3253–3276.

Conference proceedings

- P. A. Pérez-Toro, **J. C. Vásquez-Correa**, T. Arias-Vergara, N. Garcia-Ospina, J. R. Orozco-Arroyave, & E. Nöth (2018). A Non-linear Dynamics Approach to Classify Gait Signals of Patients with Parkinson's Disease. In Workshop on Engineering Applications (pp. 268-278).

Publications related with assessment of Parkinson's disease using smartphones

Journal papers

- J. R. Orozco-Arroyave, **J. C., Vásquez-Correa**, P. Klumpp, P. A. Pérez-Toro, D. Escobar-Grisales, N. Roth, et al. (2020). Apkinson: the smartphone application for telemonitoring Parkinson's patients through speech, gait and hands movement. Neurodegenerative Disease Management, 10(3), 137-157.

Conference proceedings

- **J. C. Vásquez-Correa**, T. Arias-Vergara, et al. (2019). Apkinson: A Mobile Solution for Multimodal Assessment of Patients with Parkinson's Disease. In INTERSPEECH (pp. 964-965).

- T. Arias-Vergara, **J. C. Vasquez-Correa**, J. R. Orozco-Arroyave, P. Klumpp, & E. Nöth (2018). Unobtrusive Monitoring of Speech Impairments of Parkinson'S Disease Patients Through Mobile Devices. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 6004-6008).

- P. Klumpp, T. Janu, T. Arias-Vergara, **J. C. Vásquez-Correa**, J. R. Orozco-Arroyave, & E. Nöth (2017). Apkinson-A Mobile Monitoring Solution for Parkinson's Disease. In INTERSPEECH (pp. 1839-1843).

- T. Arias-Vergara, P. Klumpp, **J. C. Vásquez-Correa**, J. R. Orozco-Arroyave, & E. Nöth (2017). Parkinson's Disease Progression Assessment from Speech Using a Mobile Device-Based Application. In International Conference on Text, Speech, and Dialogue (pp. 371-379).

Publications related with multimodal assessment of Parkinson's disease

Journal papers

- **J. C. Vásquez-Correa**, T. Arias-Vergara, J. R. Orozco-Arroyave, B. Eskofier, J. Klucken, & E. Nöth (2019). Multimodal assessment of Parkinson's disease: a deep learning approach. IEEE Journal of Biomedical and Health Informatics, 23(4), 1618-1630.

Book chapter

- J. R. Orozco-Arroyave, **J. C. Vásquez-Correa**, & E. Nöth. Current methods and new trends in signal processing and pattern recognition for the automatic assessment of motor impairments: the case of Parkinson's disease. Neurological Disorders and Imaging Physics, Volume 5.

Conference proceedings

- **J. C. Vásquez-Correa**, T. Arias-Vergara, P. Klumpp, P. A. Perez-Toro, J. R., Orozco-Arroyave, & E. Nöth (2021). End-2-End Modeling of Speech and Gait from Patients with Parkinson's Disease: Comparison Between High Quality Vs. Smartphone Data. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 7298-7302).

- **J. C. Vásquez-Correa**, T. Bocklet, J. R. Orozco-Arroyave, & E. Nöth, (2020). Comparison of User Models Based on GMM-UBM and I-Vectors for Speech, Handwriting, and Gait Assessment of Parkinson's Disease Patients. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 6544-6548).

- N. García, **J. C. Vásquez-Correa**, J. R. Orozco-Arroyave, & E. Nöth (2018). Multimodal I-vectors to Detect and Evaluate Parkinson's Disease. In INTERSPEECH (pp. 2349-2353).

- **J. C. Vasquez-Correa**, J. R. Orozco-Arroyave, R. Arora, E. Nöth, et al. (2017). Multi-view representation learning via GCCA for multimodal analysis of Parkinson's disease. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 2966-2970).

List of Acronyms

- APQ: Amplitude perturbation quotient
- BBE: Bark band energies
- CD: Correlation dimension
- CNN: Convolutional neural network
- CWT: Continuous wavelet transform
- DDK: Diadochokinetic
- DFA: detrended fluctuation analysis
- DNN: Deep neural network
- DSC: Dynamic score combination
- DWT: Dynamic time warping
- FDA: Frenchay dysarthria assessment
- FI: Freeze index
- FoG: Freezing of gait
- GMM: Gaussian mixture model
- GMM - UBM: Gaussian mixture model - universal background model
- GRU: Gated recurrent unit
- HE: Hurst exponent
- KNN: K-nearest neighbors
- LLE: Largest Lyapunov exponent
- LSTM: Long short-term memory
- LZC: Lempel-Ziv complexity
- m-FDA: modified Frenchay dysarthria assessment

- MAP: Maximum a posteriori
- MDS-UPDRS: Movement disorder society - unified Parkinson's disease rating scale
- MFCC: Mel frequency cepstral coefficients
- MSE: Mean squared error
- NLD: Non-linear dynamics
- PD: Parkinson's disease
- PLLR: Phonological log-likelihood ratio
- PPQ: Pitch perturbation quotient
- RAE: Recurrent autoencoder
- RF: random forest
- SVM: Support vector machine
- SVR: Support vector regression
- TUG: Timed up and go
- VOT: Voiced onset time

List of Figures

List of Tables

Bibliography

[Abad 16] A. Abad *et al.* "Exploiting Phone Log-Likelihood Ratio Features for the Detection of the Native Language of Non-Native English Speakers.". In: *Proceedings of INTERSPEECH*, pp. 2413–2417, 2016.

[Abra 20] A. Abrami, S. Heisig, *et al.* "Using an unbiased symbolic movement representation to characterize Parkinson's disease states". *Scientific Reports*, Vol. 10, No. 1, pp. 1–12, 2020.

[Acke 91] H. Ackermann and W. Ziegler. "Articulatory deficits in parkinsonian dysarthria: an acoustic analysis.". *Journal of Neurology, Neurosurgery & Psychiatry*, Vol. 54, No. 12, pp. 1093–1098, 1991.

[Afon 19] L. Afonso *et al.* "A recurrence plot-based approach for Parkinson's disease identification". *Future Generation Computer Systems*, Vol. 94, pp. 282–292, 2019.

[Agha 17] S. Aghanavesi, D. Nyholm, M. Senek, F. Bergquist, and M. Memedi. "A smartphone-based system to quantify dexterity in Parkinson's disease patients". *Informatics in Medicine Unlocked*, Vol. 9, pp. 11–17, 2017.

[Agha 20] S. Aghanavesi *et al.* "Motion sensor-based assessment of Parkinson's disease motor symptoms during leg agility tests: results from levodopa challenge". *IEEE Journal of Biomedical and Health Informatics*, Vol. 24, pp. 111–119, 2020.

[Ahlr 16] C. Ahlrichs *et al.* "Detecting freezing of gait with a tri-axial accelerometer in Parkinson's disease patients". *Medical & Biological Engineering & Computing*, Vol. 54, No. 1, pp. 223–233, 2016.

[Ahn 13] J. Ahn and N. Hogan. "Long-range correlations in stride intervals may emerge from non-chaotic walking dynamics". *PloS one*, Vol. 8, No. 9, 2013.

[Alha 20] A. S. Alharthi, A. J. Casson, and K. B. Ozanyan. "Gait spatiotemporal signal analysis for parkinson's disease detection and severity rating". *IEEE Sensors Journal*, Vol. 21, No. 2, pp. 1838–1848, 2020.

[Ali 19a] L. Ali *et al.* "Reliable Parkinson's Disease Detection by Analyzing Handwritten Drawings: Construction of an Unbiased Cascaded Learning System Based on Feature Selection and Adaptive Boosting Model". *IEEE Access*, Vol. 7, pp. 116480–116489, 2019.

[Ali 19b] L. Ali, C. Zhu, Z. Zhang, and Y. Liu. "Automated Detection of Parkinson's Disease Based on Multiple Types of Sustained Phonations using Linear Discriminant Analysis and Genetically Optimized Neural Network". *IEEE Journal of Translational Engineering in Health and Medicine*, 2019.

[Alib 16] L. Alibiglou *et al.* "Subliminal gait initiation deficits in rapid eye movement sleep behavior disorder: a harbinger of freezing of gait?". *Movement Disorders*, Vol. 31, No. 11, pp. 1711–1719, 2016.

[Alis 21] M. Alissa *et al.* "Parkinson's disease diagnosis using convolutional neural networks and figure-copying tasks". *Neural Computing and Applications*, pp. 1–21, 2021.

[Alkh 20] R. Alkhatib *et al.* "Machine learning algorithm for gait analysis and classification on early detection of Parkinson". *IEEE Sensors Letters*, Vol. 4, No. 6, pp. 1–4, 2020.

[Amat 21] F. Amato *et al.* "An algorithm for Parkinson's disease speech classification based on isolated words analysis". *Health Information Science and Systems*, Vol. 9, No. 1, pp. 1–15, 2021.

[Ammo 21] A. Ammour *et al.* "Online Arabic and French handwriting of Parkinson's disease: The impact of segmentation techniques on the classification results". *Biomedical Signal Processing and Control*, Vol. 66, p. 102429, 2021.

[Aria 16] T. Arias-Vergara *et al.* "Parkinson's Disease Progression Assessment from Speech Using GMM-UBM.". In: *Proceedings of INTERSPEECH*, pp. 1933–1937, 2016.

[Aria 17] T. Arias-Vergara *et al.* "Parkinson's Disease and Aging: Analysis of Their Effect in Phonation and Articulation of Speech". *Cognitive Computation*, Vol. 9, No. 6, pp. 731–748, 2017.

[Aria 18a] T. Arias-Vergara *et al.* "Speaker models for monitoring Parkinson's disease progression considering different communication channels and acoustic conditions". *Speech Communication*, Vol. 101, pp. 11–25, 2018.

[Aria 18b] T. Arias-Vergara *et al.* "Unobtrusive Monitoring of Speech Impairments of Parkinson's Disease Patients Through Mobile Devices". In: *Proceedings of ICASSP*, pp. 6004–6008, 2018.

[Aria 19] T. Arias-Vergara *et al.* "Phone-attribute posteriors to evaluate the speech of cochlear implant users". In: *Proceedings of INTERSPEECH*, pp. 3108–3112, 2019.

[Arki 18] C. Arkinson and H. Walden. "Parkin function in Parkinson's disease". *Science*, Vol. 360, No. 6386, pp. 267–268, 2018.

[Aror 19] S. Arora, L. Baghai-Ravary, and A. Tsanas. "Developing a large scale population screening tool for the assessment of Parkinson's disease using telephone-quality voice". *The Journal of the Acoustical Society of America*, Vol. 145, No. 5, pp. 2871–2884, 2019.

[Arra 20] D. Arraziqi and Others. "Detection of Parkinson's Disease at The Level of Motor Experiences of Daily Living Using Spiral Handwriting". In: *2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*, pp. 39–46, IEEE, 2020.

[Asae 16] A. Asaei, M. Cernak, and M. Laganaro. "PAoS Markers: Trajectory Analysis of Selective Phonological Posteriors for Assessment of Progressive Apraxia of Speech". In: *Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)*, pp. 50–55, 2016.

[Asho 20] A. S. Ashour *et al.* "Long short term memory based patient-dependent model for FOG detection in Parkinson's disease". *Pattern recognition letters*, Vol. 131, pp. 23–29, 2020.

[Atre 10] P. K. Atrey *et al.* "Multimodal fusion for multimedia analysis: a survey". *Multimedia systems*, Vol. 16, No. 6, pp. 345–379, 2010.

[Bach 10] M. Bachlin *et al.* "Wearable assistant for Parkinson's disease patients with the freezing of gait symptom". *IEEE Transactions on Information Technology in Biomedicine*, Vol. 14, No. 2, pp. 436–446, 2010.

[Bach 89] M. Bacher, E. Scholz, and H. C. Diener. "24 hour continuous tremor quantification based on EMG recording". *Electroencephalography and clinical neurophysiology*, Vol. 72, No. 2, pp. 176–183, 1989.

[Bagu 17] A. R. Bagudanch. "Variation and phonological change: The case of yeísmo in Spanish". *Folia Linguistica*, Vol. 51, No. 1, pp. 169–206, 2017.

[Bake 98] K. K. Baker, L. O. Ramig, E. S. Luschei, and M. E. Smith. "Thyroarytenoid muscle activity associated with hypophonia in Parkinson's disease and aging". *Neurology*, Vol. 51, No. 6, pp. 1592–1598, 1998.

[Bala 20] E. Balaji, D. Brindha, and R. Balakrishnan. "Supervised machine learning based gait classification system for early detection and stage classification of Parkinson's disease". *Applied Soft Computing*, Vol. 94, p. 106494, 2020.

[Bala 21a] E. Balaji, D. Brindha, V. K. Elumalai, and K. Umesh. "Data-driven gait analysis for diagnosis and severity rating of Parkinson's disease". *Medical Engineering & Physics*, Vol. 91, pp. 54–64, 2021.

[Bala 21b] E. Balaji *et al.* "Automatic and non-invasive Parkinson's disease diagnosis and severity rating using LSTM network". *Applied Soft Computing*, Vol. 108, p. 107463, 2021.

[Barn 16] M. S. Barnish *et al.* "Roles of cognitive status and intelligibility in everyday communication in people with Parkinson's disease: A systematic review". *Journal of Parkinson's disease*, Vol. 6, No. 3, pp. 453–462, 2016.

[Bart 11] J. Barth *et al.* "Biometric and mobile gait analysis for early diagnosis and therapy monitoring in Parkinson's disease". In: *Proceedings of EMBC*, pp. 868–871, 2011.

[Bart 12a] J. Barth *et al.* "Combined analysis of sensor data from hand and gait motor function improves automatic recognition of Parkinson's disease". In: *Proceedings of EMBC*, pp. 5122–5125, 2012.

[Bart 12b] J. Barth *et al.* "Combined analysis of sensor data from hand and gait motor function improves automatic recognition of Parkinson's disease". In: *Proceedings of EMBC*, pp. 5122–5125, 2012.

[Bart 17] J. Barth. *Development and Validation of a Mobile Gait Analysis System Providing Clinically Relevant Target Parameters in Parkinson's Disease*. Logos-Verlag, Berlin, Germany, 1st Ed., 2017.

[Baub 00] C. E. Bauby and A. D. Kuo. "Active control of lateral balance in human walking". *Journal of biomechanics*, Vol. 33, No. 11, pp. 1433–1440, 2000.

[Baye 13] A. Bayestehtashk, M. Asgari, I. Shafran, and J. McNames. "Fully automated assessment of the severity of Parkinson's disease from speech". *Computer speech & language*, Vol. 29, No. 1, pp. 172–185, 2013.

[Bela 16] E. A. Belalcázar-Bolaños, J. R. Orozco-Arroyave, *et al.* "Glottal Flow Patterns Analyses for Parkinson's Disease Detection: Acoustic and Nonlinear Approaches". In: *International Conference on Text, Speech, and Dialogue*, pp. 400–407, 2016.

[Benb 19] A. Benba, A. Jilbab, S. Sandabad, and A. Hammouch. "Voice signal processing for detecting possible early signs of Parkinson's disease in patients with rapid eye movement sleep behavior disorder". *International Journal of Speech Technology*, Vol. 22, No. 1, pp. 121–129, 2019.

[Berg 12] J. Bergstra and Y. Bengio. "Random search for hyper-parameter optimization". *The Journal of Machine Learning Research*, Vol. 13, No. 1, pp. 281–305, 2012.

[Berg 18] D. Berg and R. B. Postuma. "From prodromal to overt Parkinson's disease: towards a new definition in the year 2040". *Journal of Parkinson's Disease*, Vol. 8, No. s1, pp. S19–S23, 2018.

[Beri 17] V. Berisha *et al.* "Float Like a Butterfly Sting Like a Bee: Changes in Speech Preceded Parkinsonism Diagnosis for Muhammad Ali". In: *Proceedings of INTERSPEECH*, pp. 1809–1813, 2017.

[Beru 19] L. Berus, S. Klancnik, M. Brezocnik, and M. Ficko. "Classifying Parkinson's Disease Based on Acoustic Measures Using Artificial Neural Networks". *Sensors*, Vol. 19, No. 1, p. 16, 2019.

[Bjor 18] N. Bjorck *et al.* "Understanding batch normalization". In: *Advances in Neural Information Processing Systems*, pp. 7694–7705, 2018.

[Blan 09] P. G. Blanchet and G. J. Snyder. "Speech Rate Deficits in Individuals with Parkinson's disease: a review of the literature". *Journal of Medical Speech-Language Pathology*, Vol. 17, No. 1, pp. 1–7, 2009.

[Bock 13] T. Bocklet, S. Steidl, E. Nöth, and S. Skodda. "Automatic Evaluation of Parkinson's Speech – Acoustic, Prosodic and Voice Related Cues". In: *Proceedings of INTERSPEECH*, pp. 1149–1153, 2013.

[Boro 80] J. C. Borod, H. Goodglass, and E. Kaplan. "Normative data on the Boston diagnostic aphasia examination, parietal lobe battery, and the Boston naming test". *Journal of Clinical and Experimental Neuropsychology*, Vol. 2, No. 3, pp. 209–215, 1980.

[Borz 20] L. Borzì, M. Varrecchia, *et al.* "Smartphone-based estimation of item 3.8 of the MDS-UPDRS-III for assessing leg agility in people with Parkinson's disease". *IEEE Open Journal of Engineering in Medicine and Biology*, 2020.

[Borz 21] L. Borzì *et al.* "Prediction of freezing of gait in Parkinson's disease using wearables and machine learning". *Sensors*, Vol. 21, No. 2, p. 614, 2021.

[Bose 92] B. E. Boser, I. M. Guyon, and V. N. Vapnik. "A training algorithm for optimal margin classifiers". In: *Fifth annual workshop on Computational learning theory*, pp. 144–152, 1992.

[Bot 16a] B. M. Bot *et al.* "The mPower study, Parkinson disease mobile data collected using ResearchKit". *Scientific data*, Vol. 3, p. 160011, 2016.

[Bot 16b] B. M. Bot *et al.* "The mPower study, Parkinson disease mobile data collected using Research-Kit". *Nature Scientific Data*, Vol. 3, No. 160011, pp. 1–24, 2016.

[Brog 19] L. Brognara *et al.* "Assessing gait in Parkinson's disease using wearable motion sensors: a systematic review". *Diseases*, Vol. 7, No. 1, p. 18, 2019.

[Buzz 03] U. H. Buzzi, N. Stergiou, *et al.* "Nonlinear dynamics indicates aging affects variability during gait". *Clinical Biomechanics*, Vol. 18, No. 5, pp. 435–443, 2003.

[Camp 05] S. von Campenhausen, B. Bornschein, *et al.* "Prevalence and incidence of Parkinson's disease in Europe". *European Neuropsychopharmacology*, Vol. 15, No. 4, pp. 473–490, 2005.

[Camp 18] J. Camps *et al.* "Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit". *Knowledge-Based Systems*, Vol. 139, pp. 119–131, 2018.

[Canh 00] R. O. Canham, S. L. Smith, and A. M. Tyrrell. "Automated scoring of a neuropsychological test: the Rey Osterrieth complex figure". In: *Proceedings of the Euromicro Conference.*, pp. 406–413, 2000.

[Cant 20] I. Canturk. "Fuzzy recurrence plot-based analysis of dynamic and static spiral tests of Parkinson's disease patients". *Neural Computing and Applications*, 2020.

[Cant 21a] İ. Cantürk. "A computerized method to assess Parkinson's disease severity from gait variability based on gender". *Biomedical Signal Processing and Control*, Vol. 66, p. 102497, 2021.

[Cant 21b] İ. Cantürk. "Fuzzy recurrence plot-based analysis of dynamic and static spiral tests of Parkinson's disease patients". *Neural Computing and Applications*, Vol. 33, pp. 349–360, 2021.

[Cape 16] M. Capecci, L. Pepa, F. Verdini, and M. G. Ceravolo. "A smartphone-based architecture to detect and quantify freezing of gait in Parkinson's disease". *Gait & posture*, Vol. 50, pp. 28–33, 2016.

[Cara 18] C. Caramia *et al.* "IMU-Based Classification of Parkinson's Disease From Gait: A Sensitivity Analysis on Sensor Location and Feature Selection". *IEEE Journal of Biomedical and Health Informatics*, Vol. 22, No. 6, pp. 1765–1774, 2018.

[Caru 94] R. Caruana. "Multitask Connectionist Learning". In: *Proceedings of the Connectionist Models Summer School*, pp. 372–379, 1994.

[Cast 14] M. Castelli, L. Vanneschi, and S. Silva. "Prediction of the unified Parkinson's disease rating scale assessment using a genetic programming system with geometric semantic genetic operators". *Expert Systems with Applications*, Vol. 41, No. 10, pp. 4608–4616, 2014.

[Cast 16] R. M. Castilhos *et al.* "Genetic aspects of Huntington's disease in Latin America. A systematic review". *Clinical genetics*, Vol. 89, No. 3, pp. 295–303, 2016.

[Cast 19] R. Castrillon *et al.* "Characterization of the Handwriting Skills as a Biomarker for Parkinson's Disease". In: *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, pp. 1–5, 2019.

[Cern 16] M. Cernak and P. N. Garner. "PhonVoc: A Phonetic and Phonological Vocoding Toolkit". In: *Proceedings of INTERSPEECH*, pp. 988–992, 2016.

[Cern 17] M. Cernak *et al.* "Characterisation of voice quality of Parkinson's disease using differential phonological posterior features". *Computer Speech & Language*, Vol. 46, pp. 196–208, 2017.

[Chan 11] C. C. Chang and C. J. Lin. "LIBSVM: A library for support vector machines". *ACM Transactions on Intelligent Systems and Technology*, Vol. 2, pp. 27:1–27:27, 2011.

[Chan 16] H. C. Chang, Y. L. Hsu, S. C. Yang, J. C. Lin, and Z. H. Wu. "A wearable inertial measurement system with complementary filter for gait analysis of patients with stroke or Parkinson's disease". *IEEE Access*, Vol. 4, pp. 8442–8453, 2016.

[Chen 11] K. Chenausky, J. MacAuslan, and R. Goldhor. "Acoustic analysis of PD speech". *Parkinson's Disease*, Vol. 2011, 2011.

[Cho 14] K. Cho *et al.* "Learning phrase representations using RNN encoder-decoder for statistical machine translation". In: *Proceedings of EMNLP*, 2014.

[Chom 19] T. Chomiak *et al.* "A novel single-sensor-based method for the detection of gait-cycle breakdown and freezing of gait in Parkinson's disease". *Journal of Neural Transmission*, pp. 1–8, 2019.

[Chor 19] J. Chorowski, R. J. Weiss, S. Bengio, and A. van den Oord. "Unsupervised speech representation learning using wavenet autoencoders". *IEEE/ACM transactions on audio, speech, and language processing*, Vol. 27, No. 12, pp. 2041–2053, 2019.

[Clee 87] L. Cleeves, L. J. Findley, and M. Gresty. "Assessment of rest tremor in Parkinson's disease.". *Advances in neurology*, Vol. 45, pp. 349–352, 1987.

[Coen 13] P. M. Coen, S. A. Jubrias, *et al.* "Skeletal muscle mitochondrial energetics are associated with maximal aerobic capacity and walking speed in older adults". *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, Vol. 68, No. 4, pp. 447–455, 2013.

[Corr 18] J. Correia, B. Raj, I. Trancoso, and F. Teixeira. "Mining Multimodal Repositories for Speech Affecting Diseases". In: *Proceedings of INTERSPEECH*, pp. 2963–2967, 2018.

[Corr 19] J. Correia, I. Trancoso, and B. Raj. "In-the-Wild End-to-End Detection of Speech Affecting Diseases". In: *IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 734–741, 2019.

[Cumm 18] N. Cummins, A. Baird, and B. Schuller. "Speech analysis for health: Current state-of-the-art and the increasing impact of deep learning". *Methods*, 2018.

[Cuzz 17] F. Cuzzolin, M. Sapienza, *et al.* "Metric learning for Parkinsonian identification from IMU gait measurements". *Gait & posture*, Vol. 54, pp. 127–132, 2017.

[Damo 10] S. Damouras, M. D. Chang, *et al.* "An empirical examination of detrended fluctuation analysis for gait data". *Gait & posture*, Vol. 31, No. 3, pp. 336–340, 2010.

[Dann 19] J. Danna *et al.* "Digitalized spiral drawing in Parkinson's disease: A tool for evaluating beyond the written trace". *Human movement science*, Vol. 65, pp. 80–88, 2019.

[Das 12] S. Das *et al.* "Detecting Parkinson's symptoms in uncontrolled home environments: A multiple instance learning approach". In: *proceedings of EMBC*, pp. 3688–3691, 2012.

[De S 19] C. De Stefano, F. Fontanella, D. Impedovo, G. Pirlo, and A. S. di Freca. "Handwriting analysis to support neurodegenerative diseases diagnosis: A review". *Pattern Recognition Letters*, Vol. 121, pp. 37–45, 2019.

[Deha 11] N. Dehak *et al.* "Front–end factor analysis for speaker verification". *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 19, No. 4, pp. 788–798, 2011.

[Demo 15] M. Demonceau *et al.* "Contribution of a trunk accelerometer system to the characterization of gait in patients with mild-to-moderate Parkinson's disease". *IEEE journal of Biomedical and Health Informatics*, Vol. 19, No. 6, pp. 1803–1808, 2015.

[Deng 10] L. Deng *et al.* "Binary coding of speech spectrograms using a deep auto-encoder". In: *Proceedings of INTERSPEECH*, 2010.

[Dent 21] V. Dentamaro *et al.* "Benchmarking of Shallow Learning and Deep Learning Techniques with Transfer Learning for Neurodegenerative Disease Assessment Through Handwriting". In: *International Conference on Document Analysis and Recognition*, pp. 7–20, Springer, 2021.

[Desr 95] J. Desrosiers, R. Hebert, G. Bravo, and E. Dutil. "The Purdue Pegboard Test: normative data for people aged 60 and over". *Disability and rehabilitation*, Vol. 17, No. 5, pp. 217–224, 1995.

[Deus 96] G. Deuschl *et al.* "Clinical neurophysiology of tremor". *Journal of clinical neurophysiology*, Vol. 13, No. 2, pp. 110–121, 1996.

[Diaz 19] M. Diaz, M. A. Ferrer, D. Impedovo, G. Pirlo, and G. Vessio. "Dynamically enhanced static handwriting representation for Parkinson's disease detection". *Pattern Recognition Letters*, Vol. 128, pp. 204–210, 2019.

[Diaz 21] M. Diaz *et al.* "Sequence-based dynamic handwriting analysis for Parkinson's disease detection with one-dimensional convolutions and BiGRUs". *Expert Systems with Applications*, Vol. 168, p. 114405, 2021.

[Diel 14] S. Dieleman and B. Schrauwen. "End-to-end learning for music audio". In: *Proceedings of ICASSP*, pp. 6964–6968, 2014.

[Dier 17] F. Dierick, A. L. Nivard, *et al.* "Fractal analyses reveal independent complexity and predictability of gait". *PloS one*, Vol. 12, No. 11, 2017.

[Diez 14a] M. Diez *et al.* "New insight into the use of phone log-likelihood ratios as features for language recognition". In: *Proceedings of INTERSPEECH*, 2014.

[Diez 14b] M. Diez *et al.* "On the projection of PLLRs for unbounded feature distributions in spoken language recognition". *IEEE Signal Processing Letters*, Vol. 21, No. 9, pp. 1073–1077, 2014.

[Dima 17] G. Dimauro, V. Di-Nicola, *et al.* "Assessment of Speech Intelligibility in Parkinson's Disease Using a Speech-To-Text System". *IEEE Access*, Vol. 5, pp. 22199–22208, 2017.

[Djur 17] M. Djurić-Jovičić *et al.* "Selection of gait parameters for differential diagnostics of patients with de novo Parkinson's disease". *Neurological research*, Vol. 39, No. 10, pp. 853–861, 2017.

[Dors 18a] E. Dorsey *et al.* "The emerging evidence of the Parkinson pandemic". *Journal of Parkinson's disease*, Vol. 8, No. s1, pp. S3–S8, 2018.

[Dors 18b] E. R. Dorsey, A. Elbaz, *et al.* "Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016". *The Lancet Neurology*, Vol. 17, No. 11, pp. 939–953, 2018.

[Drot 14] P. Drotár *et al.* "Analysis of in-air movement in handwriting: A novel marker for Parkinson's disease". *Computer Methods and Programs in Biomedicine*, Vol. 117, No. 3, pp. 405–411, 2014.

[Drot 16] P. Drotár *et al.* "Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease". *Artificial intelligence in Medicine*, Vol. 67, pp. 39–46, 2016.

[Drum 93] S. S. Drummond. *Dysarthria examination battery.* Communication Skill Builders, 1993.

[Duff 13] J. R. Duffy. *Motor Speech disorders-E-Book: Substrates, differential diagnosis, and management.* Elsevier Health Sciences, 2013.

[Duma 09] B. Dumas, D. Lalanne, and S. Oviatt. "Multimodal interfaces: A survey of principles, models and frameworks". In: *Human machine interaction*, pp. 3–26, Springer, 2009.

[Eite 94] T. Eiter and H. Mannila. "Computing discrete Fréchet distance". Tech. Rep., Citeseer, 1994.

[El M 20] I. El Maachi, G. A. Bilodeau, and W. Bouachir. "Deep 1D-Convnet for accurate Parkinson disease detection and severity prediction from gait". *Expert Systems with Applications*, Vol. 143, p. 113075, 2020.

[Ende 08] P. M. Enderby and R. Palmer. *FDA-2: Frenchay Dysarthria Assessment: Examiner's Manual.* Pro-ed, 2008.

[Ertu 16] Ö. F. Ertuğrul *et al.* "Detection of Parkinson's disease by shifted one dimensional local binary patterns from gait". *Expert Systems with Applications*, Vol. 56, pp. 156–163, 2016.

[Eski 12] Ö. Eskidere, F. Ertaş, and C. Hanilçi. "A comparison of regression methods for remote tracking of Parkinson's disease progression". *Expert Systems with Applications*, Vol. 39, No. 5, pp. 5523–5528, 2012.

[Eybe 15] F. Eyben and B. Schuller. "openSMILE:): The Munich open-source large-scale multimedia feature extractor". *ACM SIGMultimedia Records*, Vol. 6, No. 4, pp. 4–13, 2015.

[Fan 18] C. Fan *et al.* "Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data". *Applied energy*, Vol. 211, pp. 1123–1135, 2018.

[Feig 17] V. L. Feigin, A. A. Abajobir, *et al.* "GBD 2015 Neurological Disorders Collaborator Group. Global, regional, and national burden of neurological disorders during 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015". *Lancet Neurol.*, Vol. 16, No. 11, pp. 877–97, 2017.

[Feig 19] V. L. Feigin *et al.* "Global, regional, and national burden of neurological disorders, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016". *The Lancet Neurology*, Vol. 18, No. 5, pp. 459–480, 2019.

[Fere 19] S. M. Fereshtehnejad *et al.* "Evolution of prodromal Parkinson's disease and dementia with Lewy bodies: a prospective study". *Brain*, Vol. 142, No. 7, pp. 2051–2067, 2019.

[Fitt 54] P. M. Fitts. "The information capacity of the human motor system in controlling the amplitude of movement.". *Journal of experimental psychology*, Vol. 47, No. 6, p. 381, 1954.

[Forr 89] K. Forrest, G. Weismer, and G. S. Turner. "Kinematic, acoustic, and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric adults". *The Journal of the Acoustical Society of America*, Vol. 85, No. 6, pp. 2608–2622, 1989.

[Frai 16] L. Fraiwan, R. Khnouf, and A. R. Mashagbeh. "Parkinson's disease hand tremor detection system for mobile application". *Journal of medical engineering & technology*, Vol. 40, No. 3, pp. 127–134, 2016.

[Gala 16] Z. Galaz, J. Mekyska, *et al.* "Prosodic analysis of neutral, stress-modified and rhymed speech in patients with Parkinson's disease". *Computer Methods and Programs in Biomedicine*, Vol. 127, pp. 301–317, 2016.

[Gala 18] Z. Galaz *et al.* "Changes in Phonation and Their Relations with Progress of Parkinson's Disease". *Applied Sciences*, Vol. 8, No. 12, p. 2339, 2018.

[Gall 18] C. Gallicchio, A. Micheli, and L. Pedrelli. "Deep Echo State Networks for Diagnosis of Parkinson's Disease". In: *26th European Symposium on Artificial Neural Networks*, pp. 397–402, 2018.

[Galn 15] B. Galna, S. Lord, D. J. Burn, and L. Rochester. "Progression of gait dysfunction in incident Parkinson's disease: impact of medication and phenotype". *Movement Disorders*, Vol. 30, No. 3, pp. 359–367, 2015.

[Gan 10] Z. Gan-Or *et al.* "LRRK2 and GBA mutations differentially affect the initial presentation of Parkinson disease". *Neurogenetics*, Vol. 11, No. 1, pp. 121–125, 2010.

[Garc 14] N. García *et al.* "Evaluation of the effects of speech enhancement algorithms on the detection of fundamental frequency of speech". In: *Proceedings of STSIVA*, pp. 1–5, 2014.

[Garc 16] A. M. García *et al.* "How language flows when movements don't: an automated analysis of spontaneous discourse in Parkinson's disease". *Brain and language*, Vol. 162, pp. 19–28, 2016.

[Garc 17] N. Garcia, J. R. Orozco-Arroyave, L. F. D'Haro, N. Dehak, and E. Nöth. "Evaluation of the neurological state of people with Parkinson's disease using i-vectors". In: *Proceedings of INTERSPEECH*, pp. 299–303, 2017.

[Garc 18a] A. M. García, Y. Bocanegra, *et al.* "Parkinson's disease compromises the appraisal of action meanings evoked by naturalistic texts". *Cortex*, Vol. 100, pp. 111–126, 2018.

[Garc 18b] N. García *et al.* "Multimodal I-vectors to Detect and Evaluate Parkinson's Disease.". In: *Proceedings of INTERSPEECH*, pp. 2349–2353, 2018.

[Garc 21] A. M. García *et al.* "Cognitive Determinants of Dysarthria in Parkinson's Disease: An Automated Machine Learning Approach". *Movement Disorders*, 2021.

[Gate 07] D. H. Gates, J. L. Su, and J. B. Dingwell. "Possible biomechanical origins of the long-range correlations in stride intervals of walking". *Physica A: Statistical Mechanics and its Applications*, Vol. 380, pp. 259–270, 2007.

[Gazd 21] M. Gazda, M. Hireš, and P. Drotár. "Multiple-Fine-Tuned Convolutional Neural Networks for Parkinson's Disease Diagnosis From Offline Handwriting". *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021.

[Gil 19] M. Gil-Martín, J. M. Montero, and R. San-Segundo. "Parkinson's Disease Detection from Drawing Movements Using Convolutional Neural Networks". *Electronics*, Vol. 8, No. 8, p. 907, 2019.

[Gilk 05] W. P. Gilks *et al.* "A common LRRK2 mutation in idiopathic Parkinson's disease". *The Lancet*, Vol. 365, No. 9457, pp. 415–416, 2005.

[Girs 15] R. Girshick. "Fast R-CNN". In: *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, 2015.

[Godi 17] J. I. Godino-Llorente *et al.* "Towards the identification of Idiopathic Parkinson's Disease from the speech. New articulatory kinetic biomarkers". *PloS one*, Vol. 12, No. 12, p. e0189583, 2017.

[Goet 08] C. Goetz *et al.* "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results". *Movement Disorders*, Vol. 23, No. 15, pp. 2129–2170, 2008.

[Gold 00] A. L. Goldberger *et al.* "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals". *circulation*, Vol. 101, No. 23, pp. e215–e220, 2000.

[Gold 14] S. M. Goldman. "Environmental toxins and Parkinson's disease". *Annual review of pharmacology and toxicology*, Vol. 54, pp. 141–164, 2014.

[Gome 21] L. F. Gomez *et al.* "Improving Parkinson Detection Using Dynamic Features From Evoked Expressions in Video". In: *Proceedings of CVPR*, pp. 1562–1570, 2021.

[Good 16] I. Goodfellow, Y. Bengio, and A. Courville. *"Deep Learning"*. MIT Press, 2016.

[Grac 14] R. Graça, R. S. Castro, and J. Cevada. "Parkdetect: Early diagnosing parkinson's disease". In: *Proceedings of MeMeA*, pp. 1–6, 2014.

[Gras 04] P. Grassberger and I. Procaccia. "Measuring the strangeness of strange attractors". In: *The Theory of Chaotic Attractors*, pp. 170–189, Springer, 2004.

[Grav 12] A. Graves. "Supervised sequence labelling". In: *Supervised sequence labelling with recurrent neural networks*, pp. 5–13, Springer, 2012.

[Gros 15] T. Grósz *et al.* "Assessing the Degree of Nativeness and Parkinson's Condition Using Gaussian Processes and Deep Rectifier Neural Networks". In: *Proceedings of INTERSPEECH*, pp. 1339–1343, 2015.

[Gupt 20] U. Gupta, H. Bansal, and D. Joshi. "An improved sex-specific and age-dependent classification model for Parkinson's diagnosis using handwriting measurement". *Computer methods and programs in biomedicine*, Vol. 189, p. 105305, 2020.

[Haki 16] D. Hakim. "This pesticide is prohibited In Britain. Why is it still being exported?". *The New York Times, December*, Vol. 20, p. 2016, 2016.

[Hann 17] J. Hannink, H. Gaßner, *et al.* "Inertial sensor-based estimation of peak accelerations during heel-strike and loading as markers of impaired gait patterns in PD patients". *Basal Ganglia*, Vol. 8, p. 1, 2017.

[Hass 12a] C. J. Hass, P. Malczak, *et al.* "Quantitative Normative Gait Data in a Large Cohort of Ambulatory Persons with Parkinson's Disease". *PLOS ONE*, Vol. 7, No. 8, pp. 1–5, 2012.

[Hass 12b] C. J. Hass, P. Malczak, *et al.* "Quantitative normative gait data in a large cohort of ambulatory persons with Parkinson's disease". *PloS one*, Vol. 7, No. 8, 2012.

[Hast 09] T. Hastie *et al.* "Multi-class adaboost". *Statistics and its Interface*, Vol. 2, No. 3, pp. 349–360, 2009.

[Haus 00] J. M. Hausdorff, A. Lertratanakul, *et al.* "Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis". *Journal of applied physiology*, Vol. 88, No. 6, pp. 2045–2053, 2000.

[Haus 96] J. M. Hausdorff, P. L. Purdon, *et al.* "Fractal Dynamics of Human Gait: Stability of Long-Range Correlations in Stride Interval Fluctuations". *Journal of applied physiology*, Vol. 80, No. 5, pp. 1448–1457, 1996.

[Haus 97] J. M. Hausdorff, S. L. Mitchell, *et al.* "Altered fractal dynamics of gait: reduced stride-interval correlations with aging and Huntington's disease". *Journal of applied physiology*, Vol. 82, No. 1, pp. 262–269, 1997.

[He 16] K. He, X. Zhang, *et al.* "Deep residual learning for image recognition". In: *Proceedings of CVPR*, pp. 770–778, 2016.

[Hemm 16] D. Hemmerling *et al.* "Automatic Detection of Parkinson's Disease Based on Modulated Vowels.". In: *Proceedings of INTERSPEECH*, pp. 1190–1194, 2016.

[Hemm 20] D. Hemmerling and M. Wojcik-Pedziwiatr. "Prediction and Estimation of Parkinson's Disease Severity Based on Voice Signal". *Journal of Voice*, 2020.

[Herm 05] T. Herman *et al.* "Gait instability and fractal dynamics of older adults with a "cautious" gait: why do certain older adults walk fearfully?". *Gait & posture*, Vol. 21, No. 2, pp. 178–185, 2005.

[Hern 14] C. D. Hernández-Mena and J. A. Herrera-Camacho. "CIEMPIESS: A new open-sourced Mexican Spanish radio corpus". In: *Proceedings of LREC*, pp. 371–375, 2014.

[Hier 93] J. L. Hieronymus. "ASCII phonetic symbols for the world's languages: Worldbet". *Journal of the International Phonetic Association*, Vol. 23, p. 72, 1993.

[Hlav 17] J. Hlavnicka, R. Cmejla, T. Tykalova, K. Sonka, E. Ruzicka, and J. Rusz. "Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behaviour disorder". *Scientific reports*, Vol. 7, No. 1, p. 12, 2017.

[Ho 99] A. K. Ho *et al.* "Speech impairment in a large sample of patients with Parkinson's disease". *Behavioural neurology*, Vol. 11, No. 3, pp. 131–137, 1999.

[Hoch 97] S. Hochreiter and J. Schmidhuber. "Long short-term memory". *Neural computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.

[Horn 98] O. Hornykiewicz. "Biochemical aspects of Parkinson's disease". *Neurology*, Vol. 51, No. 2, pp. S2–S9, 1998.

[Hssa 19] M. D. Hssayeni, J. Jimenez-Shahed, and B. Ghoraani. "Hybrid Feature Extraction for Detection of Degree of Motor Fluctuation Severity in Parkinson's Disease Patients". *Entropy*, Vol. 21, No. 2, p. 137, 2019.

[Hurs 65] H. E. Hurst. *Long term storage: An experimental study.* Constable, London, 1965.

[Iako 19] D. Iakovakis *et al.* "Early Parkinson's Disease Detection via Touchscreen Typing Analysis using Convolutional Neural Networks". In: *Proceedings of EMBC*, pp. 3535–3538, 2019.

[Iand 16] F. N. Iandola *et al.* "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size". *arXiv preprint arXiv:1602.07360*, 2016.

[Impe 19a] D. Impedovo. "Velocity-based signal features for the assessment of Parkinsonian handwriting". *IEEE Signal Processing Letters*, Vol. 26, No. 4, pp. 632–636, 2019.

[Impe 19b] D. Impedovo and G. Pirlo. "Dynamic handwriting analysis for the assessment of neurodegenerative diseases: A pattern recognition perspective". *IEEE reviews in biomedical engineering*, Vol. 12, pp. 209–220, 2019.

[Impe 19c] D. Impedovo, G. Pirlo, G. Vessio, and M. T. Angelillo. "A Handwriting-Based Protocol for Assessing Neurodegenerative Dementia". *Cognitive Computation*, pp. 1–11, 2019.

[Impe 21] D. Impedovo and Others. "Investigating the Sigma-Lognormal Model for Disease Classification by Handwriting". In: *THE LOGNORMALITY PRINCIPLE AND ITS APPLICATIONS IN E-SECURITY, E-LEARNING AND E-HEALTH*, pp. 195–209, World Scientific, 2021.

[Ioff 15] S. Ioffe and C. Szegedy. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". In: *Proceedings of ICML*, pp. 448–456, 2015.

[Iran 18] A. Iranzo. "Dissecting premotor Parkinson's disease with multimodality neuroimaging". *The Lancet Neurology*, Vol. 17, No. July, pp. 574–576, 2018.

[Isen 14] M. E. Isenkul, B. E. Sakar, and O. Kursun. "Improved spiral test using digitized graphics tablet for monitoring Parkinson's disease". In: *International Conference on e-Health and Telemedicine*, pp. 171–175, 2014.

[Jean 20] L. Jeancolas *et al.* "X-vectors: New Quantitative Biomarkers for Early Parkinson's Disease Detection from Speech". *arXiv preprint arXiv:2007.03599*, 2020.

[Jean 21] L. Jeancolas *et al.* "X-Vectors: new quantitative biomarkers for early Parkinson's disease detection from speech". *Frontiers in Neuroinformatics*, Vol. 15, p. 4, 2021.

[Jerk 18] V. Jerkovic *et al.* "Analysis of on-surface and in-air movement in handwriting of subjects with Parkinson's disease and atypical parkinsonism". *Biomedical Engineering/Biomedizinische Technik*, Vol. 64, No. 2, pp. 1–8, 2018.

[Jian 08] Z. Jian-Jun, N. Xin-Bao, *et al.* "Decrease in Hurst exponent of human gait with aging and neurodegenerative diseases". *Chinese Physics B.*, Vol. 17, No. 3, p. 852, 2008.

[Jiao 17] Y. Jiao, V. Berisha, and J. Liss. "Interpretable phonological features for clinical applications". In: *Proceedings of ICASSP*, pp. 5045–5049, 2017.

[John 13] S. J. Johnson *et al.* "An economic model of Parkinson's disease: Implications for slowing progression in the United States". *Movement Disorders*, Vol. 28, No. 3, pp. 319–326, 2013.

[Jord 07] K. Jordan, J. H. Challis, and K. M. Newell. "Walking speed influences on gait cycle variability". *Gait & posture*, Vol. 26, No. 1, pp. 128–134, 2007.

[Kama 16] C. Kamath. "Analysis of altered complexity of gait dynamics with aging and Parkinson's disease using ternary Lempel-Ziv complexity". *Cogent engineering*, Vol. 3, No. 1, p. 1177924, 2016.

[Kara 19] B. Karan, S. S. Sahu, and K. Mahto. "Parkinson disease prediction using intrinsic mode function based features from speech signal". *Biocybernetics and Biomedical Engineering*, 2019.

[Kara 20] B. Karan *et al.* "Hilbert spectrum analysis for automatic detection and evaluation of Parkinson's speech". *Biomedical Signal Processing and Control*, Vol. 61, p. 102050, 2020.

[Kara 21a] O. Karaman *et al.* "Robust automated Parkinson disease detection based on voice signals with transfer learning". *Expert Systems with Applications*, Vol. 178, p. 115013, 2021.

[Kara 21b] N. Karan *et al.* "Non-negative matrix factorization-based time-frequency feature extraction of voice signal for Parkinson's disease prediction". *Computer Speech & Language*, Vol. 69, p. 101216, 2021.

[Kasp 87] F. Kaspar and H. G. Schuster. "Easily calculable measure for the complexity of spatiotemporal patterns". *Physical review A*, Vol. 36, No. 2, pp. 842–848, 1987.

[Kell 12] V. E. Kelly, A. J. Eusterbrock, and A. Shumway-Cook. "A review of dual-task walking deficits in people with Parkinson's disease: motor and cognitive contributions, mechanisms, and clinical implications". *Parkinson's Disease*, Vol. 2012, 2012.

[Kenn 92] M. B. Kennel, R. Brown, and H. D. Abarbanel. "Determining embedding dimension for phase-space reconstruction using a geometrical construction". *Physical review A*, Vol. 45, No. 6, pp. 3403–3411, 1992.

[Kher 21] P. Khera and N. Kumar. "Age-gender specific prediction model for Parkinson's severity assessment using gait biomarkers". *Engineering Science and Technology, an International Journal*, 2021.

[Kim 09] Y. Kim, G. Weismer, R. D. Kent, and J. R. Duffy. "Statistical models of F2 slope in relation to severity of dysarthria". *Folia Phoniatrica et Logopaedica*, Vol. 61, No. 6, pp. 329–335, 2009.

[Kim 15] H. Kim *et al.* "Unconstrained detection of freezing of Gait in Parkinson's disease patients using smartphone". In: *Proceedings of EMBC*, pp. 3751–3754, 2015.

[King 14] D. P. Kingma and J. Ba. "Adam: A method for stochastic optimization". *arXiv preprint arXiv:1412.6980*, 2014.

[Kisl 17] T. Kisler, U. Reichel, and F. Schiel. "Multilingual processing of speech via web services". *Computer Speech & Language*, Vol. 45, pp. 326–347, 2017.

[Kluc 13] J. Klucken, J. Barth, *et al.* "Unbiased and mobile gait analysis detects motor impairment in Parkinson's disease". *PloS one*, Vol. 8, No. 2, p. e56956, 2013.

[Kodr 20] I. Kodrasi and H. Bourlard. "Spectro-Temporal Sparsity Characterization for Dysarthric Speech Detection". *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 28, pp. 1210–1222, 2020.

[Korz 19] D. Korzekwa *et al.* "Interpretable Deep Learning Model for the Detection and Reconstruction of Dysarthric Speech". In: *Proceedings of INTERSPEECH*, pp. 1–5, 2019.

[Kost 15] N. Kostikis, D. Hristu-Varsakelis, M. Arnaoutoglou, and C. Kotsavasiloglou. "A smartphone-based tool for assessing parkinsonian hand tremor". *IEEE journal of biomedical and health informatics*, Vol. 19, No. 6, pp. 1835–1842, 2015.

[Kots 17] C. Kotsavasiloglou *et al.* "Machine learning-based classification of simple drawing movements in Parkinson's disease". *Biomedical Signal Processing and Control*, Vol. 31, pp. 174–180, 2017.

[Kriz 12] A. Krizhevsky, I. Sutskever, and G. Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[Kuhn 17] A. Kuhner *et al.* "correlations between Motor symptoms across Different Motor Tasks, Quantified via random Forest Feature classification in Parkinson's Disease". *Frontiers in neurology*, Vol. 8, p. 607, 2017.

[Kupf 20] M. Kupfer. "Automated analysis of Parkinson's Disease on the basis of evaluation of handwriting". In: *Bachelor thesis in computer science, Friedrich Alexander University, Erlangen-Nuremberg.*, 2020.

[Laar 17] I. Laaridh, W. B. Kheder, C. Fredouille, and C. Meunier. "Automatic prediction of speech evaluation metrics for dysarthric speech". In: *Proceedings of INTERSPEECH*, pp. 1834–1838, 2017.

[Lamb 21] R. Lamba *et al.* "A systematic approach to diagnose Parkinson's disease through kinematic features extracted from handwritten drawings". *Journal of Reliable Intelligent Environments*, pp. 1–10, 2021.

[LeCu 98] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-based learning applied to document recognition". *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278–2324, 1998.

[Lee 16] W. Lee, A. Evans, and D. R. Williams. "Validation of a smartphone application measuring motor function in Parkinson's disease". *Journal of Parkinson's disease*, Vol. 6, No. 2, pp. 371–382, 2016.

[Lemp 76] A. Lempel and J. Ziv. "On the complexity of finite sequences". *IEEE Transactions on Information Theory*, Vol. 22, No. 1, pp. 75–81, 1976.

[Leta 14] A. Letanneux, J. Danna, J. Velay, F. Viallet, and S. Pinto. "From micrographia to Parkinson's disease dysgraphia". *Movement Disorders*, Vol. 29, No. 12, pp. 1467–1475, 2014.

[Li 18] F. Li, J. Johnson, and S. Yeung. "Lecture notes in Convolutional Neural Networks for Visual Recognition". 2018.

[Lin 98] T. Lin, B. G. Horne, P. Tino, and C. L. Giles. "Learning long-term dependencies is not as difficult with NARX recurrent neural networks". Tech. Rep., 1998.

[Loge 78] J. A. Logemann, H. B. Fisher, B. Boshes, and E. R. Blonsky. "Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients". *Journal of Speech and Hearing Disorders*, Vol. 43, No. 1, pp. 47–57, 1978.

[Lokk 11] J. Lökk. "Lack of information and access to advanced treatment for Parkinson's disease patients". *Journal of multidisciplinary healthcare*, Vol. 4, p. 433, 2011.

[Lope 19] J. V. E. López, J. R. Orozco-Arroyave, and G. Gosztolya. "Assessing Parkinson's Disease From Speech Using Fisher Vectors". *Proc. Interspeech 2019*, pp. 3063–3067, 2019.

[Ma 12] A. J. Ma, P. C. Yuen, and J. H. Lai. "Linear dependency modeling for classifier fusion and feature combination". *IEEE transactions on pattern analysis and machine intelligence*, Vol. 35, No. 5, pp. 1135–1148, 2012.

[Maas 13] A. L. Maas, A. Hannun, and A. Ng. "Rectifier Nonlinearities Improve Neural Network Acoustic Models". In: *Proceedings of ICML*, 2013.

[Mall 20] J. Mallela, A. Illa, *et al.* "Voice based classification of patients with Amyotrophic Lateral Sclerosis, Parkinson's Disease and Healthy Controls with CNN-LSTM using transfer learning". In: *Proceedings of ICASSP*, pp. 6784–6788, 2020.

[Mazi 12] S. Mazilu *et al.* "Online detection of freezing of gait with smartphones and machine learning techniques". In: *International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, pp. 123–130, 2012.

[Mazi 16] S. Mazilu, U. Blanke, *et al.* "The role of wrist-mounted inertial sensors in detecting gait freeze episodes in Parkinson's disease". *Pervasive and Mobile Computing*, Vol. 33, pp. 1–16, 2016.

[Meky 16] J. Mekyska, Z. Smekal, *et al.* "Perceptual features as markers of Parkinson's Disease: the issue of clinical interpretability". In: *Recent Advances in Nonlinear Speech Processing*, pp. 83–91, Springer, 2016.

[Midd 09] C. Middag *et al.* "Automated intelligibility assessment of pathological speech using phonological features". *EURASIP Journal on Advances in Signal Processing*, Vol. 2009, No. 1, p. 629030, 2009.

[Mill 20] G. F. Miller, J. C. Vásquez-Correa, and E. Nöth. "Assessing the Dysarthria Level of Parkinson's Disease Patients with GMM-UBM Supervectors Using Phonological Posteriors and Diadochokinetic Exercises". In: *International Conference on Text, Speech, and Dialogue*, pp. 356–365, Springer, 2020.

[Moet 19] M. Moetesum *et al.* "Assessing visual attributes of handwriting for prediction of neurological disorders – A case study on Parkinson's disease". *Pattern Recognition Letters*, Vol. 121, pp. 19–27, 2019.

[Moet 20] M. Moetesum *et al.* "Dynamic Handwriting Analysis for Parkinson's Disease Identification using C-BiGRU Model". In: *2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pp. 115–120, IEEE, 2020.

[Moha 18] N. Mohammadian Rad, T. van Laarhoven, C. Furlanello, and E. Marchiori. "Novelty Detection using Deep Normative Modeling for IMU-Based Abnormal Movement Monitoring in Parkinson's Disease and Autism Spectrum Disorders". *Sensors*, Vol. 18, No. 10, p. 3533, 2018.

[Monr 57] G. H. Monrad-Krohn. "The third element of speech: prosody in the neuro-psychiatric clinic". *Journal of Mental Science*, Vol. 103, No. 431, pp. 326–331, 1957.

[Mont 18] D. Montaña, Y. Campos-Roca, and C. J. Pérez. "A Diadochokinesis-based expert system considering articulatory features of plosive consonants for early detection of Parkinson's disease". *Computer Methods and Programs in Biomedicine*, Vol. 154, pp. 89–97, 2018.

[Moon 20] S. Moon *et al.* "Classification of Parkinson's disease and essential tremor based on balance and gait characteristics from wearable motion sensors via machine learning techniques: a data-driven approach". *Journal of NeuroEngineering and Rehabilitation*, Vol. 17, No. 1, pp. 1–8, 2020.

[Moor 08] S. T. Moore, H. G. MacDougall, and W. G. Ondo. "Ambulatory monitoring of freezing of gait in Parkinson's disease". *Journal of neuroscience methods*, Vol. 167, No. 2, pp. 340–348, 2008.

[More 03] R. Moretti *et al.* "Speech Initiation Hesitation's following Subthalamic Nucleus Stimulation in a Patient with Parkinson's Disease". *European Neurology*, Vol. 49, No. 4, pp. 251–253, 2003.

[More 09] D. M. Morens, G. K. Folkers, and A. S. Fauci. "What is a pandemic?". *The Journal of Infectious Diseases*, Vol. 200, pp. 1018–1021, 2009.

[Moro 18] L. Moro-Velázquez *et al.* "Analysis of speaker recognition methodologies and the influence of kinetic changes to automatically detect Parkinson's Disease". *Applied Soft Computing*, Vol. 62, pp. 649–666, 2018.

[Moro 19a] L. Moro-Velazquez *et al.* "A forced gaussians based methodology for the differential evaluation of Parkinson's Disease by means of speech processing". *Biomedical Signal Processing and Control*, Vol. 48, pp. 205–220, 2019.

[Moro 19b] L. Moro-Velazquez *et al.* "Phonetic relevance and phonemic grouping of speech in the automatic detection of Parkinson's Disease". *Scientific Reports*, Vol. 9, No. 1, pp. 1–16, 2019.

[Moro 20] L. Moro-Velazquez, J. Villalba, and N. Dehak. "Using X-vectors to automatically detect Parkinson's disease from speech". In: *Proceedings of ICASSP*, pp. 1–5, 2020.

[Moro 21] L. Moro-Velazquez *et al.* "Advances in Parkinson's Disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects". *Biomedical Signal Processing and Control*, Vol. 66, p. 102418, 2021.

[Moze 92] M. C. Mozer. "Induction of multiscale temporal structure". In: *Advances in neural information processing systems*, pp. 275–282, 1992.

[Much 18a] J. Mucha *et al.* "Fractional derivatives of online handwriting: A new approach of parkinsonic dysgraphia analysis". In: *41st International Conference on Telecommunications and Signal Processing (TSP)*, pp. 1–4, 2018.

[Much 18b] J. Mucha *et al.* "Identification and Monitoring of Parkinson's Disease Dysgraphia Based on Fractional-Order Derivatives of Online Handwriting". *Applied Sciences*, Vol. 8, No. 12, p. 2566, 2018.

[Nagh 21] N. Naghavi and E. Wade. "Towards Real-time Prediction of Freezing of Gait in Patients with Parkinsons Disease: A Novel Deep One-class Classifier". *IEEE Journal of Biomedical and Health Informatics*, 2021.

[Nair 10] V. Nair and G. E. Hinton. "Rectified Linear Units Improve Restricted Boltzmann Machines". In: *Proceedings of ICML*, pp. 807–814, 2010.

[Nara 16] L. Naranjo, C. J. Pérez, Y. Campos-Roca, and J. Martín. "Addressing voice recording replications for Parkinson's disease detection". *Expert Systems with Applications*, Vol. 46, pp. 286–292, 2016.

[Nase 20] A. Naseer *et al.* "Refining Parkinson's neurological disorder identification through deep transfer learning". *Neural Computing and Applications*, pp. 1–16, 2020.

[Nati 06] National Collaborating Centre for Chronic Conditions (Great Britain). "Parkinson's disease: national clinical guideline for diagnosis and management in primary and secondary care". Royal College of Physicians, 2006.

[Nila 18] M. Nilashi, O. Ibrahim, H. Ahmadi, L. Shahmoradi, and M. Farahmand. "A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques". *Biocybernetics and Biomedical Engineering*, Vol. 38, No. 1, pp. 1–15, 2018.

[Nila 19] M. Nilashi, O. Ibrahim, S. Samad, H. Ahmadi, L. Shahmoradi, and E. Akbari. "An analytical method for measuring the Parkinson's disease progression: A case on a Parkinson's telemonitoring dataset". *Measurement*, Vol. 136, pp. 545–557, 2019.

[Nola 21] J. A. Nolazco-Flores *et al.* "Exploiting spectral and cepstral handwriting features on diagnosing Parkinson's disease". *IEEE Access*, 2021.

[Nomm 20] S. Nõmm *et al.* "Deep CNN Based Classification of the Archimedes Spiral Drawing Tests to Support Diagnostics of the Parkinson's Disease". *IFAC-PapersOnLine*, Vol. 53, No. 5, pp. 260–264, 2020.

[Nore 20] R. Norel, C. Agurto, *et al.* "Speech-based characterization of dopamine replacement therapy in people with Parkinson's disease". *npj Parkinson's Disease*, Vol. 6, No. 1, pp. 1–8, 2020.

[Novo 14] M. Novotnỳ, J. Rusz, *et al.* "Automatic evaluation of articulatory disorders in Parkinson's disease". *IEEE/ACM Trans. on Audio, Speech and Language Processing*, Vol. 22, No. 9, pp. 1366–1378, 2014.

[Novo 20] M. Novotný, P. Dusek, I. Daly, E. Ruzicka, and J. Rusz. "Glottal Source Analysis of Voice Deficits in Newly Diagnosed Drug-naïve Patients with Parkinson's Disease: Correlation Between Acoustic Speech Characteristics and Non-Speech Motor Performance". *Biomedical Signal Processing and Control*, Vol. 57, p. 101818, 2020.

[Np 21] N. Np, B. Schuller, and P. Alku. "The detection of Parkinsons disease from speech using voice source information". *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2021.

[Nutt 11] J. G. Nutt, B. R. Bloem, N. Giladi, M. Hallett, F. B. Horak, and A. Nieuwboer. "Freezing of gait: moving forward on a mysterious clinical phenomenon". *The Lancet Neurology*, Vol. 10, No. 8, pp. 734–744, 2011.

[Oord 16] A. van den Oord *et al.* "WaveNet: A Generative Model for Raw Audio". In: *9th ISCA Speech Synthesis Workshop*, pp. 125–125, 2016.

[ORei 09] C. O'Reilly and R. Plamondon. "Development of a Sigma–Lognormal representation for on-line signatures". *Pattern Recognition*, Vol. 42, No. 12, pp. 3324–3337, 2009.

[Orne 17] C. Ornelas-Vences *et al.* "Fuzzy inference model evaluating turn for Parkinson's disease patients". *Computers in biology and medicine*, Vol. 89, pp. 379–388, 2017.

[Orne 19] C. Ornelas-Vences *et al.* "Computer model for leg agility quantification and assessment for Parkinson's disease patients". *Medical & biological engineering & computing*, Vol. 57, No. 2, pp. 463–476, 2019.

[Oroz 14] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, M. C. Gonzalez-Rátiva, and E. Nöth. "New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease". In: *Language Resources and Evaluation Conference, (LREC)*, pp. 342–347, 2014.

[Oroz 15] J. R. Orozco-Arroyave, E. A. Belalcázar-Bolaños, *et al.* "Characterization Methods for the Detection of Multiple Voice Disorders: Neurological, Functional, and Laryngeal Diseases". *IEEE Journal of Biomedical and Health Informatics*, Vol. 19, No. 6, pp. 1820–1828, 2015.

[Oroz 16a] J. R. Orozco-Arroyave, J. C. Vásquez-Correa, *et al.* "Towards an automatic monitoring of the neurological state of the Parkinson's patients from speech". In: *Proceedings of ICASSP*, pp. 6490–6494, 2016.

[Oroz 16b] J. Orozco-Arroyave. *Analysis of speech of people with Parkinson's disease.* Logos-Verlag, Berlin, Germany, 1st Ed., 2016.

[Oroz 18] J. R. Orozco-Arroyave, J. C. Vásquez-Correa, *et al.* "NeuroSpeech: An open-source software for Parkinson's speech analysis". *Digital Signal Processing*, Vol. 77, pp. 207–221, 2018.

[Oroz 20a] J. R. Orozco-Arroyave, J. C. Vásquez-Correa, *et al.* "Apkinson: the smartphone application for telemonitoring Parkinson's patients through speech, gait, and hands movement". *Neurodegenerative Disease Management*, Vol. 10, No. 3, pp. 137–157, 2020.

[Oroz 20b] J. R. Orozco-Arroyave, J. C. Vásquez-Correa, and E. Nöth. "Current methods and new trends in signal processing and pattern recognition for the automatic assessment of motor impairments: the case of Parkinson's disease". *Neurological Disorders and Imaging Physics, Volume 5*, pp. 8.1–8.57, 2020.

[Oung 15] Q. W. Oung, H. Muthusamy, *et al.* "Technologies for assessment of motor disorders in Parkinson's disease: a review". *Sensors*, Vol. 15, No. 9, pp. 21710–21745, 2015.

[Oung 18] Q. W. Oung *et al.* "Empirical Wavelet Transform Based Features for Classification of Parkinson's Disease Severity". *Journal of medical systems*, Vol. 42, No. 2, p. 29, 2018.

[Para 08] I. Paradisi *et al.* "Huntington disease mutation in Venezuela: age of onset, haplotype analyses and geographic aggregation". *Journal of human genetics*, Vol. 53, No. 2, pp. 127–135, 2008.

[Pari 15] F. Parisi, G. Ferrari, *et al.* "Body-Sensor-Network-Based Kinematic Characterization and Comparative Outlook of UPDRS Scoring in Leg Agility, Sit-to-Stand, and Gait Tasks in Parkinson's Disease". *IEEE Journal of Biomedical and Health Informatics*, Vol. 19, No. 6, pp. 1777–1793, 2015.

[Park 17] J. Parkinson. *An essay on the shaking palsy.* Whittingham and Rowland, London, UK, 1817.

[Parr 18] L. F. Parra-Gallego, T. Arias-Vergara, J. C. Vásquez-Correa, N. Garcia-Ospina, J. R. Orozco-Arroyave, and E. Nöth. "Automatic Intelligibility Assessment of Parkinson's Disease with Diadochokinetic Exercises". In: *Workshop on Engineering Applications*, pp. 223–230, Springer, 2018.

[Parz 21] A. Parziale *et al.* "Cartesian genetic programming for diagnosis of Parkinson disease through handwriting analysis: Performance vs. interpretability issues". *Artificial Intelligence in Medicine*, Vol. 111, p. 101984, 2021.

[Pasc 19] S. Pascual, M. Ravanelli, J. Serrà, A. Bonafonte, and Y. Bengio. "Learning Problem-Agnostic Speech Representations from Multiple Self-Supervised Tasks". *Proceedings of INTERSPEECH*, pp. 161–165, 2019.

[Past 13] M. Pastorino, M. T. Arredondo, J. Cancela, and S. Guillen. "Wearable sensor network for health monitoring: The case of Parkinson's disease". In: *Journal of Physics: Conference Series*, p. 012055, 2013.

[Pasz 17] A. Paszke, S. Gross, S. Chintala, *et al.* "Automatic differentiation in PyTorch". In: *Conference on Neural Information Processing Systems (NIPS)*, pp. 1–4, 2017.

[Pate 16] S. Patel, S. Parveen, and S. Anand. "Prosodic changes in Parkinson's disease". *The Journal of the Acoustical Society of America*, Vol. 140, No. 4, pp. 3442–3442, 2016.

[Pell 06] M. D. Pell, H. S. Cheang, and C. L. Leonard. "The impact of Parkinson's disease on vocal-prosodic communication from the perspective of listeners". *Brain and language*, Vol. 97, No. 2, pp. 123–134, 2006.

[Pere 16a] C. R. Pereira *et al.* "Convolutional neural networks applied for parkinson's disease identification". In: *Machine Learning for Health Informatics*, pp. 377–390, Springer, 2016.

[Pere 16b] C. R. Pereira *et al.* "A new computer vision-based approach to aid the diagnosis of Parkinson's disease". *Computer methods and programs in biomedicine*, Vol. 136, pp. 79–88, 2016.

[Pere 18a] C. R. Pereira *et al.* "Handwritten dynamics assessment through convolutional neural networks: An application to Parkinson's disease identification". *Artificial intelligence in Medicine*, Vol. 87, pp. 67–77, 2018.

[Pere 18b] J. Pereira and M. Silveira. "Unsupervised anomaly detection in energy time series data using variational recurrent autoencoders with attention". In: *IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1275–1282, IEEE, 2018.

[Pere 18c] P. A. Pérez-Toro *et al.* "A Non-linear Dynamics Approach to Classify Gait Signals of Patients with Parkinson's Disease". In: *Workshop on Engineering Applications*, pp. 268–278, Springer, 2018.

[Pere 19] P. A. Pérez-Toro *et al.* "Natural language analysis to detect Parkinson's disease". In: *International Conference on Text, Speech, and Dialogue*, pp. 82–90, Springer, 2019.

[Pere 20a] J. C. Perez-Ibarra, A. A. Siqueira, and H. I. Krebs. "Identification of gait events in healthy and Parkinson's disease subjects using inertial sensors: A supervised learning approach". *IEEE Sensors Journal*, Vol. 20, No. 24, pp. 14984–14993, 2020.

[Pere 20b] P. Perez-Toro, J. C. Vasquez-Correa, *et al.* "Nonlinear dynamics and Poincaré sections to model gait impairments in different stages of Parkinson's disease". *Nonlinear Dynamics (IN PRESS)*, Vol. 100, No. , pp. 3253–3276, 2020.

[Pere 21a] P. A. Pérez-Toro *et al.* "Acoustic and linguistic analyses to assess early-onset and genetic Alzheimer's disease". In: *Proceedings of ICASSP (Under review)*, pp. 1–5, 2021.

[Pere 21b] P. A. Pérez-Toro *et al.* "Emotional State Modeling for the Assessment of Depression in Parkinson's Disease". In: *International Conference on Text, Speech, and Dialogue*, pp. 457–468, Springer, 2021.

[Pere 21c] P. A. Perez-Toro *et al.* "User State Modeling Based on the Arousal-Valence Plane: Applications in Customer Satisfaction and Health-Care". *IEEE Transactions on Affective Computing*, No. 01, pp. 1–1, 2021.

[Pere 21d] P. A. Pérez-Toro *et al.* "User State Modeling Based on the Arousal-Valence Plane: Applications in Customer Satisfaction and Health-Care". *IEEE Transactions on Affective Computing (Under review)*, Vol. , pp. 1–13, 2021.

[Pfis 20] F. M. Pfister, T. T. Um, *et al.* "High-Resolution Motor State Detection in Parkinson's Disease Using convolutional neural networks". *Scientific reports*, Vol. 10, No. 1, pp. 1–11, 2020.

[Phin 20] A. Phinyomark, R. Larracy, and E. Scheme. "Fractal Analysis of Human Gait Variability via Stride Interval Time Series". *Frontiers in Physiology*, Vol. 11, p. 333, 2020.

[Pine 06] N. Pineda-Trujillo *et al.* "A genetic cluster of early onset Parkinson's disease in a Colombian population". *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, Vol. 141, No. 8, pp. 885–889, 2006.

[Pint 04] S. Pinto, C. Ozsancak, E. Tripoliti, S. Thobois, P. Limousin-Dowsey, and P. Auzou. "Treatments for dysarthria in Parkinson's disease". *The Lancet Neurology*, Vol. 3, No. 9, pp. 547–556, 2004.

[Pomp 20] A. Pompili, R. Solera-Ureña, *et al.* "Assessment of Parkinson's Disease Medication State through Automatic Speech Analysis". *arXiv preprint arXiv:2005.14647*, 2020.

[Post 09] R. B. Postuma *et al.* "Quantifying the risk of neurodegenerative disease in idiopathic REM sleep behavior disorder". *Neurology*, Vol. 72, No. 15, pp. 1296–1300, 2009.

[Post 15] R. B. Postuma, A. Iranzo, *et al.* "Risk factors for neurodegeneration in idiopathic rapid eye movement sleep behavior disorder: a multicenter study". *Annals of neurology*, Vol. 77, No. 5, pp. 830–839, 2015.

[Post 18] R. B. Postuma, W. Poewe, *et al.* "Validation of the MDS clinical diagnostic criteria for Parkinson's disease". *Movement Disorders*, Vol. 33, No. 10, pp. 1601–1608, 2018.

[Post 19] G. Postolache *et al.* "Smartphone Sensing Technologies for Tailored Parkinson's Disease Diagnosis and Monitoring". In: *Innovations in Communication and Computing*, pp. 251–273, Springer, 2019.

[Prab 20] P. Prabhu, A. Karunakar, H. Anitha, and N. Pradhan. "Classification of gait signals into different neurodegenerative diseases using statistical analysis and recurrence quantification analysis". *Pattern Recognition Letters*, Vol. 139, pp. 10–16, 2020.

[Prin 14] B. P. Printy *et al.* "Smartphone application for classification of motor impairment severity in Parkinson's disease". In: *Proceedings of EMBC*, pp. 2686–2689, 2014.

[Quan 21] C. Quan, K. Ren, and Z. Luo. "A Deep Learning Based Method for Parkinson's Disease Detection Using Dynamic Features of Speech". *IEEE Access*, Vol. 9, pp. 10239–10252, 2021.

[Rame 17] H. Ramezani, H. Khaki, E. Erzin, and O. B. Akan. "Speech features for telemonitoring of Parkinson's disease symptoms". In: *Proceedings of (EMBC)*, pp. 3801–3805, 2017.

[Rava 18] M. Ravanelli and Y. Bengio. "Speaker recognition from raw waveform with sincnet". In: *2018 IEEE Spoken Language Technology Workshop (SLT)*, pp. 1021–1028, 2018.

[Rava 20a] V. Raval, K. P. Nguyen, *et al.* "Prediction of individual progression rate in Parkinson's disease using clinical measures and biomechanical measures of gait and postural stability". In: *Proceedings of ICASSP*, pp. 1319–1323, 2020.

[Rava 20b] M. Ravanelli, J. Zhong, S. Pascual, P. Swietojanski, J. Monteiro, J. Trmal, and Y. Bengio. "Multi-task self-supervised learning for Robust Speech Recognition". In: *Proceedings of ICASSP*, 2020.

[Rehm 19] R. Z. U. Rehman, S. Del Din, Y. Guan, A. J. Yarnall, J. Q. Shi, and L. Rochester. "Selecting clinically relevant gait characteristics for classification of early Parkinson's disease: A comprehensive machine learning approach". *Scientific reports*, Vol. 9, No. 1, pp. 1–12, 2019.

[Rehm 20] R. Z. U. Rehman, C. Buckley, *et al.* "Accelerometry-Based Digital Gait Characteristics for Classification of Parkinson's Disease: What Counts?". *IEEE Open Journal of Engineering in Medicine and Biology*, Vol. 1, pp. 65–73, 2020.

[Ren 16] P. Ren *et al.* "Analysis of gait rhythm fluctuations for neurodegenerative diseases by phase synchronization and conditional entropy". *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 24, No. 2, pp. 291–299, 2016.

[Reyn 00] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn. "Speaker verification using adapted Gaussian mixture models". *Digital signal processing*, Vol. 10, No. 1-3, pp. 19–41, 2000.

[Rezv 16] S. Rezvanian and T. Lockhart. "Towards real-time detection of freezing of gait using wavelet transform on wireless accelerometer data". *Sensors*, Vol. 16, No. 4, p. 475, 2016.

[Ribe 19] L. C. Ribeiro, L. C. Afonso, and J. P. Papa. "Bag of Samplings for computer-assisted Parkinson's disease diagnosis based on Recurrent Neural Networks". *Computers in biology and medicine*, Vol. 115, p. 103477, 2019.

[Rich 00] J. S. Richman and J. R. Moorman. "Physiological time-series analysis using approximate entropy and sample entropy". *American Journal of Physiology-Heart and Circulatory Physiology*, Vol. 278, No. 6, pp. H2039–H2049, 2000.

[Riga 12] G. Rigas *et al.* "Assessment of tremor activity in the Parkinson's disease using a set of wearable sensors". *IEEE Transactions on Information Technology in Biomedicine*, Vol. 16, No. 3, pp. 478–487, 2012.

[Rios 19] C. D. Rios-Urrego, J. C. Vásquez-Correa, J. F. Vargas-Bonilla, E. Nöth, F. Lopera, and J. R. Orozco-Arroyave. "Analysis and Evaluation of Handwriting in Patients with Parkinson's Disease Using kinematic, Geometrical, and Non-linear Features". *Computer Methods and Programs in Biomedicine*, Vol. 173, pp. 43–52, 2019.

[Rios 20] C. D. D. Rios-Urrego *et al.* "Transfer Learning to Detect Parkinson's Disease from Speech In Different Languages Using Convolutional Neural Networks with Layer Freezing". In: *International Conference on Text, Speech, and Dialogue*, pp. 331–339, Springer, 2020.

[Rios 21] C. D. Rios-Urrego *et al.* "Is There Any Additional Information in a Neural Network Trained for Pathological Speech Classification?". In: *International Conference on Text, Speech, and Dialogue*, pp. 435–447, Springer, 2021.

[Rizz 16] G. Rizzo, M. Copetti, S. Arcuti, D. Martino, A. Fontana, and G. Logroscino. "Accuracy of clinical diagnosis of Parkinson disease: a systematic review and meta-analysis". *Neurology*, Vol. 86, No. 6, pp. 566–576, 2016.

[Robe 82] S. J. Robertson. "Robertson Dysarthria Profile". *Buckinghamshire: Winslow*, 1982.

[Rodr 09] M. C. Rodriguez-Oroz *et al.* "Initial clinical manifestations of Parkinson's disease: features and pathophysiological mechanisms". *The Lancet Neurology*, Vol. 8, No. 12, pp. 1128–1139, 2009.

[Rodr 17] D. Rodríguez-Martín *et al.* "Home detection of freezing of gait using support vector machines through a single waist-worn triaxial accelerometer". *PloS one*, Vol. 12, No. 2, p. e0171764, 2017.

[Roma 21] A. Romana *et al.* "Automatically Detecting Errors and Disfluencies in Read Speech to Predict Cognitive Impairment in People with Parkinson's Disease". *Proc. Interspeech 2021*, pp. 1907–1911, 2021.

[Rose 13] S. Rosenblum, M. Samuel, S. Zlotnik, I. Erikh, and I. Schlesinger. "Handwriting as an objective tool for Parkinson's disease diagnosis". *Journal of Neurology*, Vol. 260, No. 9, pp. 2357–2361, 2013.

[Rose 93] M. T. Rosenstein, J. J. Collins, and C. J. De Luca. "A practical method for calculating largest Lyapunov exponents from small data sets". *Physica D: Nonlinear Phenomena*, Vol. 65, No. 1-2, pp. 117–134, 1993.

[Rude 16] S. Ruder. "An overview of gradient descent optimization algorithms". *arXiv preprint arXiv:1609.04747*, 2016.

[Rued 19] A. Rueda *et al.* "Feature Representation of Pathophysiology of Parkinsonian Dysarthria". In: *Proceedings of INTERSPEECH*, pp. 3048–3052, 2019.

[Rusz 11] J. Rusz, R. Cmejla, H. Ruzickova, and E. Ruzicka. "Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease". *journal of the Acoustical Society of America*, Vol. 129, No. 1, pp. 350–367, 2011.

[Rusz 13] J. Rusz, R. Cmejla, *et al.* "Imprecise vowel articulation as a potential early marker of Parkinson's disease: Effect of speaking task". *The Journal of the Acoustical Society of America*, Vol. 134, No. 3, pp. 2171–2181, 2013.

[Rusz 15] J. Rusz, C. Bonnet, *et al.* "Speech disorders reflect differing pathophysiology in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy". *Journal of Neurology*, Vol. 262, No. 4, pp. 992–1001, 2015.

[Rusz 16] J. Rusz, T. Tykalová, *et al.* "Effects of dopaminergic replacement therapy on motor speech disorders in Parkinson's disease: longitudinal follow-up study on previously untreated patients". *Journal of Neural Transmission*, Vol. 123, No. 4, pp. 379–387, 2016.

[Rusz 17] J. Rusz *et al.* "Comparative analysis of speech impairment and upper limb motor dysfunction in Parkinson's disease". *Journal of Neural Transmission*, Vol. 124, No. 4, pp. 463–470, 2017.

[Rusz 18a] J. Rusz, J. Hlavnicka, *et al.* "Smartphone allows capture of speech abnormalities associated with high risk of developing Parkinson's disease". *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2018.

[Rusz 18b] J. Rusz. *Detecting speech disorders in early Parkinson's disease by acoustic analysis*. Habilitation thesis, Czech Technical University in Prague, 2018.

[Sade 00] H. Sadeghi *et al.* "Symmetry and limb dominance in able-bodied gait: a review". *Gait & posture*, Vol. 12, No. 1, pp. 34–45, 2000.

[Saka 13] B. E. Sakar *et al.* "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings". *IEEE Journal of Biomedical and Health Informatics*, Vol. 17, No. 4, pp. 828–834, 2013.

[Saka 19] C. O. Sakar *et al.* "A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform". *Applied Soft Computing*, Vol. 74, pp. 255–263, 2019.

[Sala 04] A. Salarian *et al.* "Gait assessment in Parkinson's disease: toward an ambulatory system for long-term monitoring". *IEEE transactions on Biomedical Engineering*, Vol. 51, No. 8, pp. 1434–1443, 2004.

[Sama 18] A. Samà *et al.* "Determining the optimal features in freezing of gait detection through a single waist accelerometer in home environments". *Pattern Recognition Letters*, Vol. 105, pp. 135–143, 2018.

[Sanc 18] L. A. Sanchez-Perez *et al.* "Rest tremor quantification based on fuzzy inference systems and wearable sensors". *International journal of medical informatics*, Vol. 114, pp. 6–17, 2018.

[Sarb 13] Y. Sarbaz, F. Towhidkhah, V. Mosavari, A. Janani, and A. Soltanzadeh. "Separating Parkinsonian patients from normal persons using handwriting features". *Journal of Mechanics in Medicine and Biology*, Vol. 13, No. 03, p. 1350030, 2013.

[Saun 08] R. Saunders-Pullman *et al.* "Validity of spiral analysis in early Parkinson's disease". *Movement disorders*, Vol. 23, No. 4, pp. 531–537, 2008.

[Scha 03] J. D. Schaafsma *et al.* "Characterization of freezing of gait subtypes and the response of each to levodopa in Parkinson's disease". *European journal of neurology*, Vol. 10, No. 4, pp. 391–398, 2003.

[Sche 18] F. Scheperjans, P. Derkinderen, and P. Borghammer. "The gut and Parkinson's disease: hype or hope?". *Journal of Parkinson's disease*, Vol. 8, No. s1, pp. S31–S39, 2018.

[Schi 99] F. Schiel. "Automatic phonetic transcription of non-prompted speech". In: *Proceedings of the ICPhS*, pp. 607–610, 1999.

[Schu 15] B. Schuller, S. Steidl, A. Batliner, S. Hantke, F. Hönig, J. R. Orozco-Arroyave, E. Nöth, Y. Zhang, and F. Weninger. "The INTERSPEECH 2015 computational paralinguistics challenge: Nativeness, Parkinson's & eating condition". In: *Proceedings of INTERSPEECH*, pp. 478–482, 2015.

[Schu 97] M. Schuster and K. K. Paliwal. "Bidirectional recurrent neural networks". *IEEE transactions on Signal Processing*, Vol. 45, No. 11, pp. 2673–2681, 1997.

[Sech 21] K. Sechidis *et al.* "A machine learning perspective on the emotional content of Parkinsonian speech". *Artificial Intelligence in Medicine*, Vol. 115, p. 102061, 2021.

[Sejd 14] E. Sejdić, K. A. Lowry, J. Bellanca, M. S. Redfern, and J. S. Brach. "A comprehensive assessment of gait accelerometry signals in time, frequency and time-frequency domains". *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 22, No. 3, p. 603, 2014.

[Sena 19] R. Senatore, A. Della-Cioppa, and A. Marcelli. "Automatic Diagnosis of Neurodegenerative Diseases: An Evolutionary Approach for Facing the Interpretability Problem". *Information*, Vol. 10, No. 1, p. 30, 2019.

[Seti 21] F. Setiawan and C. W. Lin. "Implementation of a Deep Learning Algorithm Based on Vertical Ground Reaction Force Time–Frequency Features for the Detection and Severity Classification of Parkinson's Disease". *Sensors*, Vol. 21, No. 15, p. 5207, 2021.

[Shin 06] M. Shin, S. Park, *et al.* "Clinical and empirical applications of the Rey–Osterrieth complex figure test". *Nature Protocols*, Vol. 1, No. 2, pp. 892–899, 2006.

[Shul 14] P. B. Shull *et al.* "Quantified self and human movement: a review on the clinical impact of wearable sensing and feedback for gait analysis and intervention". *Gait & posture*, Vol. 40, No. 1, pp. 11–19, 2014.

[Silv 17] A. L. Silva-De-Lima *et al.* "Freezing of gait and fall detection in Parkinson's disease using wearable sensors: a systematic review". *Journal of neurology*, Vol. 264, No. 8, pp. 1642–1654, 2017.

[Skod 11a] S. Skodda, W. Grönheit, and U. Schlegel. "Gender-related patterns of dysprosody in Parkinson disease and correlation between speech variables and motor symptoms". *Journal of Voice*, Vol. 25, No. 1, pp. 76–82, 2011.

[Skod 11b] S. Skodda, W. Grönheit, and U. Schlegel. "Intonation and speech rate in Parkinson's disease: General and dynamic aspects and responsiveness to levodopa admission". *Journal of Voice*, Vol. 25, No. 4, pp. e199–e205, 2011.

[Smit 14] E. J. Smits *et al.* "Standardized handwriting to assess bradykinesia, micrographia and tremor in Parkinson's disease". *PloS one*, Vol. 9, No. 5, p. e97614, 2014.

[Smit 17a] K. M. Smith, J. R. Williamson, and T. F. Quatieri. "Vocal markers of motor, cognitive, and depressive symptoms in Parkinson's disease". In: *International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 71–78, 2017.

[Smit 17b] E. J. Smits *et al.* "Graphical tasks to measure upper limb function in patients with Parkinson's disease: Validity and response to dopaminergic medication". *IEEE journal of biomedical and health informatics*, Vol. 21, No. 1, pp. 283–289, 2017.

[Snyd 18] D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, and S. Khudanpur. "X-vectors: Robust DNN embeddings for speaker recognition". In: *Proceedings of ICASSP*, pp. 5329–5333, 2018.

[Spen 05] K. A. Spencer and M. A. Rogers. "Speech motor programming in hypokinetic and ataxic dysarthria". *Brain and Language*, Vol. 94, No. 3, pp. 347–366, 2005.

[Stam 18] C. Stamate *et al.* "The cloudUPDRS app: A medical device for the clinical assessment of Parkinson's Disease". *Pervasive and mobile computing*, Vol. 43, pp. 146–166, 2018.

[Star 90] S. E. Starkstein *et al.* "Depression in Parkinson's disease.". *Journal of Nervous and Mental Disease*, 1990.

[Take 81] F. Takens *et al.* "Detecting strange attractors in turbulence". *Lecture notes in mathematics*, Vol. 898, No. 1, pp. 366–381, 1981.

[Tale 20] C. Taleb, L. Likforman-Sulem, and C. Mokbel. "Improving Deep Learning Parkinson's Disease Detection Through Data Augmentation Training". In: *Mediterranean Conference on Pattern Recognition and Artificial Intelligence*, pp. 79–93, Springer, 2020.

[Talk 95] D. Talkin and W. B. Kleijn. "A robust algorithm for pitch tracking (RAPT)". *Speech coding and synthesis*, Vol. 495, p. 518, 1995.

[Tana 11] Y. Tanaka, M. Nishio, and S. Niimi. "Vocal acoustic characteristics of patients with Parkinson's disease". *Folia Phoniatrica et logopaedica*, Vol. 63, No. 5, pp. 223–230, 2011.

[Tava 05] T. Tavares *et al.* "Quantitative measurements of alternating finger tapping in Parkinson's disease correlate with UPDRS motor disability and reveal the improvement in fine motor control from medication and deep brain stimulation". *Movement disorders*, Vol. 20, No. 10, pp. 1286–1298, 2005.

[Toos 15] N. Toosizadeh, J. Mohler, H. Lei, S. Parvaneh, S. Sherman, and B. Najafi. "Motor performance assessment in Parkinson's disease: association between objective in-clinic, objective in-home, and subjective/semi-objective measures". *PloS one*, Vol. 10, No. 4, p. e0124763, 2015.

[Torv 18] V. G. Torvi, A. Bhattacharya, and S. Chakraborty. "Deep Domain Adaptation to Predict Freezing of Gait in Patients with Parkinson's Disease". In: *IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1001–1006, 2018.

[Trav 17] C. M. Travieso *et al.* "Detection of different voice diseases based on the nonlinear characterization of speech signals". *Expert Systems With Applications*, No. 82, pp. 184–195, 2017.

[Trip 13] E. E. Tripoliti *et al.* "Automatic detection of freezing of gait events in patients with Parkinson's disease". *Computer Methods and Programs in Biomedicine*, Vol. 110, No. 1, pp. 12–26, 2013.

[Tron 09] R. Tronci, G. Giacinto, and F. Roli. "Dynamic score combination: A supervised and unsupervised score combination method". In: *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, pp. 163–177, Springer, 2009.

[Tsan 10] A. Tsanas *et al.* "Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests". *IEEE transactions on Biomedical Engineering*, Vol. 57, No. 4, pp. 884–893, 2010.

[Tsan 12] A. Tsanas, M. A. Little, *et al.* "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease". *IEEE Transactions on Biomedical Engineering*, Vol. 59, No. 5, pp. 1264–1271, 2012.

[Tsan 14] A. Tsanas, M. A. Little, C. Fox, and L. O. Ramig. "Objective automatic assessment of rehabilitative speech treatment in Parkinson's disease". *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 22, No. 1, pp. 181–190, 2014.

[Tu 17a] M. Tu, V. Berisha, and J. Liss. "Interpretable Objective Assessment of Dysarthric Speech Based on Deep Neural Networks". In: *Proceedings of INTERSPEECH*, pp. 1849–1853, 2017.

[Tu 17b] M. Tu, V. Berisha, and J. Liss. "Objective assessment of pathological speech using distribution regression". In: *Proceedings of ICASSP*, pp. 5050–5054, 2017.

[Tyka 17] T. Tykalova *et al.* "Distinct patterns of imprecise consonant articulation among Parkinson's disease, progressive supranuclear palsy and multiple system atrophy". *Brain and language*, Vol. 165, pp. 1–9, 2017.

[Tzal 14] A. T. Tzallas, M. G. Tsipouras, *et al.* "PERFORM: a system for monitoring, assessment and management of patients with Parkinson's disease". *Sensors*, Vol. 14, No. 11, pp. 21329–21357, 2014.

[Vaic 17] E. Vaiciukynas, A. Verikas, A. Gelzinis, and M. Bacauskiene. "Detecting Parkinson's disease from sustained phonation and speech signals". *PloS one*, Vol. 12, No. 10, p. e0185613, 2017.

[Varr 21] T. Varrecchia, S. F. Castiglia, A. Ranavolo, C. Conte, A. Tatarelli, G. Coppola, C. Di Lorenzo, F. Draicchio, F. Pierelli, and M. Serrao. "An artificial neural network approach to detect presence and severity of Parkinson's disease via gait parameters". *Plos one*, Vol. 16, No. 2, p. e0244396, 2021.

[Vasq 17a] J. C. Vásquez-Correa, J. R. Orozco-Arroyave, and E. Nöth. "Convolutional Neural Network to Model Articulation Impairments in Patients with Parkinson's Disease". In: *Proceedings of INTERSPEECH*, pp. 314–318, 2017.

[Vasq 17b] J. C. Vásquez-Correa, J. Serra, J. R. Orozco-Arroyave, J. F. Vargas-Bonilla, and E. Nöth. "Effect of acoustic conditions on algorithms to detect Parkinson's disease from speech". In: *Proceedings of ICASSP*, pp. 5065–5069, 2017.

[Vasq 18a] J. C. Vásquez-Correa, T. Arias-Vergara, J. R. Orozco-Arroyave, and E. Nöth. "A Multitask Learning Approach to Assess the Dysarthria Severity in Patients with Parkinson's Disease". In: *Proceedings of INTERSPEECH*, pp. 456–460, 2018.

[Vasq 18b] J. C. Vásquez-Correa, J. R. Orozco-Arroyave, T. Bocklet, and E. Nöth. "Towards an automatic evaluation of the dysarthria level of patients with Parkinson's disease". *Journal of Communication Disorders*, Vol. 76, pp. 21–36, 2018.

[Vasq 19a] J. C. Vásquez-Correa *et al.* "Apkinson: a Mobile Solution for Multimodal Assessment of Patients with Parkinson's Disease". In: *Proceedings of INTERSPEECH*, pp. 964–965, 2019.

[Vasq 19b] J. C. Vásquez-Correa, P. Klumpp, J. R. Orozco-Arroyave, and E. Nöth. "Phonet: a Tool Based on Gated Recurrent Neural Networks to Extract Phonological Posteriors from Speech". In: *Proceedings of INTERSPEECH*, pp. 549–553, 2019.

[Vasq 19c] J. C. Vásquez-Correa, T. Arias-Vergara, J. Orozco-Arroyave, B. Eskofier, J. Klucken, and E. Nöth. "Multimodal assessment of Parkinson's disease: a deep learning approach". *IEEE journal of biomedical and health informatics*, Vol. 23, No. 4, pp. 1618–1630, 2019.

[Vasq 20a] J. C. Vasquez-Correa, T. Arias-Vergara, M. Schuster, J. R. Orozco-Arroyave, and N. E. "Parallel Representation Learning for the Classification of Pathological Speech: Studies on Parkinson's Disease and Cleft Lip and Palate". *Speech Communication*, Vol. 122, pp. 56–67, 2020.

[Vasq 20b] J. C. Vasquez-Correa *et al.* "Comparison of User Models Based on GMM-UBM and I-Vectors for Speech, Handwriting, and Gait Assessment of Parkinson's Disease Patients". In: *Proceedings of ICASSP*, pp. 6544–6548, 2020.

[Vasq 21a] J. C. Vasquez-Correa *et al.* "End-2-End Modeling of Speech and Gait from Patients with Parkinson's Disease: Comparison Between High Quality Vs. Smartphone Data". In: *Proceedings of ICASSP*, pp. 7298–7302, IEEE, 2021.

[Vasq 21b] J. C. Vásquez-Correa *et al.* "On Modeling Glottal Source Information for Phonation Assessment in Parkinson's Disease". *Proc. Interspeech 2021*, pp. 26–30, 2021.

[Vasq 21c] J. C. Vásquez-Correa *et al.* "Transfer learning helps to improve the accuracy to classify patients with different speech disorders in different languages". *Pattern Recognition Letters*, 2021.

[Vess 19] G. Vessio. "Dynamic Handwriting Analysis for Neurodegenerative Disease Assessment: A Literary Review". *Applied Sciences*, Vol. 9, No. 21, p. 4666, 2019.

[Vidy 21] B. Vidya and P. Sasikumar. "Gait based Parkinson's disease diagnosis and severity rating using multi-class support vector machine". *Applied Soft Computing*, p. 107939, 2021.

[Vill 15] T. Villa-Cañas *et al.* "Low-Frequency Components Analysis in Running Speech for the Automatic Detection of Parkinson's Disease". In: *Proceedings of INTERSPEECH*, pp. 100–104, 2015.

[Visw 20] R. Viswanathan, S. P. Arjunan, *et al.* "Complexity Measures of Voice Recordings as a Discriminative Tool for Parkinson's Disease". *Biosensors*, Vol. 10, No. 1, p. 1, 2020.

[Wals 12] B. Walsh and A. Smith. "Basic parameters of articulatory movements and acoustics in individuals with Parkinson's disease". *Movement Disorders*, Vol. 27, No. 7, pp. 843–850, 2012.

[Wenz 00] R. Wenzelburger *et al.* "Kinetic tremor in a reach-to-grasp movement in Parkinson's disease". *Movement disorders*, Vol. 15, No. 6, pp. 1084–1094, 2000.

[West 94] J. R. Westbury. "X-ray microbeam speech production database user's handbook version 1.0". *Waisman Center on Mental Retardation & Human Development, University of Wisconsin, Madison, WI*, 1994.

[Will 11] A. Willis, M. Schootman, B. Evanoff, J. Perlmutter, and B. Racette. "Neurologist care in Parkinson's disease: a utilization, outcomes, and survival study". *Neurology*, Vol. 77, No. 9, pp. 851–857, 2011.

[Wodz 19] M. Wodzinski *et al.* "Deep Learning Approach to Parkinson's Disease Detection Using Voice Recordings and Convolutional Neural Network Dedicated to Image Classification". In: *Proceedings of EMBC*, pp. 717–720, 2019.

[Wold 07] B. Woldert-Jokisz. "Saarbruecken voice database". 2007.

[Wort 13] P. F. Worth. "How to treat Parkinson's disease in 2013". *Clinical medicine*, Vol. 13, No. 1, p. 93, 2013.

[Wu 04] Y. Wu *et al.* "Optimal multimodal fusion for multimedia data analysis". In: *Proceedings of ACM*, pp. 572–579, 2004.

[Wu 18] K. Wu, D. Zhang, G. Lu, and Z. Guo. "Learning acoustic features to detect Parkinson's disease". *Neurocomputing*, Vol. 318, pp. 102–108, 2018.

[Xia 15] Y. Xia, Q. Gao, and Q. Ye. "Classification of gait rhythm signals between patients with neuro-degenerative diseases and normal subjects: Experiments with statistical features and different classification models". *Biomedical Signal Processing and Control*, Vol. 18, pp. 254–262, 2015.

[Xia 18] Y. Xia *et al.* "Evaluation of deep convolutional neural networks for detection of freezing of gait in Parkinson's disease patients". *Biomedical Signal Processing and Control*, Vol. 46, pp. 221–230, 2018.

[Yaha 19] G. Yahalom *et al.* "Carriers of both GBA and LRRK2 mutations, compared to carriers of either, in Parkinson's disease: Risk estimates and genotype-phenotype correlations". *Parkinsonism & related disorders*, Vol. 62, pp. 179–184, 2019.

[Yang 08] Y. Yang, Y. Lee, S. Cheng, P. Lin, and R. Wang. "Relationships between gait and dynamic balance in early Parkinson's disease". *Gait & posture*, Vol. 27, No. 4, pp. 611–615, 2008.

[You 10] C. H. You, K. A. Lee, and H. Li. "GMM-SVM kernel with a Bhattacharyya-based distance for speaker recognition". *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 18, No. 6, pp. 1300–1312, 2010.

[Zach 15] H. Zach *et al.* "Identifying freezing of gait in Parkinson's disease during freezing provoking tasks using waist-mounted accelerometry". *Parkinsonism & related disorders*, Vol. 21, No. 11, pp. 1362–1366, 2015.

[Zeng 19] W. Zeng, C. Yuan, Q. Wang, F. Liu, and Y. Wang. "Classification of gait patterns between patients with Parkinson's disease and healthy controls using phase space reconstruction (PSR), empirical mode decomposition (EMD) and neural networks". *Neural Networks*, Vol. 111, pp. 64–76, 2019.

[Zham 18] P. Zham *et al.* "Efficacy of guided spiral drawing in the classification of Parkinson's disease". *IEEE Journal of Biomedical and Health Informatics*, Vol. 22, No. 5, pp. 1648–1652, 2018.

[Zham 19] P. Zham *et al.* "Effect of levodopa on handwriting tasks of different complexity in Parkinson's disease: a kinematic study". *Journal of neurology*, pp. 1–7, 2019.

[Zhan 16] X. Y. Zhang, G. S. Xie, C. L. Liu, and Y. Bengio. "End-to-end online writer identification with recurrent neural network". *IEEE Transactions on Human-Machine Systems*, Vol. 47, No. 2, pp. 285–292, 2016.

[Zhan 17] Y. N. Zhang. "Can a Smartphone Diagnose Parkinson Disease? A Deep Neural Network Method and Telediagnosis System Implementation". *Parkinson's Disease*, Vol. 2017, pp. 1–11, 2017.

[Zhan 18] A. Zhan *et al.* "Using smartphones and machine learning to quantify Parkinson disease severity: the mobile Parkinson disease score". *JAMA neurology*, Vol. 75, No. 7, pp. 876–880, 2018.

[Zhan 19] H. Zhang, C. Song, A. Wang, C. Xu, D. Li, and W. Xu. "PDVocal: Towards Privacy-preserving Parkinson's Disease Detection using Non-speech Body Sounds". In: *Proceedings of Mobicom*, pp. 1,17, 2019.

[Zhan 20] Y. Zhang *et al.* "Prediction of Freezing of Gait in Patients with Parkinson's Disease by Identifying Impaired Gait Patterns". *IEEE transactions on neural systems and rehabilitation engineering*, Vol. 28, No. 3, pp. 591–600, 2020.

[Zhao 14] S. Zhao, F. Rudzicz, L. G. Carvalho, C. Márquez-Chin, and S. Livingstone. "Automatic detection of expressed emotion in Parkinson's disease". In: *Proceedings of ICASSP*, pp. 4813–4817, 2014.

[Zhao 15] S. Zhao and F. Rudzicz. "Classifying phonological categories in imagined and articulated speech". In: *Proceedings of ICASSP*, pp. 992–996, 2015.

[Zhao 18] A. Zhao, L. Qi, J. Li, J. Dong, and H. Yu. "A hybrid spatio-temporal model for detection and severity rating of Parkinson's Disease from gait data". *Neurocomputing*, Vol. 315, pp. 1–8, 2018.

[Zwic 80] E. Zwicker and E. Terhardt. "Analytical expressions for critical-band rate and critical bandwidth as a function of frequency". *The Journal of the Acoustical Society of America*, Vol. 68, No. 5, pp. 1523–1525, 1980.

Index