Data-driven classification algorithms for identification and characterization of early neurodegeneration

Cesari, Matteo

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Technical University of Denmark

Data-driven classification algorithms for identification and characterization of early neurodegeneration

Matteo Cesari

Ph.D. Thesis, October 2019

DTU Health Tech
Department of Health Technology

PARACELSUS ELENA-KLINIK Kassel

REGION Rigshospitalet

Data-driven classification algorithms for identification and characterization of early neurodegeneration

Matteo Cesari

Ph.D. Thesis, October 2019

Supervisors:

Helge Bjarup Dissing Sørensen, Main Supervisor, Assoc. Prof., MSK, Ph.D., Head of Biomedical Signal Processing Research Group, Department of Health Technology, Technical University of Denmark, Kgs. Lyngby, Denmark

Poul Jørgen Jennum, Co-supervisor, Prof., M.D., Ph.D., Chief Physician, Danish Center for Sleep Medicine, Department of Clinical Neurophysiology, Rigshospitalet Glostrup, Denmark

Julie Anja Engelhard Christensen, Co-supervisor, Assoc. Clinical Prof., Ph.D., Department of Health Technology, Technical University of Denmark, Kgs. Lyngby, Denmark and Danish Center for Sleep Medicine, Department of Clinical Neurophysiology, Rigshospitalet Glostrup, Denmark

DTU - Technical University of Denmark, Kongens Lyngby - 2019

Data-driven classification algorithms for identification and characterization of early neurodegeneration

This report was prepared by:
Matteo Cesari

Supervisors:
Helge Bjarup Dissing Sørensen, Main Supervisor, Assoc. Prof., MSK, Ph.D., Head of Biomedical Signal Processing Research Group, Department of Health Technology, Technical University of Denmark, Kgs. Lyngby, Denmark
Poul Jørgen Jennum, Co-supervisor, Prof., M.D., Ph.D., Chief Physician, Danish Center for Sleep Medicine, Department of Clinical Neurophysiology, Rigshospitalet Glostrup, Denmark
Julie Anja Engelhard Christensen, Co-supervisor, Assoc. Clinical Prof., Ph.D., Department of Health Technology, Technical University of Denmark, Kgs. Lyngby, Denmark and Danish Center for Sleep Medicine, Department of Clinical Neurophysiology, Rigshospitalet Glostrup, Denmark

DTU Health Tech
Digital Health
Technical University of Denmark
Ørsted Plads, Building 349
2800 Kgs. Lyngby
Denmark

Remarks: Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Technical University of Denmark.

To my family
Alla mia famiglia

"This sign means negative!"; "This sign means positive!"; "Can you borrow me the pencil?"; "And what does this mean?"; "It means that it is always needed to underline the positive!".

Changes from original version

Abstract (pag. xiii)

Original: The method showed higher performances for identifying RBD patients than previously developed methods (accuracy and sensitivity over 80% and specificity over 70%)

Modified to: The method showed higher performances for identifying RBD patients than previously developed methods (accuracy and specificity over 80% and sensitivity over 70%)

Resume(pag. xv)

Original: Metoden viste højere performance til identifikation af RBD-patienter end tidligere udviklede metoder (nøjagtighed og sensitivitet over 80% og specificitet over 70%)

Modified to: Metoden viste højere performance til identifikation af RBD-patienter end tidligere udviklede metoder (nøjagtighed og specificitet over 80% og sensitivitet over 70%)

Chapter 4 (pag. 41)

Original: This chapter is based upon Paper II [94] and contributes in addressing thesis objective 2.

Modified to: This chapter is based upon Paper II [115] and contributes in addressing thesis objective 2.

Table of Contents

Preface

The research providing the foundation of this Ph.D. thesis has been carried out in the Biomedical Engineering Section at the Department of Electrical Engineering (from October 2016 to December 2018) and in the Digital Health Section at the Department of Health Technology (from January to October 2019) at Technical University of Denmark (DTU) in cooperation with the Danish Center for Sleep Medicine, Rigshospitalet Glostrup, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Engineering.

The research work presented in this dissertation has been carried out over a period of three years, along with other activities such as teaching assistant activity in the course Personal Portable Health Technologies, participating in conferences and supervising students through bachelor and master projects.

The dissertation consists of a summary report and six research papers written during the period 2016-2019. Five of them are published, and the last one is submitted.

Kgs. Lyngby, October 2019
Matteo Cesari

Acknowledgements

First of all, I would like to express my deepest appreciation to my main supervisor Assoc. Prof. *Helge B. D. Sørensen*, for believing in me as the right candidate for this project and for providing support and technical guidance throughout the three years. My deepest appreciation also goes to my clinical co-supervisor Prof. *Poul J. Jennum*, for providing support and guidance and for making me passionate about sleep research. My profound gratefulness goes to my technical and clinical co-supervisor Clinical Assoc. Prof. *Julie A. E. Christensen*, for transmitting me her passion and commitment in sleep research and for being always helpful both as a co-supervisor as well as a great friend. I would like to thank all three supervisors for always encouraging me and for the scientific collaboration and scientifically interesting discussions we had. All of you made this project exciting.

I would like to express my profound gratefulness to Prof. *Claudia Trenkwalder*, Assoc. Prof. *Friederike Sixel-Döring* and M.D. *Maria-Lucia Muntean* for their useful inputs in the project, for giving me the opportunity of working on their unique database and for hosting me during my external research stay in Kassel. I would like to thank also Assoc. Prof. *Geert Mayer* and Prof. *Wolfgang H. Oertel* for the useful discussions we had during the project. All of you contributed substantially with ideas and research questions in these three years, and I learned a lot from you.

A big thank goes to *all my colleagues, Ph.D. students and postdocs* from the Biomedical Signal Processing and A.I. Research Group, the Section of Digital Health and the Section of Biomedical Engineering. Our scientific discussions have been valuable and inspiring. Moreover, our lunch and coffee breaks made me getting to know different, new and interesting cultures.

I would like to also thank the staff at the *Danish Center for Sleep Medicine*, in particular *Helle* and *Sara*, for the interesting discussions and for the enjoyable time we spent together.

I want to thank all the students I have supervised and I have been teaching to in these three years. I have learned a lot from all of you, both professionally and personally.

A special thank goes to the many friends I met in these years in Copenhagen. In particular, I would like to thank *Aaron*, *Andrea L.*, *Andrea D.*, *Anna*, *Beatrice*, *Giulia*, *Ilaria*, *Lana*, *Lorenzo*, *Markus* and *Martina* for being patient and for always supporting me when I needed your help.

I want to express my gratitude also to all the Italian friends, both the ones in Sansepolcro and the ones I met during my University studies. All of you have been always helpful and supportive.

A great thank goes to all the members of the *Italian Catholic Choir of Copenhagen*. The time spent with you has given me more energy to carry on my Ph.D. project.

Lastly, my deepest gratitude goes to my family, in particular to my parents *Doriana* and *Marco*. From the beginning of my studies, you have always believed in me and you have always sustained me with love. I will be grateful to you forever.

Abstract

Parkinson's disease (PD) is a common neurodegenerative disease and is diagnosed in presence of motor symptoms, which appear when the neurodegeneration has affected a large part of the brain. No treatment is currently available to slow down or stop the neurodegeneration. Rapid eye movement (REM) sleep behavior disorder (RBD) is a sleep disorder characterized by abnormal increase in muscular activity during REM sleep and dream enactment. RBD is the strongest early biomarker of PD. Evidence show that RBD might be preceded by a prodromal phase, where minor REM behavioral events can be appreciated but not enough abnormal muscular tone is seen to diagnose RBD. Currently, identification of RBD and prodromal RBD is based on visual inspection of video, electroencephalography (EEG), electrooculography (EOG) and electromyography (EMG) recorded during sleep. Such a process is slow and not objective.

This thesis, divided into three parts, presents new data-driven automated methods for fast and objective identification and characterization of RBD and prodromal RBD.

First, a comparison of the currently available automated methods for RBD detection was performed. The results showed that none of them could be considered *the optimal* one, due to varying performances when they were applied to different patient groups.

Second, a new data-driven method for RBD identification, based on machine learning techniques applied to EMG signals, was developed. The method showed higher performances for identifying RBD patients than previously developed methods (accuracy and specificity over 80% and sensitivity over 70%). Moreover, it was found that muscular activity in non-REM sleep contributed for more accurate RBD identification. When the new method was applied to data recorded in another clinic, it identified RBD patients with similar performances, thus proving its robustness.

Finally, a data-driven method based on machine learning applied to EEG and EOG signals was developed to identify RBD and prodromal RBD in PD patients. RBD was identified with over 80% accuracy, sensitivity and specificity. Moreover, the algorithm could identify PD patients with prodromal RBD with accuracy and specificity over 80%. It was also found that micro-sleep instability could be a biomarker for RBD and prodromal RBD in PD patients. These findings are potentially applicable to patients without overt PD.

In conclusion, this thesis proposes new, fast and automatic methods and biomarkers to identify patients with RBD and prodromal RBD. Compared to current visual-based methods, these algorithms have the potential to be used to identify patients in early stages of neurodegeneration objectively, consistently and significantly faster. Thanks to the objective identification, such patients could constitute a homogeneous target for future neuroprotective trials aiming at slowing down or even stopping the ongoing neurodegeneration.

Resumé

Parkinsons sygdom (PD) er en almindelig neurodegenerativ sygdom og diagnosticeres på baggrund af motoriske symptomer, der opstår, når neurodegenerationen har påvirket en stor del af hjernen. Der findes i øjeblikket ingen behandling for at bremse eller stoppe neurodegenerationen. Rapid eye movement (REM) søvns adfærdsforstyrrelse (RBD) er en søvnforstyrrelse, der er kendetegnet ved unormal høj muskulær aktivitet under REM søvn og drømme. RBD er den mest lovende tidlige biomarkør for PD. Det er vist, at derfør RBD muligvis er en prodromal fase, hvor mindre REM-adfærdsbegivenheder kan observeres, men hvor der ikke ses tilstrækkelig motorisk aktivitet til at kunne stille RBD diagnosen. I øjeblikket er identifikation af RBD og prodromal RBD baseret på manuel og visuel inspektion af video, elektroencefalografi (EEG), elektrookulografi (EOG) og elektromyografi (EMG) registreret under søvn. Denne proces er langsom og ikke-objektiv.

Denne afhandling er delt op i tre dele og præsenterer nye datadrevne automatiserede metoder til hurtig og objektiv identifikation og karakterisering af RBD og prodromal RBD.

Første del omhandler en sammenligning af de aktuelt tilgængelige automatiserede metoder til RBD-detektion. Resultaterne viste at ingen af de aktuelle metoder kan betragtes som den overordnede optimale, da de præsterede forskelligt når de blev anvendt til forskellige patientgrupper.

Dernæst blev der udviklet en ny datadreven metode til RBD-identifikation, der var baseret på machine learning og anvendt på EMG-signaler. Metoden viste højere performance til identifikation af RBD-patienter end tidligere udviklede metoder (nøjagtighed og specificitet over 80% og sensitivitet over 70%). Derudover blev det konstateret, at muskelaktivitet i ikke-REM-søvn bidrog til en mere nøjagtig RBD-identifikation. Da den nye metode blev anvendt på data opsamlet fra en anden klinik, blev RBD-patienterne identificeret med en lignende performance, hvilket beviser metodens robusthed.

Sidste del af afhandlingen omhandler en datadreven metode til identifikation af RBD og prodromal RBD hos PD patienter. Metoden er baseret på machine learning anvendt på EEG og EOG-signaler. RBD blev identificeret med over 80% nøjagtighed, sensitivitet og specificitet. Derudover kunne algoritmen identificere PD-patienter med prodromal RBD med nøjagtighed og specificitet over 80%. Under udviklingen af metoden fremkom det vigtige fund at ustabilitet i mikrosøvn kan være en biomarkør for RBD og prodromal RBD hos PD patienter. Dette kan potentielt også være en biomarkør for patienter der endnu ikke har PD diagnosen.

Afslutningsvis foreslår denne afhandling nye, hurtige og automatiske metoder og biomarkører, der kan identificere patienter med RBD og prodromal RBD. De udviklede algoritmer har potentialet til at blive brugt til at identificere patienter i tidlige stadier af nerodegenerationen objektivt, mere præcist og hurtigere end nuværende visuelle metoder. Disse patienter kan udgøre et homogent mål

for fremtidige kliniske forsøg der har til formål at bremse eller endda stoppe den igangværende neurodegeneration.

List of Figures

List of Tables

Acronyms

AASM American Academy for Sleep Medicine.

AD Alzheimer's disease.

AHI Apnea/hypopnea index.

BF Basal forebrain.

CNN Convolutional neural network.

CV Cross-validation.

CWT Continuous wavelet transform.

DCSM Danish Center for Sleep Medicine.

DLB Dementia with Lewy bodies.

DMH Dorsomedial hypothalamic nucleus.

DRN Dorsal raphe nucleus.

DS Dempster-Shafer.

DT Decision tree.

DWT Discrete wavelet transform.

ECG Electrocardiography.

EEG Electroencephalography.

EMG Electromyography.

EOG Electrooculography.

FDS Flexor digitorum superficialis.

FRI Frandsen index.

FU Follow-up.

GP Gaussian process.

HC Healthy control.

ICSD International classification of sleep disorders.

iRBD Isolated/idiopathic RBD.

iRSWA Isolated RSWA.

KDE Kernel density estimator.

KEI Kempfner index.

LB Lewy bodies.

LC Locus coeruleus.

LDT Laterodorsal tegmental nucleus.

LHA Lateral hypothalamic area.

LM Limb movement.

lMAI Long muscle activity index.

LPT Lateral pontine tegmentum.

MA Muscular activity.

MAF Muscular activity feature.

MnPO Median preoptic nucleus.

MSA Multiple system atrophy.

N1 NREM 1.

N2 NREM 2.

N3 NREM 3.

NDD Neurodegenerative disease.

nPLM Non PLM.

NREM Non-REM.

PB Parabrachial nucleus.

PC Precoeruleus nucleus.

PD Parkinson's disease.

PD+RBD PD patient(s) with RBD.

PD+RBE PD patient(s) with RBEs.

PDD PD dementia.

PDnonRBD PD patient(s) without RBD.

PLM Periodic limb movements.

PLMD PLM disorder.

PLMS PLM during sleep.

POA Preoptic area.

PPT Pedunculopontine tegmental nucleus.

PSG Polysomnography.

RAI REM atonia index.

RBD REM sleep behavior disorder.

RBE REM behavioral event.

RBF Radial basis function.

REM Rapid eye movement.

RF Random forest.

RSWA REM sleep without atonia.

SCN Suprachiasmatic nucleus.

SLD Sublaterodorsal nucleus.

sMAI Short muscle activity index.

STREAM Supra-threshold REM activity metric.

SVM Support vector machine.

TIB Time in bed.

TIBL Anterior tibialis left.

TIBR Anterior tibialis right.

TMN Tuberomammillary nucleus.

TST Total sleep time.

v-PSG Video PSG.

vlPAG Ventrolateral periaqueductal gray.

VLPO Ventrolateral preoptic nucleus.

vM Ventral medulla.

vPAG Ventral periaqueductal gray.

W Wakefulness.

WT Wavelet transform.

Chapter 1

Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disease (NDD) after Alzheimer's disease (AD) and it has a prevalence of 3% in the elderly over 80 years old [1]. Currently, no drugs or neuroprotective agents are available to cure PD [2]. Because of the increasing number of elderly worldwide, the global burden of PD has more than doubled in the last three decades, thus making PD one of the leading sources of disability in the world [3].

Currently, no objective tests are available to diagnose PD, but the diagnosis is based on the presence of motor symptoms, including bradikynesia, resting tremor and/or rigidity [4, 5]. The manifestation of motor symptoms is due to the loss of at least 70% of the dopaminergic neurons located in the substantia nigra [6] and evidence has shown that the neurodegeneration is already in an advanced stage at the time of motor manifestations [7, 8].

In the last decades, research studies have shown that PD has a relatively long pre-motor phase in which some autonomic, olfactory, sleep, cognitive and neuropsychiatric abnormalities might be observed [9]. Among these, a sleep disorder known as rapid eye movement (REM) sleep behavior disorder (RBD), in its idiopathic/isolated form (iRBD), has proven to be by far *the strongest* biomarker of later development of PD [10, 11, 12]. RBD is a disease characterized by abnormal increase in muscular activity during REM sleep (REM sleep without atonia - RSWA) and by dream enactment [10]. To diagnose RBD, a video-polysomnography (v-PSG) is required and this comprises the simultaneous recording of video and electrophysiological signals, including electroencephalography (EEG), electrooculography (EOG), electromyography (EMG), electrocardiography (ECG) and respiratory signals. Currently, data recorded during v-PSG are visually analyzed and scored manually by following non-uniquely defined international rules [13]. Because of this, diagnosis of RBD is not only time-consuming, but also prone to subjectiveness.

Recent studies have also found that RBD in PD patients might be preceded by a stage where some minor abnormal behaviors are present in REM sleep, but no abnormal muscular activity is seen [14, 15]. This stage is referred as *prodromal RBD*. Despite this finding has been reported in PD patients only, it supports the idea that neurodegeneration is a continuous process and has the potential to enable the identification of neurodegeneration in its very early stages in the future. Also in this case, prodromal RBD can be identified only by visual time-consuming and subjective analysis.

Automatic algorithms applied to signals recorded during v-PSG can dramatically help to provide fast, precise and objective identification of RBD and prodromal RBD. Moreover, automatic methods have the potential to reveal hidden signal patterns and identify new biomarkers, thus providing a better characterization of the diseases under investigation. Fig. 1.1 shows an overview of some

possible approaches to design automatic algorithms for identification of patients with RBD and prodromal RBD from v-PSG. The approaches that were investigated in this thesis are highlighted.

The fast and objective identification and characterization of patients with early neurodegeneration has the potential to identify a homogeneous patient group which can be the target for administration of neuroprotective agents, with the aim of slowing down or even stopping the evolving neurodegeneration.

The research presented in this thesis focuses only on identification of RBD as an early neurodegeneration by using electrophysiological signals recorded during sleep. It has to be acknowledged that many other research groups do focus on identification of early neurodegeneration with other biomarkers, including clinical observations, brain imaging, and blood/cerebrospinal fluid/tissue biomarkers [16, 17].

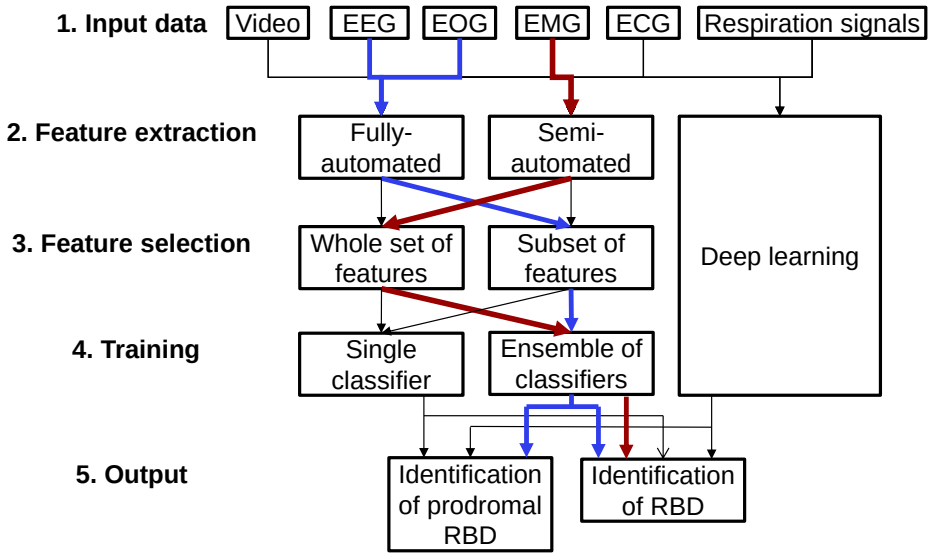


Figure 1.1: Possible approaches for identification of patients with RBD and prodromal RBD with automatic analysis of data recorded during v-PSG. The data acquisition is followed by feature extraction, which can be either fully or semi-automated (meaning that input from humans are required in the latter case). The whole set of features or a subset of it is then given as input to train a machine learning system, which can consist of a single classifier or an ensemble of classifiers. Deep learning techniques combine the steps of feature extraction, feature selection and training. The output corresponds to the identification of a patient with prodromal RBD or with RBD. The red line represents the path followed in current thesis to identify patients with RBD. The blue line represents instead the path followed in current thesis to identify RBD and prodromal RBD in PD patients.

1.1 Problem statement

Prodromal RBD and RBD are considered early stages of PD and identification of patients suffering from these disorders is of utmost importance to define target groups for neuroprotective trials aiming at slowing down or even stopping the ongoing neurodegeneration. Current methods for their identification rely on visual and manual analysis of video and electrophysiological signals recorded during sleep, making diagnosis of RBD and prodromal RBD time-consuming and subjective. Automated methods have the potential to identify patients with RBD and prodromal RBD faster

and objectively, helping to characterize these target groups more clearly for future neuroprotective trials.

1.2 Thesis research questions and objectives

Taking into account the motivation and the problem statement described above, the following thesis research questions and objectives were defined:

- ***Thesis research question 1:*** Some automatic methods have already been developed to identify patients with RBD [18, 19, 20, 21, 22, 23, 24, 25, 26] based on the automatically detected level of RSWA. However, a formal comparison of them is lacking. The following thesis research question was therefore defined:
  Which one of the available automated methods is the optimal one to identify patients with RBD?
  Thesis objective 1: To perform an objective comparison of the currently available automated methods to identify the one(s) achieving the highest performances for identifying patients with RBD.

- ***Thesis research question 2:*** Most of the previously described automated methods are based on traditional programming techniques. This means that they implement a set of human-defined rules for RBD identification. Moreover, their robustness to inter-clinical variability has rarely been proven. Machine learning techniques, due to their data-driven nature, are potentially superior to traditional programming approaches [27], therefore the following research question was defined:
  Can a new automated method, based on machine learning techniques applied to EMG signals, overcome the currently available ones to identify RBD? Is such a method robust to inter-clinical variability?
  Thesis objective 2: To develop and validate on different databases an automated data-driven method that identifies abnormal muscular activity using machine learning techniques applied to EMG signals. Such method should be robust and should diagnose RBD with higher performances compared to previously developed automated methods.

- ***Thesis research question 3:*** The progression from prodromal RBD to RBD is a recent finding and has been proven in PD patients [14], but has the potential to be valid also in subjects without overt PD [15]. As prodromal RBD has been shown not to be characterized by RSWA, an objective method for its identification and characterization could be based on EEG and EOG signals. From this, the following research question was defined:
  Can an automated machine learning-based method, built on EEG and EOG signals, identify and characterize RBD and prodromal RBD in PD patients?
  Thesis objective 3: To develop an automated data-driven method that, based on EEG and EOG signals, can identify and characterize RBD and prodromal RBD in PD patients.

1.3 Thesis outline and contributions

The research content of this thesis is structured in three parts, each of them attempting to answer one of the three thesis research questions. These parts are articulated in four Chapters.

Besides these chapters, the thesis contains this introduction, a clinical background, a conclusive Chapter and a final Chapter outlining possible ways for future work.

Chapter 2 provides a clinical background on PD, sleep and RBD to help the understanding of the rest of the thesis.

Chapter 3 describes the formal comparison of the currently available automated methods for identifying RBD. This Chapter addresses thesis objective 1.

Chapter 4 presents a preliminary study for the development of a new machine learning based method to identify RBD. In particular, this study focuses on the identification of limb movements during sleep. This Chapter partly addresses thesis objective 2.

Chapter 5 presents the development of a new data-driven machine learning method to identify RBD based on abnormal muscular activity during sleep, and its validation on different databases. This Chapter addresses thesis objective 2.

Chapter 6 presents the development and validation of a fully automated machine learning method that, based on EEG and EOG signals, can identify and characterize RBD and prodromal RBD in PD patients. This Chapter addresses thesis objective 3.

Chapter 7 concludes the thesis by summing up the main results.

Chapter 8 outlines future research perspectives in the field of automatic identification and characterization of patients with RBD and prodromal RBD.

The scientific content of this thesis is mainly based on four journal papers (three accepted and one submitted) and two accepted conference papers, listed here below:

Journal papers:

- **M. Cesari**, J. A. E. Christensen, L. Kempfner, A. N. Olesen, G. Mayer, K. Kesper, W. H. Oertel, F. Sixel-Döring, C. Trenkwalder, H. B. D. Sorensen, and P. Jennum, "Comparison of computerized methods for rapid eye movement sleep without atonia detection", *Sleep*, vol. 41, no. 10, pp. zsy133, 2018.

- **M. Cesari**, J. A. E. Christensen, F. Sixel-Döring, C. Trenkwalder, G. Mayer, W. H. Oertel, P. Jennum, and H. B. D. Sorensen, "Validation of a new data-driven automated algorithm for muscular activity detection in REM sleep behavior disorder", *Journal of Neuroscience Methods*, vol. 312, pp. 53-64, 2019.

- **M. Cesari**, J. A. E. Christensen, H. B. D. Sorensen, P. Jennum, B. Mollenhauer, M.-L. Muntean, C. Trenkwalder, and F. Sixel-Döring, "External validation of a data-driven algorithm for muscular activity identification during sleep", *Journal of Sleep Research*, In press, 2019.

- **M. Cesari**, J. A. E. Christensen, M.-L. Muntean, B. Mollenhauer, F. Sixel-Döring, H. B. D. Sorensen, C. Trenkwalder, and P. Jennum, "A data-driven system to identify REM sleep behavior disorder and to predict its progression from the prodromal stage in Parkinson's disease", Submitted, 2019.

Conference papers:

- **M. Cesari**, J. A. E. Christensen, P. Jennum, and H. B. D. Sorensen, "Probabilistic data-driven method for limb movement detection during sleep", *Proceedings of the 40^{th} International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2018, pp. 163-166, 2018.

- **M. Cesari**, J. A. E. Christensen, F. Sixel-Döring, M.-L. Muntean, B. Mollenhauer, C. Trenkwalder, P. Jennum, and H. B. D. Sorensen, "A clinically applicable interactive micro and macro-sleep staging algorithm for elderly and patients with neurodegeneration", *Proceedings of the 41st International Conference of the IEEE Engineering in Medicine and Biology Society*, In Press, 2019.

In addition, during the Ph.D. project the following publications were co-authored:

- **M. Cesari** and P. Jennum, "Selective polysomnographic findings in REM sleep behavior disorder (RBD) and Parkinson's disease", *Rapid-eye-movement Sleep Behavior Disorder, Springer, Cham*, pp. 271-279, 2018.
- A. N. Olesen, **M. Cesari**, J. A. E. Christensen, H. B. D. Sorensen, E. Mignot, and P. Jennum, "A comparative study of methods for automatic detection of rapid eye movement abnormal muscular activity in narcolepsy", *Sleep Medicine*, vol. 44, pp. 97-105, 2018.
- A. B. Klok, J. Edin, **M. Cesari**, A. N. Olesen, P. Jennum, and H. B. D. Sorensen, "A new fully automated random-forest algorithm for sleep staging", *Proceedings of the 40th International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2018, pp. 4920-4923, 2018.

Chapter 2

Clinical background

Chapter main objectives: α-synucleinopathies are the second most common neurodegenerative diseases in elderly population and pharmaceutical research is focusing on developing effective neuroprotective treatments. Idiopathic/isolated rapid eye movement (REM) sleep behavior disorder (RBD) is the strongest early biomarker of α-synucleinopathies. Therefore, its identification plays an important role in the perspective of slowing down or even blocking α-synucleinopathies in their early stages.

This chapter aims to briefly introduce the pathophysiology of Parkinsonian neurodegenerative diseases, the structure and physiology of normal sleep and how this is disrupted in patients suffering from RBD. Finally, the current challenges in identification and characterization of RBD are presented.

2.1 Parkinson's disease and α-synucleinopathies

Neurodegenerative diseases (NDDs) are a common and growing cause of mortality, morbidity and cognitive impairment in the elderly population [1]. Among NDDs, Alzheimer's disease (AD) is the most common one, affecting up to 50% of the people older than 85 years old [28]. Following AD, Parkinson's disease (PD) is the second most common NDD with a prevalence of 0.3% in the general population, 1% in the people over 60 years old, and 3% in the elderly over 80 years old [1]. PD belongs to the group of NDDs known as α-synucleinopathies, which include also dementia with Lewy bodies (DLB), PD dementia (PDD) and multiple system atrophy (MSA) [1]. Clinically, α-synucleinopathies are characterized by chronic and progressive decline in motor, cognitive, behavioural, and autonomic functions [29]. From a pathological point of view, they are defined by abnormal neural depositions of Lewy bodies (LB), which are aggregates of a protein known as α-synuclein, of which the role has not been cleared yet [30].

PD is the most common among the α-synucleinopathies and its core pathologic feature is the loss of dopaminergic neurons in the substantia nigra pars compacta due to LB depositions [1]. No objective tests for diagnosing PD are currently available, and the diagnosis is based on clinical motor symptoms, including bradykinesia, in combination with either rest tremor, rigidity, or both [4, 5]. For giving PD diagnosis, these clinical factors should be accompanied by additional supporting criteria defined by international standards [4, 5]. It has been observed that the typical motor symptoms of PD occur when at least 70% of the dopaminergic neurons in the substantia nigra have been lost [6].

Treatment includes mainly dopaminergic drugs and deep brain stimulation devices which, however, are purely symptomatic and aim to reduce the burden of the motor symptoms of

Parkinsonism [1]. However, pharmaceutical research is ongoing to develop neuroprotective drugs aiming to slow down the neurodegeneration and, potentially, to stop it [31, 32].

2.1.1 Progression of Parkinson's disease from its prodromal stages

Patients suffering from PD do not show motor symptoms only, but are often characterized by autonomic, olfactory, sleep, cognitive and neuropsychiatric abnormalities which can occur either after disease onset, or even before. The evolution and progression of these non-motor symptoms, together with pathophysiological findings, support the so-called *Braak's staging*, a model describing the temporal progression of α-synuclein pathology [7, 8, 9]. According to Braak and colleagues, PD progression follows a brainstem path, starting in the dorsal motor nucleus and progressing to the higher structures of the brain. This progression could be divided into six different stages, which are depicted in Fig. 2.1 [7, 8, 9, 33, 34]:

- Stage 1: The dorsal motor nucleus of the vagus nerve and the anterior olfactory nucleus complex are affected by the α-synuclein pathology. This stage is thought to correspond to the development of gastrointestinal and cardiovascular dysfunctions and hyposmia.
- Stage 2: The neurodegeneration spreads to the sublaterodorsal and precoeruleus region, to the magocellular reticular formation, the raphe nucleus and the locus coeruleus. This is considered to be related to the development of rapid eye movement (REM) sleep behavior disorder (RBD), obesity and depression.
- Stage 3: The substantia nigra, pedunculopontine nucleus and amygdala are involved and the typical clinical features of unilateral tremor, rigidity, akinesia and postural instability are thought to appear.
- Stage 4: The neurodegeneration spreads to the temporal mesocortex, which is likely to cause bilateral manifestations of the motor symptoms.
- Stage 5: The α-synuclein pathology involves some structures of the neocortex. Such an involvement could be related to the development of autonomic dysfunction, gait disturbances and impaired balance.
- Stage 6: Finally, the neurodegeneration spreads to the entire neocortex, which may cause cognitive decline and development of dementia.

Stages 1 and 2 are considered the stages of the pre-motor phase of PD (also known as *prodromal PD*), as the key motor signs of Parkinsonism have not appeared yet. The challenge for investigators and scientists is to identify patients with high accuracy and precision in the early stages of neurodegeneration, before the appearance of Parkinsonism. Such identification will allow to extend the ongoing (and future) neuroprotective trials, with the aim of slowing down or even stopping the α-synuclein pathology in its earliest stages [9].

Braak's staging is considered the most robust model of progression of neurodegeneration in α-synucleinopathies [36]. However, PD is a highly heterogeneous disease, with respect to early biomarkers as well as later clinical outcomes [37]. Recent studies have pointed out that one of the possible causes of the heterogeneity is that the progression of α-synuclein pathology does not necessary follow the brainstem route that Braak and colleagues have described. Currently, it is thought that PD may follow three different routes, depicted in Fig. 2.2 [37]. Besides the brainstem

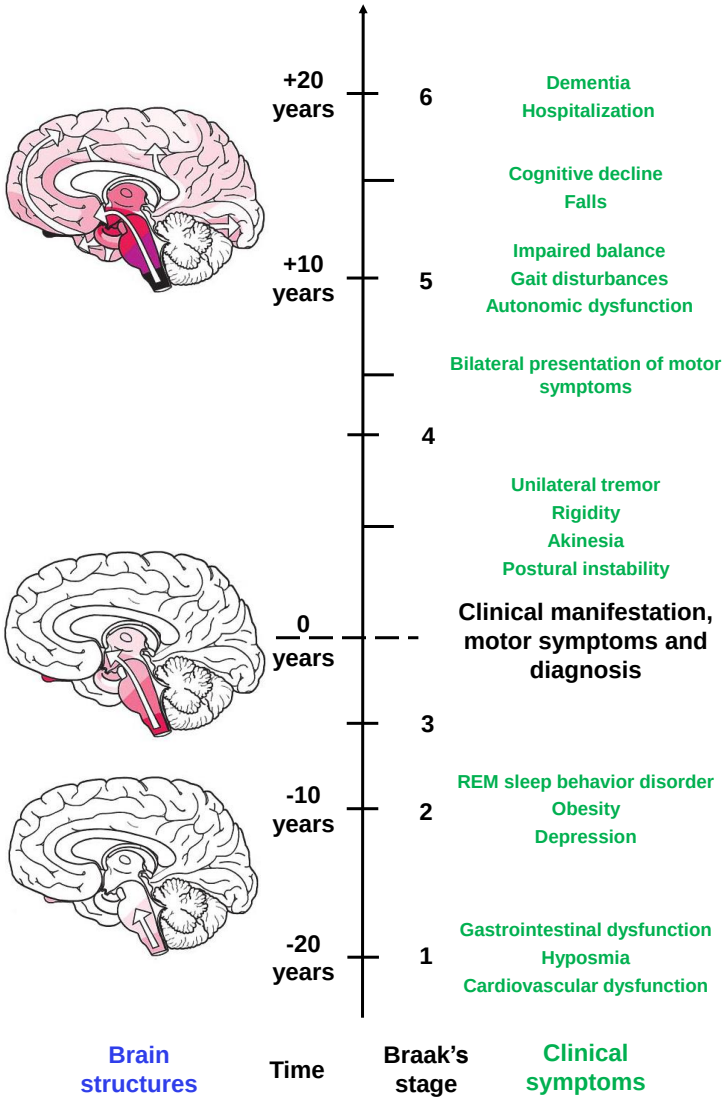
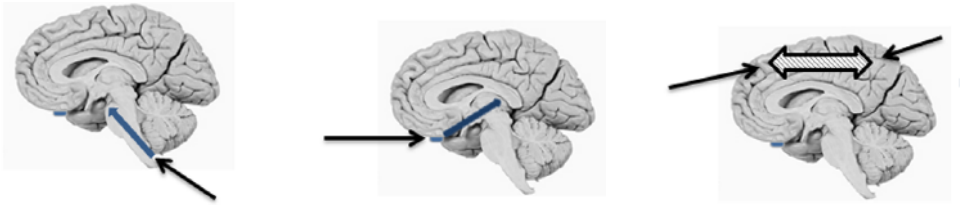


Figure 2.1: Braak's hypothesis of PD development. Adapted from [33] with permission from Elsevier. Adapted from [35] with permission from Springer.

route described by Braak and colleagues [7, 8] (Fig. 2.2a), the olfactory to limbic route [38] (Fig. 2.2b), and the cognitive route [39] (Fig. 2.2c) have been proposed. These different routes can explain, at least in part, the heterogeneous picture of PD and support the idea that different PD phenotypes exist [37, 40].

2.1.2 Isolated/idiopathic RBD as the strongest biomarker of prodromal PD

Despite different possible routes of spreading of α-synuclein, the brainstem route is thought to be the most common [36]. Following this route, RBD in its idiopathic/isolated form (iRBD) is one of the non-motor features expected to appear in prodromal stages of PD. RBD is a sleep disorder characterized by dream enactment and loss of muscular atonia during REM sleep, which has been first described in 1986 [10]. Among other biomarkers, iRBD is considered by far *the strongest* biomarker of prodromal stage of α-synucleinopathies [41] (including PD, DLB and MSA [15, 42]). This has been confirmed by the following findings. First, several follow-up studies have



(a) Brainstem route. (b) Limbic to olfactory route. (c) Cognitive route.

Figure 2.2: Possible routes of spread of α-synuclein pathology. Adapted from [37].

shown that iRBD patients, especially men > 50 years old, convert to a α-synucleinopathy with a mean conversion time of around 10 years from RBD onset [11, 12, 43, 44]. A recent multicenter study has observed that the overall conversion rate from iRBD to a α-synucleinopathy was 6.3% per year and that 73.5% of the patients converted after 12 years from RBD onset [12]. Second, post-mortem neurophysiological studies have shown deposition of α-synuclein in the brainstem in iRBD patients [45].

Evidence suggests that correct identification of patients with iRBD has the potential to identify a target group of patients with early neurodegeneration. This thesis will focus on this sleep disorder, with the aim of providing efficient tools for identifying and characterizing this target patient group. To better understand this sleep disorder, the following sections will describe the physiology of sleep more in detail, in particular, how it is analyzed and the clinical appearance and pathophysiology of RBD.

2.2 Sleep structure and physiology

Significant amount of research in neurology and psychiatry has been focusing on trying to understand the neuroanatomy and physiology of sleep [46]. However, our current knowledge still lacks the understanding of many mechanisms regulating sleep, and the overall sleep purpose has not been fully understood [46].

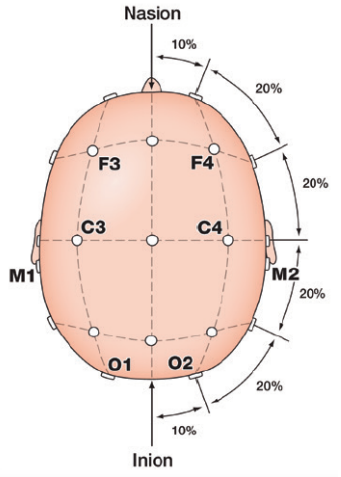
2.2.1 Electrophysiological structure of sleep

The clinical tool used to analyze sleep is polysomnography (PSG) which consists of the concomitant recording of electrophysiological signals including electroencephalography (EEG), electrooculography (EOG), chin and leg electromyography (EMG), airflow signals, respiratory effort signals, oxygen saturation and electrocardiography (ECG). Table 2.1 provides an overview of the technical requirements for EEG, EOG and EMG in a PSG recording, according to the most recent version of the American Academy for Sleep Medicine (AASM) scoring manual [13] and Fig. 2.3 shows the placement of the electrodes. The channels M1 and M2 are usually referred also as A1 and A2, respectively, and the channels E1 and E2 as LOC (left outer canthus) and ROC (right outer canthus), respectively.

For identification of some specific pathologies, PSG can be accompanied by video recording (video-PSG, v-PSG). According to AASM, data recorded during a PSG are arbitrarily divided into non-overlapping 30-s sleep epochs which are manually scored either as wakefulness (W), REM sleep or one of the three stages of non-REM (NREM) sleep (N1, N2, N3) [13]. The manual scoring of sleep stages relies on the identification of specific patterns in the EEG, EOG and chin EMG

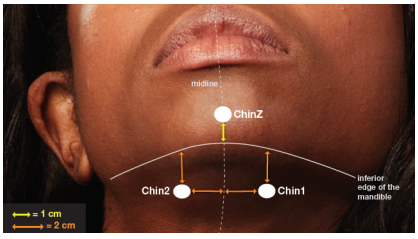
Signals	Channels	Min Fs	Filter
EEG	Required: F4-M1, C4-M1, O2-M1 (Fig. 2.3a). Backup: C3-M2, F3-M2, O1-M2 (Fig. 2.3a).	200 Hz	0.3-35 Hz
EOG	Required: E1-M2 and E2-M2 (accepted E2-M1) (Figs. 2.3a and 2.3b).	200 Hz	0.3-35 Hz
EMG	Chin EMG (required): three electrodes (Fig. 2.3c). Anterior tibialis EMG (required): two electrodes on both legs (Fig. 2.3d). Flexor digitorum superficialis EMG (optional): two electrodes on both arms (Fig. 2.3e).	200 Hz	10-100 Hz

Table 2.1: Technical specification for recording of EEG, EOG and EMG in PSG according to AASM [13]. *Min Fs* represents the minimum sampling frequency and *filter* the usual band-pass filters applied to the signals for their analysis.

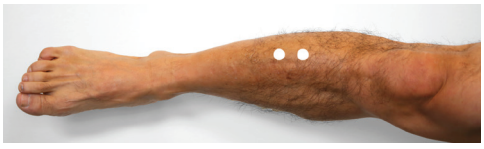


(a) EEG derivations.

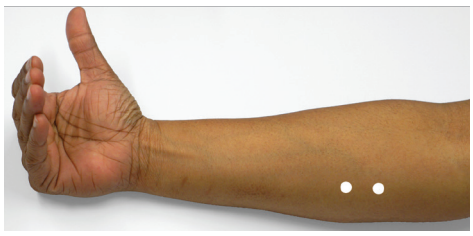
(b) EOG channels.



(c) Chin EMG channels.



(d) Anterior tibialis EMG channels.



(e) Flexor digitorum superficialis EMG channels.

Figure 2.3: Locations of the EEG, EOG and EMG electrodes during PSG according to AASM [13]. All figures are taken from the AASM manual [13] with permission from AASM publisher.

signals. Fig. 2.4 shows an example of the typical patterns seen in the five sleep stages. The electrophysiological patterns and physiological aim of each stage are [13, 47]:

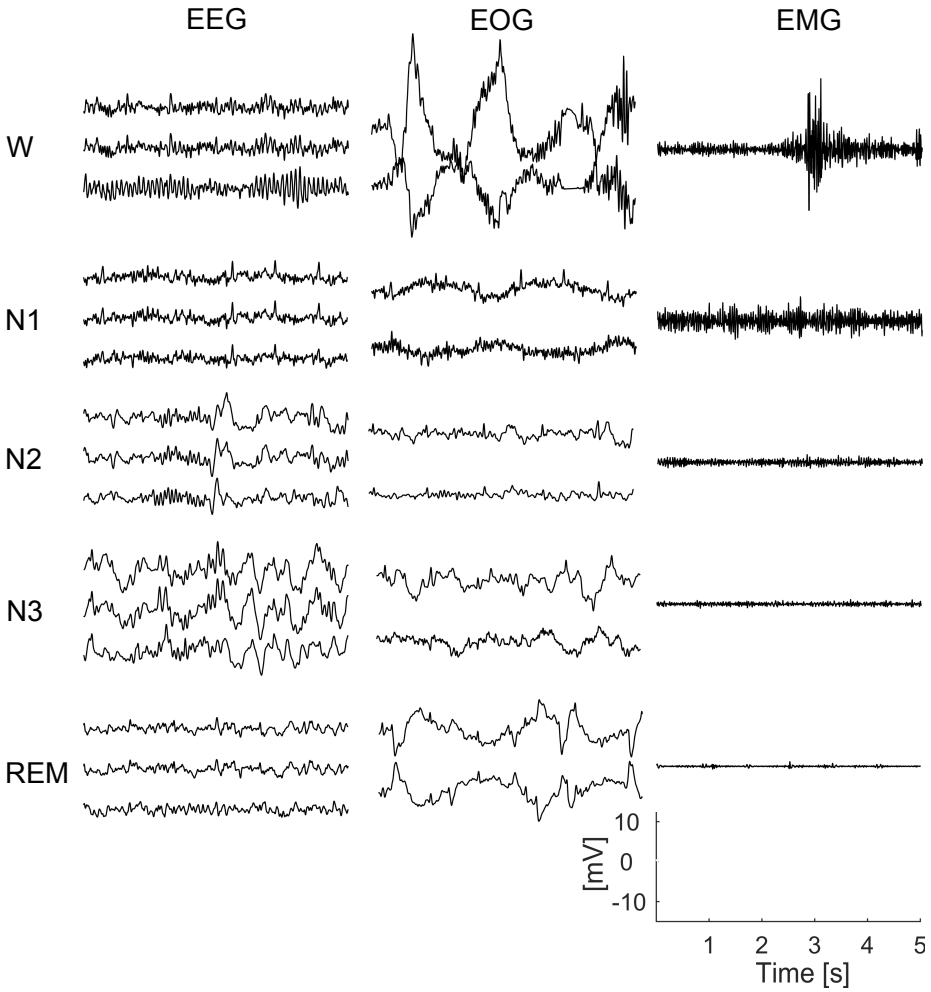


Figure 2.4: Typical patterns characterizing the five sleep stages. The three EEG derivations are F4-A1, C4-A1 and O2-A1 (Fig. 2.3a), and the two EOG channels are ROC-A1 and LOC-A2 from top to bottom (Fig. 2.3b). The EMG channel is the chin EMG derived as the difference between chin1 and chin2 signals (Fig. 2.3c).

- W: During wakefulness the EEG is typical low in amplitude and alpha rhythm (8-13 Hz) is dominant over the occipital region with closed eyes. EOG signals are characterized by eye blinks, rapid or reading eye movements. The chin EMG is typically high in tone with variable amplitude.
- N1: This sleep stage has a transitional role from wakefulness to deeper sleep and constitutes 2-5% of the total sleep time. The EEG is typically characterized by a reduction of alpha waves, low-amplitude waves with theta rhythm (4-7 Hz) and vertex sharp waves (sharply contoured waves with duration lower than 0.5 s). EOG signals present slow eye movements and there is a reduction of tone in the chin EMG compared to W.
- N2: Essential restorative functions and memory consolidation are thought to occur during N2 sleep, which constitutes approximately 45-55% of the total sleep time, and usually follows N1 sleep. The EEG is characterized by mixed frequencies, K-complexes (evident negative

waves followed directly by a positive component and lasting more than 0.5 s, more evident in frontal derivations) and sleep spindles (trains of sinusoidal waves with frequency 11-16 Hz and longer than 0.5 s, which are more evident in frontal and central derivations). The EOG signals usually show slow eye movements or absence of eye movements and the chin EMG has lower amplitude than in N1 sleep.

- N3: This stage corresponds to deep sleep and usually follows N2 sleep, constituting about 13-25% of the total sleep time. The EEG is characterized by high-amplitude slow-waves (0.5-2 Hz), typically more evident in frontal derivations. The EOG shows no eye movements, but artifacts from the frontal EEG might be present. Chin EMG usually exhibits a lower tone than N2 sleep.

- REM: This is the stage associated with most of night dreaming and constitutes 20-25% of the total sleep time. The EEG signals are characterized by low-amplitude, mixed-frequency activity without K-complexes or sleep spindles. Sometimes, sawtooth waves (trains of sharply contoured or triangular waves at 2-6 Hz) are present. Rapid eye movements characterize the EOG signals and the chin EMG signal has lower activity than any other sleep stages (referred as *atonia*). Short transient of muscular activity (< 0.25 s) might be present.

Typically, during one night sleep a healthy subject experiences 3 to 5 sleep cycles, consisting of the sequence of N1, N2, N3 and REM sleep. The sequence of scored 30-s sleep epochs constitutes a *hypnogram*, which is a compact representation of the overall sleep structure. Fig. 2.5 shows an example of a hypnogram of a healthy control subject, where the sequence of sleep cycles can be appreciated. In a healthy young subject, it is normally seen that N3 sleep is more present at the beginning of the night, while REM sleep more is more present towards the end [47].

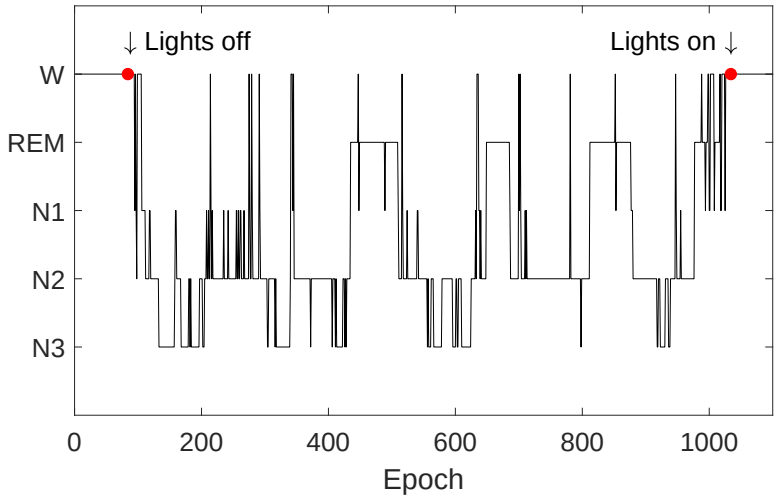


Figure 2.5: Hypnogram of a healthy male subject of 40 years. "Lights off" and "Lights on" labels are subject-reported time-points.

2.2.2 Neural regulation of sleep

Sleep is regulated by a number of neural structures interacting between them in a complex manner, suggesting that sleep is a globally coordinated but locally regulated phenomenon [46, 48, 49]. Fig. 2.6 shows the brain areas that are involved in wakefulness and sleep regulation located in the

brainstem and midbrain areas. Current research and evidence suggest the presence of two switch systems that control transitions i) between wakefulness and sleep, and ii) between NREM and REM sleep, respectively [49].

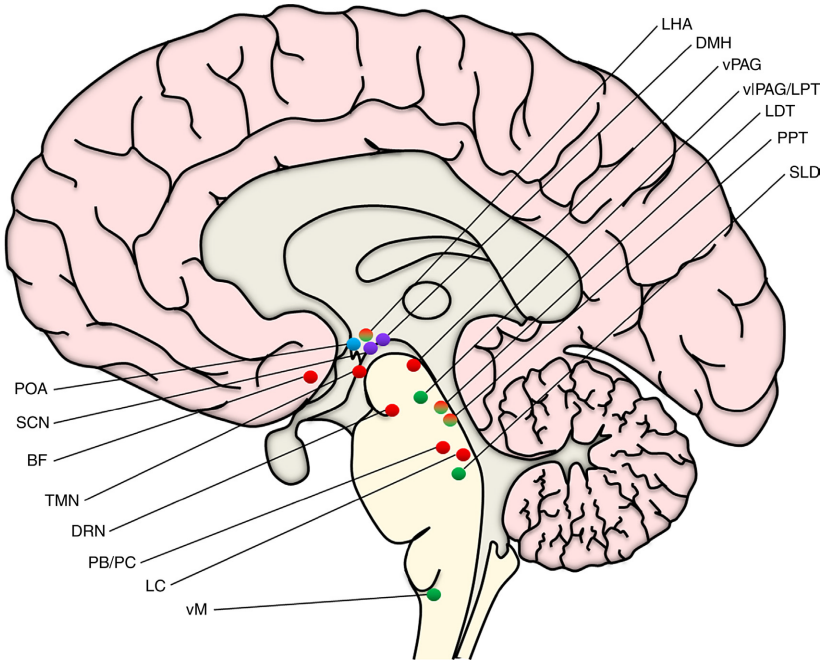


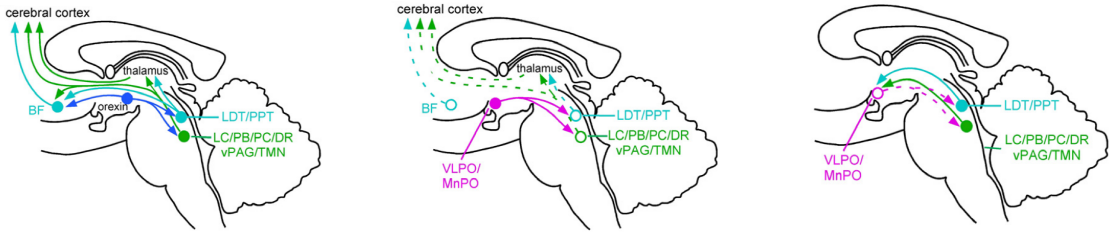
Figure 2.6: Location of the brain structures involved in the wake/sleep regulation. Colors indicate the predominant role played by each structure: red for wake and arousal, blue for sleep, green for REM sleep, purple for circadian regulation, and multicolored for mixed activity. BF: basal forebrain; DMH: dorsomedial hypothalamic nucleus; DRN: dorsal raphe nucleus; LC: locus coeruleus; LDT: laterodorsal tegmental nucleus; LHA: lateral hypothalamic area; PB/PC: parabrachial/precoeruleus nucleus; POA: preoptic area (containing ventrolateral and median preoptic nuclei - VLPO and MnPO); PPT: pedunculopontine tegmental nucleus; SCN: suprachiasmatic nucleus; SLD: sublaterodorsal nucleus; TMN: tuberomammillary nucleus; vlPAG/LPT: ventrolateral periaqueductal gray/lateral pontine tegmentum; vM: ventral medulla; vPAG: ventral periaqueductal gray. Reprinted from [46] with permission from Elsevier.

Wake-sleep switch

Fig. 2.7 shows the structures that are involved in the wake-sleep switch. Two pathways promoting wakefulness can be identified (Fig 2.7a). More specifically, the dorsal cholinergic pathway originates in the laterodorsal and pedunculopontine tegmental nuclei (LDT and PPT) and proceeds up to the thalamus, while the ventral monoaminergic pathway projects from the locus coeruleus (LC), parabrachial nucleus (PB), precoeruleus area (PC), dorsal raphe nucleus (DRN), ventral peroaqueductal gray (vPAG), and tuberomammillary nucleus (TMN) to the hypothalamus, basal forebrain (BF) and cerebral cortex. Orexin neurons located in the lateral hypothalamus reinforce the wakefulness promotion by directly exciting both pathways as well as BF and the cerebral cortex [46, 48, 49].

The main pathway promoting sleep originates in the ventrolateral (VLPO) and median (MnPO) preoptic nuclei, which are located in the preoptic area (POA). These nuclei project to the wake-promoting pathways and are able to inhibit them (Fig. 2.7b), but at the same time the two wake-promoting systems are also able to inhibit the VLPO and MnPO. Overall, these mutual

inhibition pathways are the basis of the wake-sleep switch, which is able to generate fast and complete transitions between these two states (Fig. 2.7c) [46, 48, 49].



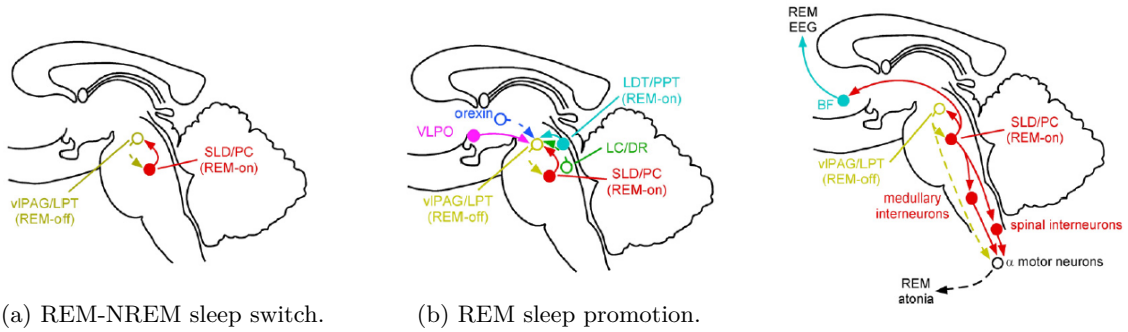
(a) Wake promotion. (b) Sleep promotion. (c) Wake-sleep switch.

Figure 2.7: Structures and connections involved in the wake-sleep switch. (a) Two wake-promoting pathways can be recognized. The dorsal one (aqua) originates in the LDT/PPT and projects to the thalamus, and the ventral one (dark green) originates in the LC/PB/PC/DRN/vPAG/TMN and projects to the BF, hypothalamus and cerebral cortex. The orexin neurons (blue) reinforce both pathways and excite the cerebral cortex and BF. (b) VLPO and MnPO inhibit the wake-promoting pathways, thus inducing sleep. (c) The wake-promoting structure can also inhibit the sleep-promoting ones, thus creating a wake-sleep switch. Reprinted from [49] with permission from Elsevier.

REM-NREM switch

Fig. 2.8 shows how the REM-NREM sleep switch is organized and regulated. The upper pons contains two populations of neurons that mutually inhibit each other to promote transitions between REM and NREM sleep (Fig. 2.8a). More specifically, REM-off neurons located in the ventrolateral periaqueductal gray (vlPAG) and in the lateral pontine tegmentum (LPT) fire to promote NREM sleep and have an inhibitory effect on the REM-on neurons located in the sublaterodorsal nucleus (SLD) and in the PC, which reciprocally fire to promote REM sleep and inhibit the NREM sleep promoting structures. This mutual inhibition is the basis of the REM-NREM sleep switch, which generates fast and complete transitions between the two sleep states. The switch to REM sleep is additionally regulated by a series of other structures and mechanisms (Fig. 2.8b). In particular, neurons located in the LC and DRN inhibit REM sleep by exciting REM-off neurons and by inhibiting REM-on neurons, and become silent during REM sleep. Opposite effects are produced by the neurons located in the LDT and PPT nuclei. Moreover, orexin neurons inhibit entering REM sleep by exciting REM-off neurons, while VLPO neurons contribute to the initiation of REM sleep by inhibiting the same target. Once REM sleep is entered, the REM-on neurons located in the SLD and PC activate pathways leading to the typical REM EEG waveforms (Fig. 2.4) and activate medullary and spinary interneurons, which have the effect of inhibiting motor neurons, thus causing the typical REM atonia (Figs. 2.8c and 2.4) [46, 49].

This brief overview of the mechanisms involved in sleep regulation documents that sleep is a complex system that is globally coordinated but locally regulated [46]. Many wake-sleep disorders (including insomnia, narcolepsy and other disorders associated with fragmented sleep) have been related to disruption or damage of this system, but knowledge in this field is still limited [50].



(a) REM-NREM sleep switch.

(b) REM sleep promotion.

(c) REM sleep atonia generation.

Figure 2.8: Structures and connections involved in the REM-NREM sleep switch. (a) REM-on (red) and REM-off (gold) neurons have mutually inhibition roles to promote transitions from REM to NREM sleep and vice-versa. (b) A series of neural regulations are involved to promote REM sleep. These include: neurons located in LC and DRN (activating REM-off neurons and inhibiting REM-on neurons), neurons located in LDT and PPT (having the opposite effect), orexin neurons (inhibiting REM sleep by activating REM-off neurons) and neurons located in the VLPO (inhibiting REM-off neurons). (c) During REM sleep, REM-on neurons fire neurons in the BF, generating REM EEG waves. At the same time,they activate medullary and spinal inter-neurons, which inhibit α-motor neurons, thus producing the typical REM atonia. Reprinted from [49] with permission from Elsevier.

2.2.3 Sleep changes related to aging

Sleep patterns and structure change continuously and considerably with age. The main changes include how sleep is initiated and maintained, the percentage of sleep spent in each sleep stage and sleep efficiency (i.e. how well sleep is maintained) [47]. Fig. 2.9 shows the trend of sleep evolution related to aging [51]. Elderly people are characterized by longer sleep latency (the time necessary to fall asleep), shorter overall sleep duration, increased sleep fragmentation (more transitions to wakefulness and lighter sleep stages), increased time spent awake after sleep onset, reduction of time spent in N3 sleep (which mainly affects men [52]), and shorter and fewer sleep cycles [53]. Neurophysiological studies have documented disruption of the structures involved in the wake-sleep regulation, including the orexin neurons and neurons located in the LC [53], which might be the cause of the increased sleep instability. However, changes in circadian regulation, as well as exercise reduction and irregular meal times might also contribute to the sleep disruption seen in elderly [47].

In addition to the changes seen at a macro-sleep level, several changes occur also at a micro-sleep level, including morphological changes in slow waves and sleep spindles during NREM sleep [53]. More specifically, a reduction of amplitude and density of slow wave activity is seen and it is particularly appreciated in the frontal lobe [54]. Moreover, elderly are characterized by a reduction in density and duration of sleep spindles [53].

All the sleep changes related to aging at both macro- and micro-structural level are thought to indicate a natural age-related neurodegeneration process, involving brain areas controlling and regulating sleep [53].

2.3 REM sleep behavior disorder

REM sleep behavior disorder (RBD) is a parasomnia characterized by vocalizations, jerks and motor behaviors during REM sleep, often associated with REM-related dream content [10]. Typical

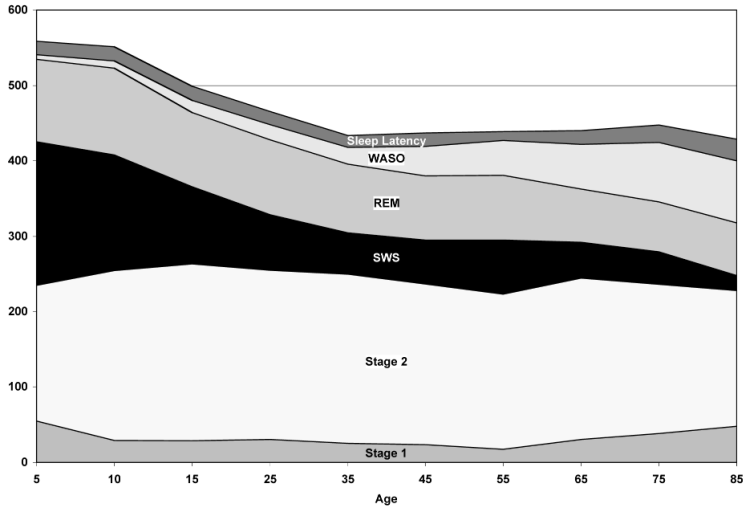


Figure 2.9: Sleep changes in relation to ageing. The graph shows the time of sleep latency, time spent awake after sleep onset (WASO) and the time in different sleep stages (SWS is slow wave sleep, corresponding to N3 sleep). The time (ordinates) is expressed in minutes and the age in years. Reprinted from [51] with permission of Oxford University Press.

clinical signs of RBD range from unnoticed sleep disruption to severe self-injurious behaviors and/or injuries to bed partners [15]. A recent study has found that RBD has a prevalence of 1.06% in the general population [55].

RBD has been categorized into idiopathic (or isolated [15]) RBD (iRBD) and symptomatic (or secondary) RBD. The former is so called because it occurs in the absence of associated comorbidities, and the latter is so called because it is associated with neurodegenerative disorders, medications (in particular antidepressants), narcolepsy type 1 or brainstem lesions [15, 42]. Among neurodegenerative disorders, secondary RBD is associated with α-synucleinopathies, including PD (33-46% of patients [56, 57]), DLB (80% of patients [58]) and MSA (up to 100% of patients [59]).

2.3.1 Diagnosis according to international standards

According to the third edition of the International Classification for Sleep Disorders (ICSD-3), RBD can be diagnosed if the following criteria are met [60]:

- Repeated episodes of sleep related vocalization and/or complex motor behaviors. [1] [2]
- These behaviors are documented by PSG to occur during REM sleep or, based on clinical history of dream enactment, are presumed to occur during REM sleep.
- PSG recording demonstrates REM sleep without atonia (RSWA). [3]
- The disturbance is not better explained by another sleep disorder, mental disorder, medication, or substance use.

[1] This criterion can be fulfilled by observations of repetitive episodes during a single night of v-PSG

[2] The observed vocalizations or behaviors often correlated with simultaneously occurring dream mentation, leading to the frequent report of "acting out one's dreams".

[3] As defined by the guidelines for scoring PSG features of RBD in the AASM manual for the scoring of sleep and associated events [13].

From the ICSD-3 criteria, it emerges that the electrophysiological hallmark of RBD is the lack of atonia during REM sleep. According to AASM manual [13], two types of muscular activity can be identified during REM sleep:

- Tonic EMG activity (Fig. 2.10a): a REM sleep epoch is scored as having tonic EMG activity if at least 50% of the duration has chin EMG amplitude greater than the minimum amplitude demonstrated during NREM sleep.
- Phasic EMG activity (Fig. 2.10b): a REM sleep epoch is scored as having phasic EMG activity if at least five out of ten 3-s mini-epoch contain bursts of transients activity with duration of 0.1-5 seconds and amplitude at least 4 times higher than the background EMG activity.

Finally, the ICSD-3 recommends RBD diagnosis when more than 27% of REM 30-s sleep epochs are identified as having any (i.e. tonic/phasic) chin EMG activity combined with bilateral phasic activity of the flexor digitorum superficialis (FDS) muscles [60]. However, the recording of FDS EMG activity is considered optional in the AASM manual [13].

2.3.2 Pathophysiology of RBD

Boeve et al. first proposed a pathophysiological basis of RBD based on animal models [62]. Fig. 2.11 shows the brain structures involved in RBD pathophysiology. In particular, Fig. 2.11a shows the circuits involved in REM sleep generation and REM sleep atonia: the pre-coeruleus and the SLD nuclei activate two inhibitory pathways (direct and indirect routes) that inhibit the spinal motor neuron, thus generating atonia. In Fig. 2.11b, it is shown that in case of dysfunction of the SLD nucleus and/or its afferent or efferent pathways, the normal inhibition of the spinal motor neuron is lost, thus permitting activation of the skeletal muscles during REM sleep [63].

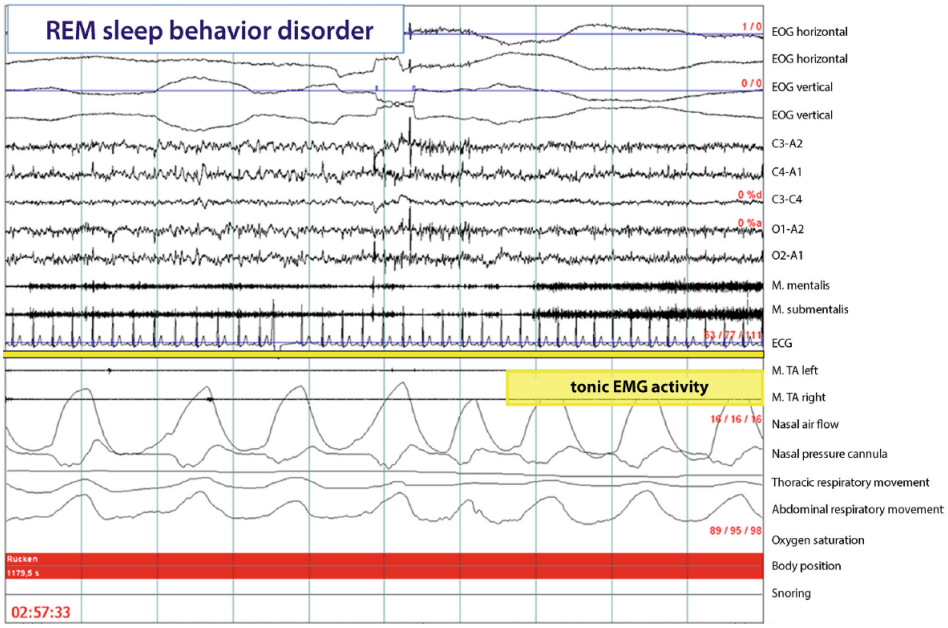
This overview of RBD pathophysiology makes clearer why iRBD is a prodromal stage of α-synucleinopathies according to Braak's staging [7, 8, 9]. RBD is triggered when α-synuclein pathology, on its brainstem pathway from the dorsal motor nucleus to the neocortex, affects the SLD nucleus.

Concerning secondary RBD, the lack of atonia in REM sleep could be triggered by a similar neurodegenerative process when RBD accompanies PD, DLB and MSA. In other secondary RBD conditions, it could instead be triggered by lesions and pharmacologically induced dysfunctions of the SLD nucleus and its afferent or efferent pathways, or both [63].

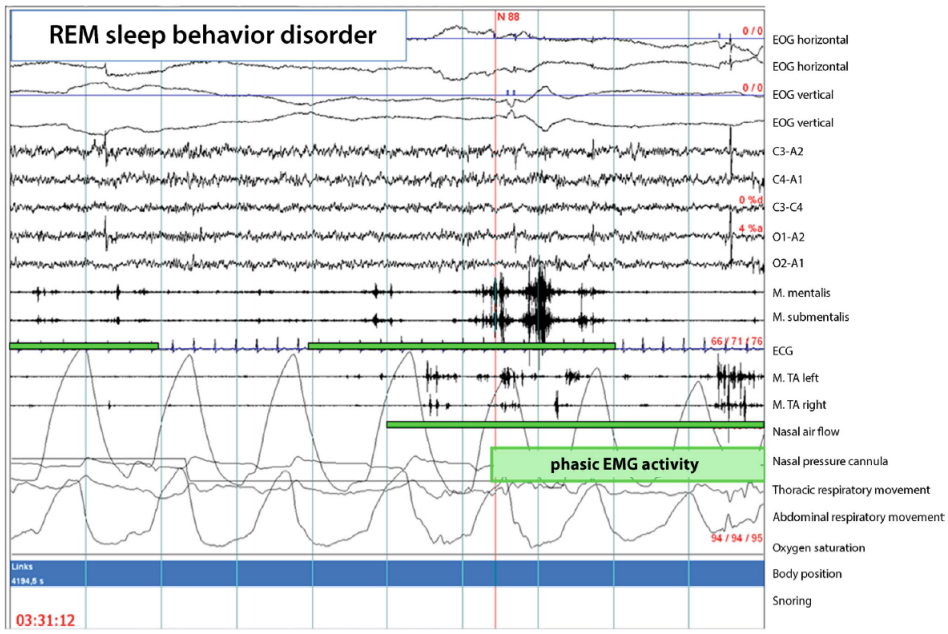
2.3.3 Polysomnographic changes in iRBD patients as biomarkers of prodromal α-synucleinopathies

The main electrophysiological hallmark of iRBD is RSWA. However, in recent years, research has been conducted to identify other electrophysiological and polysomnographic changes that could be used as biomarkers to strengthen detection of neurodegeneration in its early stages [64, 65].

Concerning brain activity, EEG slowing during wakefulness and REM sleep has been observed in iRBD patients when compared to age-matched healthy controls [66]. The slowing was found to be more pronounced in iRBD patients who later developed mild cognitive impairment [67, 68] and α-synucleinopathies [69, 70]. Moreover, alterations in EEG stability during REM sleep have been recently found in iRBD patients when compared to healthy controls [71]. In addition, EEG spectral power in REM sleep has been used to successfully distinguish these two groups [72]. In

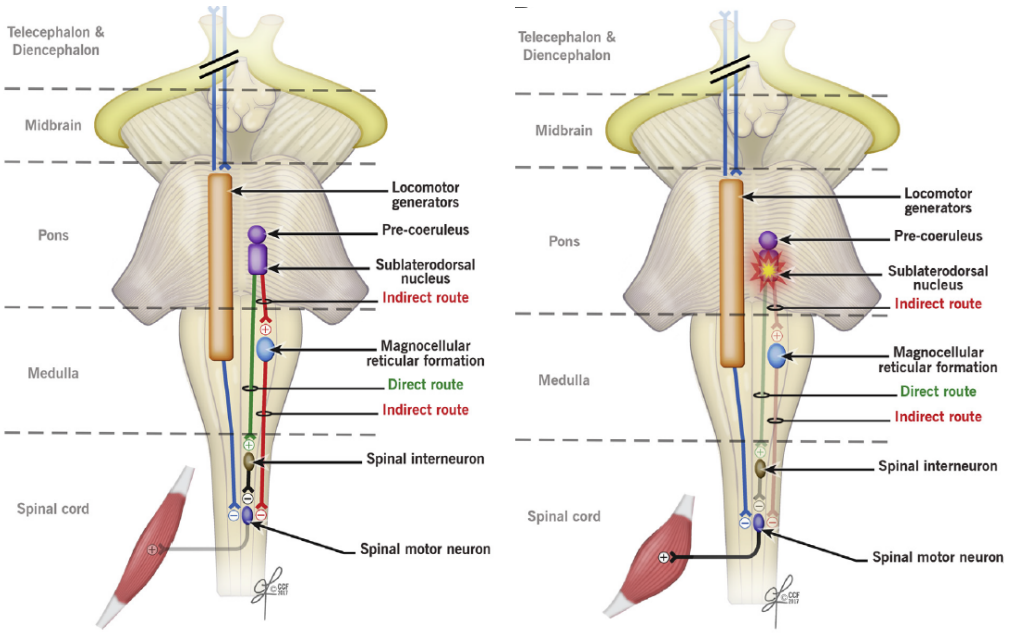


(a) Tonic EMG activity in the chin muscles (i.e. mentalis and submentalis).



(b) Phasic EMG activity in the chin (i.e. mentalis and submentalis) and tibialis muscles.

Figure 2.10: Tonic and phasic EMG activity. Reprinted from [61], under the Open Access agreement.



(a) REM sleep atonia generation. (b) RBD pathophysiology.

Figure 2.11: Physiology of generation of atonia during REM sleep (a) and RBD pathophysiology (b). Reprinted from [63] with permission from Elsevier.

regards to NREM sleep, no significant changes in overall NREM EEG content have been found in iRBD patients when compared to healthy controls [73]. However, significant reduction of sleep spindle density has been reported [74, 75], as well as alterations in cycling alternating patterns in NREM sleep, which are considered a marker of conversion to α-synucleinopathies [76]. Overall, these changes demonstrate alterations in the cortical electrophysiology of iRBD patients. Moreover, these findings are concordant with the EEG slowing [77] and with the reduction of sleep spindle density [74, 78] seen in PD patients compared to healthy controls, thus further proving the common neurodegenerative process.

Concerning EOG, differences in time-frequency patterns and coverage of eye movements could differentiate iRBD and PD patients from healthy controls [79, 80]. Moreover, significant changes in the morphology and/or timely distribution of nocturnal eye movements evaluated with a data-driven method have been found in iRBD and PD patients when compared to healthy controls [81]. As regulation of eye movements is complex and involves the brainstem, midbrain and cerebral cortex [82], these findings reflect the spreading of neurodegeneration in iRBD and PD patients.

In addition, reduction of heart rate variability during sleep has been observed in iRBD and PD patients when compared to healthy controls [83, 84, 85]. This suggests abnormalities of sympathetic and parasympathetic functions in iRBD and PD patients, which have also been confirmed by cardiac scintigraphy [86].

In addition to these micro-sleep alterations, abnormalities have been found also at a macro-structural level [87, 88]. More specifically, by applying an automated method for scoring sleep stages [89], increased sleep instability and transitions have been observed in iRBD and PD patients compared to controls [87, 88].

In summary, these findings show that iRBD and PD have similar abnormalities in both micro and macro-sleep structure when compared to age-matched healthy controls. These further confirm

the presence of ongoing neurodegeneration in iRBD patients and that the brainstem and the structures controlling sleep are similarly altered in iRBD and PD patients.

2.3.4 The emerging concept of prodromal RBD

In recent years, emerging evidence suggests that RBD associated with α-synucleinopathies is not a disease with a clearcut beginning, but is preceded by a phase with slightly increased EMG tone in REM sleep accompanied by minor REM behavioral events (RBEs) [14, 15]. RBEs are defined as vocalization and/or motor behaviors with a purposeful component [56]. This stage preceding definite RBD development has been named *prodromal RBD*. When in the prodromal RBD stage, patients do not meet the cutoff criteria of abnormal EMG tone for RBD diagnosis (Section 2.3.1), but show at least two RBEs in one night.

Fig. 2.12 shows a theoretical scheme of the changes in EMG tone and behaviors during REM sleep in the progression from normal REM sleep to development of α-synucleinopathy [15]. Currently, conversion from RBEs, as prodromal RBD, to definite RBD has been proven only in PD patients [14], but a similar progression is thought to occur in patients without already developed overt PD [15].

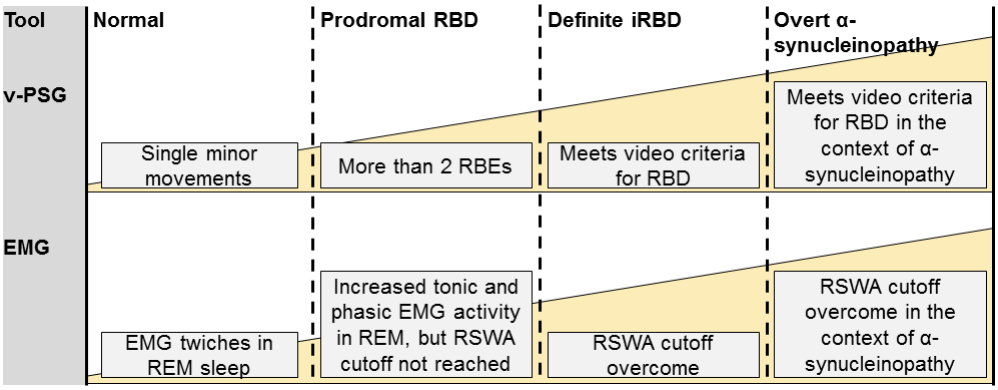


Figure 2.12: Progression from normal REM sleep to an overt α-synucleinopathy. From normal REM sleep, a patient enters the prodromal stage of RBD and then progresses to definite iRBD and finally to an α-synucleinopathy. Figure inspired from [15].

As a final remark, it should be mentioned that recent research is also focusing on patients showing isolated RSWA (iRSWA), who are characterized by EMG tone above RBD-cutoffs, but do not have any typical clinical sign of RBD [90, 91]. A follow-up study has proven that iRSWA patients progress to neurodegeneration [92], however more research is needed to understand the role of iRSWA as an early α-synucleinopathy biomarker.

2.4 Thesis motivation

The previous sections have highlighted that iRBD and prodromal RBD constitutes early stages of neurodegeneration. However, current methods for their identification are based on visual and manual analysis, which include sleep staging (Section 2.2.1), identification of RSWA (Section 2.3.1) and video inspection (Section 2.3.4). As the rules for scoring sleep and for identifying RSWA and RBEs are not uniquely defined, diagnosis of RBD and recognition of RBEs is not only time-consuming, but also subjective. When neuroprotective trials will be available, it will be

very important to have objectively defined and homogeneous groups of patients in early stages of neurodegeneration.

In this scenario, the development of automatic algorithms to identify and characterize patients with early neurodegeneration has a crucial importance, as this will lead to faster and more homogeneous identification in sleep labs across different countries. Despite important advancements in recent years in this field, current research faces a number of unmet needs. This thesis focuses on three unmet needs, from which the three thesis research questions and objectives were defined (stated in Section 1.2):

- ***Thesis unmet need 1***: Manual identification of RSWA and patients with RBD is heavily influenced by subjectiveness, which results in a high inter-rater variability [93]. To overcome the time consumption and the subjectivity of manual scoring of EMG activity, some research groups have proposed automated methods for identifying RBD patients [18, 19, 20, 21, 22, 23, 24, 25, 26]. However, no study has made an objective evaluation of these automated methods on the same database, to understand which of them the optimal one for RBD identification is.

- ***Thesis unmet need 2***: Many of the currently available automated methods to identify RBD are based on traditional programming techniques, implementing human-defined rules for scoring RSWA and identify RBD patients. Machine learning techniques are data-driven, and they are able to automatically learn data patterns, which might not be visible to human eyes. Because of the data-driven approach, machine learning techniques are potentially superior to traditional programming [27]. Moreover, most of the previously developed automated methods to identify patients with RBD have been validated only on one single database, thus they might suffer from a bias [18, 19, 20, 21, 22, 23, 24, 25, 26]. A generalized and robust to inter-clinical variability machine learning algorithm for automatic identification of patients with RBD is currently lacking.

- ***Thesis unmet need 3***: Prodromal RBD has been shown to be precursor of definite RBD in PD patients [14] and it is thought that the same process might occur also in absence of PD [15]. However, prodromal RBD can be currently identified only with video, as these patients show RBEs, but not elevated RSWA. Analyses of objective electrophysiological abnormalities of patients with prodromal RBD are lacking and no automated method has been developed to identify prodromal RBD.

Chapter 3

Comparison of automated methods for RBD detection

Chapter main objectives: Manual methods to identify RBD patients are subjective and time-consuming. To overcome these problems, a number of different automated methods has already been proposed. However, a systematic comparative analysis of them is lacking. Moreover, it is also not clear which effect muscular activity related to other sleep events (such as arousals and apneas) has on such methods, and whether such muscular activity influences RBD diagnosis.

This chapter aims to present a systematic comparison of currently available automated methods for RBD detection and to investigate the influence of muscular activity related to apneas and arousals on them.

This chapter is based upon Paper I [94] and addresses thesis objective 1.

3.1 Research background

Despite the fact that AASM manual proposes some rules to identify RSWA for RBD diagnosis [13], different research groups have proposed alternative visual methods, which are presented in Table 3.1. Visual and manual methods for identifying RSWA are time-consuming and prone to subjective interpretation, as the rules to score muscular activity are not uniquely interpretable. A recent study has shown that, when AASM rules to score phasic and tonic EMG activity during REM sleep were employed, unanimous agreement between 6 scorers was almost never achieved, and the agreement was found to be higher in phasic than in tonic activity [93].

To overcome these issues related to manual scoring, some research groups have already developed a number of automated methods for identification of RBD. All the developed methods require manual identification of REM sleep and manual removal of muscular activity (MA) related to arousals[1] and apneas[2]. Because of this, these methods are *actually* semi-automated, but they will be referred as automated in this thesis. An overview of these methods is presented in Table 3.2.

Comparative analyses of manual and automated methods have been performed [95, 96, 97], as well as comparative analyses of the automated methods on patients suffering from narcolepsy [98]. However, no study has performed a systematic comparison of the automated methods to identify which of them has the highest performances in identifying patients with RBD. In addition, only

[1] Defined as abrupt shift of EEG frequency including alpha, theta and/or frequencies greater than 16 Hz, that lasts at least 3 s with at least 10 s of stable sleep preceding the change. An arousal in REM sleep should be accompanied by an increment in EMG tone for at least 1 s [13].

[2] Defined as the cessation of the airflow for at least 10 s [13].

Method name	Muscular activity definition	RSWA criteria for RBD diagnosis
AASM [13, 60]	*Tonic EMG activity*: a 30-s REM epoch has tonic EMG activity if at least 50% of the duration has chin EMG amplitude greater than the minimum amplitude in NREM sleep. *Phasic EMG activity*: a 30-s REM sleep epoch has phasic EMG activity if at least five out of ten 3-s mini-epochs contain bursts of transient activity with duration of 0.1-5 s and amplitude at least 4 times higher than the background EMG activity.	$>$ 27% of the 30-s REM sleep epochs have any (i.e. phasic/tonic) EMG activity combined with bilateral phasic activity in the FDS muscles.
Montréal [99]	*Tonic EMG activity*: a 20-s REM sleep epoch has tonic EMG activity if at least 50% of the epoch has chin EMG amplitude twice the background or greater than 10 μV. *Phasic EMG activity*: Phasic chin EMG density is scored as the percentage of 2-s REM sleep mini-epochs containing EMG events lasting 0.1-10 s and amplitude exceeding 4 times the background EMG activity.	Either $\geq$30% of 20-s REM sleep epochs have tonic chin EMG activity or $\geq$15% of the 2-s REM sleep mini-epochs have phasic EMG activity.
SINBAR [100]	*Tonic EMG activity*: a 30-s REM sleep epoch has tonic EMG activity if at least 50% of the epoch has chin EMG amplitude twice the background or greater than 10 μV. *Phasic EMG activity*: Phasic chin EMG density is scored as the percentage of 3-s REM sleep mini-epochs containing EMG events lasting 0.1-5 s and amplitude exceeding 2 times the background EMG activity.	$>$32% of 3-s REM sleep mini-epochs with any (tonic/phasic) EMG activity combined with bilateral phasic EMG activity in FDS muscles.
Mayo [101]	*Tonic EMG activity*: a 30-s REM sleep epoch has tonic EMG activity if at least 50% of the epoch has chin EMG amplitude twice the background or greater than 10 μV. *Phasic EMG activity*: Phasic chin EMG density is scored as the percentage of 3-s REM sleep mini-epochs containing EMG events lasting 0.1-14.9 s and amplitude exceeding 4 times the background EMG activity. The end of a phasic burst corresponds to return to baseline for at least 200 ms.	$>$15.5% of 3-s REM sleep mini-epochs with phasic EMG activity in the chin muscle; $>$21.6% of 3-s REM sleep mini-epochs with any (i.e. phasic/tonic) EMG activity in chin muscle; $>$30.2% of 3-s REM sleep mini-epochs with any or phasic EMG activity in the anterior tibialis muscles; $>$37.9% of 3-s REM sleep mini-epochs with phasic EMG activity for combined chin and tibialis muscles; $>$43.4% of 3-s REM sleep mini-epochs with any EMG activity for combined chin and tibialis muscles.

Table 3.1: Overview of manual methods for RSWA identification.

Method	Procedure
REM atonia index (RAI) [18, 19]	(1) Signal: chin EMG (2) *amp* = average rectified amplitude calculated for each 1-s mini-epoch (3) Noise-reduction: *amp* = *amp* - minimum amp in surrounding moving window (± 30 s) (4) Division of the obtained amp during REM in 20 categories: $amp \leq 1$ μV, 1 μV$<amp\leq$2 μV,..., 18 μV$<amp\leq$19 μV and $amp \geq 20$ μV (5) RAI = (%$amp\leq$1 μV) /(100- (%1 μV$<amp\leq$2 μV))
Supra-threshold REM activity metric (STREAM) [20]	(1) Signal: chin EMG (2) Signal divided in 3-s mini-epochs and variance calculated for each of them (3) STREAM = % of 3-s mini-epochs in REM with variance $> 5^{\text{th}}$ percentile of variances during NREM
Short and long muscle activity index (sMAI and lMAI) [21, 22]	(1) Signal: chin EMG (2) Calculation of smoothed (0.025 s) lower and upper envelopes (3) Amplitude = difference between upper and lower envelopes (4) Amplitude smoothed over 200 s with moving window (5) Threshold defined as 2 times the smoothed amplitude + 10 μV (6) Movements identified when amplitude > threshold (7) Clusters of movements at distance < 1 s merged (8) sMAI = (# movements with length < 0.5 s during REM) / (hours of REM sleep) (9) lMAI = (# movements with length $\geq$ 0.5 s during REM) / (hours of REM sleep)
Frandsen index (FRI) [23]	(1) Signal: chin EMG (2) Amplitude curve (AC) calculated as difference between highest and lowest EMG value in a 200 ms window. (3) AC divided in 30-s segments and median amplitude (MAmp) calculated for each of them (4) Baseline identified for each segment as the lowest MAmp in the moving window surrounding the segment (±30 min) (5) Muscle activity identified when AC > 4 times the baseline and minimum duration of 0.3 s (6) Clusters of muscle activity at distance <0.5 s merged (7) FRI = % of REM 3-s mini-epochs with muscle activity in more of 50% of the time
Kempfner index (KEI) [24, 25]	(1) Signals: chin EMG, left and right tibialis EMG (2) Bandpass filtering at 30-65 Hz (3) Envelope of the EMG signals obtained with normalized Blackman window (number of samples corresponding to half of the sampling frequency) (4) Envelope divided in 3-s mini-epochs. For each 3-s mini epoch (x_{test}) a reference window (x_{ref}) defined as the 5 mini-epochs before and after x_{test} (5) For each mini-epoch n in REM sleep, calculation of $G(n) = mean(x_{test}(n))/(min(x_{ref}(n)) + 10^{-5})$ (6) Classification performed by using the features G from the 3 EMG signals. Model: a one-class support vector machine (OC-SVM) to identify muscular activity as outliers. The model was trained on healthy controls (HC) using the assumption that HC show atonia in REM sleep. (7) KEI = 100 × #outliers/(#outliers+#inliers)
Automated SINBAR [26]	(1) Signals: chin EMG, left and right FDS EMG (2) The background EMG was adaptively calculated over 15 s. (3) The onset criterion for phasic EMG activity is selected with twice the background EMG or 2 μV, whichever is higher; the offset criterion is selected when the amplitude falls below 10% of the average EMG signal amplitude or below 2 μV, whichever is higher (4) For any EMG activity, the duration was increased to 15 s (5) The tonic activity of the mentalis muscle is evaluated in 1-s windows over the entire night and divided into four categories ranging from low level to high level EMG activity. Tonic EMG activity is detected when at least 50% of the respective 30-s epoch contains a background tone higher than the range of the low level background category augmented by 1 μV (6) The same cut-off and criteria as the manual method (Table 3.1) are applied

Table 3.2: Overview of automated methods for RBD identification.

one study has analyzed systematically the influence of MA related to apneas and arousals on the outcomes of an automated method and whether these have an influence on the correct identification of RBD patients [26].

3.2 Research questions and objectives of this chapter

From this background overview, the following research questions were defined:

- ***Chapter research question 3.1:*** Which one(s) of the available automated methods can best identify patients with RBD?
- ***Chapter research question 3.2:*** To which extent does MA related to apneas and arousals influence the outcomes of the automated methods? Is removal of such MA necessary?

These research questions led to the definition of the following objectives:

- ***Chapter research objective 3.1:*** To compare systematically, on the same cohort, the available automated methods for RBD identification, to identify the one(s) performing best.
- ***Chapter research objective 3.2:*** To analyze the influence of apneas and arousal-related MA in the identification of patients with RBD.

3.3 Paper I: Comparison of computerized methods for rapid eye movement without atonia detection

To compare the different methods, their capability of separating different groups of participants with and without RBD was evaluated by means of accuracy, sensitivity and specificity. The groups included healthy controls, patients with iRBD, patients with PD with and without RBD, and patients with periodic limb movement disorder (PLMD). This last group was included as PLMD patients have abnormal increased MA in the limbs [60], but this disease is not related to any type of neurodegeneration and therefore it should be carefully distinguished from RBD. The mentioned performance metrics were computed in different configurations, which included removal or not of MA related to apneas and arousals. By comparing these performances, it was possible to identify the best method(s) for RBD identification and whether the outcomes of the automated methods were affected by arousal and apnea-related MA. The overview of the methodology is presented in Fig. 3.1.

3.3.1 Methods: Comparative analysis by means of logistic regression

Participants and recordings

The study cohort included 27 healthy control participants (HC), 25 patients with PD without RBD (PDnonRBD), 29 patients with PD and RBD (PD+RBD), 29 patients with iRBD, and 36 patients with PLMD. All of them were investigated with one night v-PSG at the Danish Center for Sleep Medicine (DCSM). All participants were advised to discontinue drugs interfering with PSG recordings (i.e. hypnotics and antidepressants) two weeks before the study. Sleep diagnoses were made by expert medical doctors in accordance with international criteria [60]. An overview of the demographics and sleep information is provided in Table 3.3.

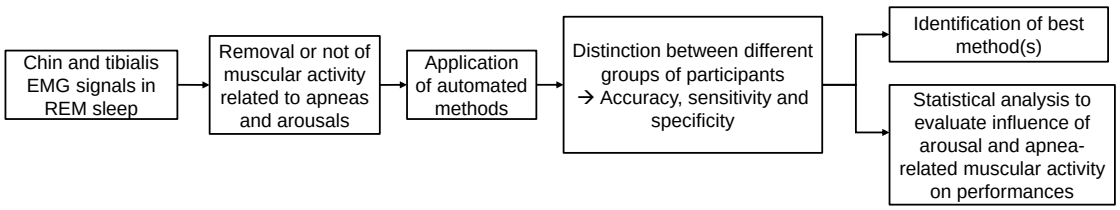


Figure 3.1: Overview of the methodology of Paper I [94]. Sleep chin and tibialis EMG signals from HCs, PLMD patients, iRBD patients, PD patients with RBD, and PD patients without RBD were included. The automated methods were applied to them in different configurations, which included removal or not of MA related to arousals and apneas. The performances in distinguishing different combinations of participant groups were compared, to elect the best method(s) for RBD identification and to evaluate whether the methods outcomes were influenced or not by arousal and apnea-related MA.

Parameter	**HC**	**PDnonRBD**	**PD+RBD**	**iRBD**	**PLMD**	**p-val**
Total count	27	25	29	29	36	-
Fraction of men	0.48	0.68	0.72	0.72	0.61	0.28
Age [years, $\mu \pm \sigma$]	56.6±9.2	63.7±8.0	63.1±5.8	57.7±17.2	58.8±14.8	0.11
AHI [#apneas/hsleep, $\mu \pm \sigma$]	5.9±5.4	15.7±19.4	9.9±16.6	13.8±17.5	9.8±12.2	0.51
PLMS index [#PLM/hsleep, $\mu \pm \sigma$]*	6.3±11.2	1.5±2.3	10.4±22.6	26.7±34.4	50.9±42.5	**<0.001**
#nPLM/hsleep [$\mu \pm \sigma$]*	5.5±5.7	6.2±7.3	4.9±3.3	6.7±6.1	7.2±6.6	0.69
#arousals/hREM [$\mu \pm \sigma$]	11.0±8.6	2.7±3.7	2.7±5.6	16.7±17.2	8.5±5.9	**<0.001**
#apneas/hREM [$\mu \pm \sigma$]	7.9±13.3	9.3±17.3	3.1±11.6	14.7±22.7	10.7±14.7	**<0.001**
#PLM/hREM [$\mu \pm \sigma$]*	4.9±9.3	0.6±1.5	15.8±33.9	37.1±46.4	25.5±33.1	**<0.001**
#nPLM/hREM [$\mu \pm \sigma$]*	7.5±8.9	5.1±7.9	3.1±4.3	9.1±8.6	10.2±9.3	**0.04**
#arousals/hNREM [$\mu \pm \sigma$]	9.8±5.6	3.9±5.6	4.4±7.7	14.7±15.2	12.8±9.4	**<0.001**
#apneas/hNREM [$\mu \pm \sigma$]	2.2±3.4	7.0±12.1	4.0±14.6	11.6±16.6	7.0±8.8	**<0.001**
#PLM/hNREM [$\mu \pm \sigma$]*	6.3±12.4	1.5±2.3	9.5±20.1	26.6±36.2	55.12±46.9	**<0.001**
#nPLM/hNREM [$\mu \pm \sigma$]*	4.9±5.3	6.3±7.3	5.1±3.9	5.4±5.5	5.4±5.8	0.96
Sleep efficiency [%, $\mu \pm \sigma$]	86.9±9.3	70.3±23.0	70.6±16.8	75.5±24.0	74.4±25.3	**<0.001**
Time in bed [min, $\mu \pm \sigma$]	500.8±71.8	446.4±133.8	518.7±189.8	449.3±86.8	447.7±76.9	**0.01**
REM latency [min, $\mu \pm \sigma$]	94.0±42.9	163.4±205.1	195.2±143.2	154.7±99.0	114.7±73.8	**0.009**
W [%,$\mu \pm \sigma$]	13.0±9.2	29.7±23.0	29.4±16.8	20.1±15.8	19.7±13.4	**<0.001**
REM [%, $\mu \pm \sigma$]	20.1±5.9	10.1±8.4	10.8±8.9	14.1±7.9	15.6±6.6	**<0.001**
N1 [%, $\mu \pm \sigma$]	8.0±4.4	7.7±5.5	13.2±10.4	11.3±9.3	10.2±8.7	0.18
N2 [%, $\mu \pm \sigma$]	44.8±8.8	32.0±16.5	37.6±14.9	35.3±16.1	37.8±17.1	0.05
N3 [%, $\mu \pm \sigma$]	14.0±7.6	20.6±18.5	8.9±8.2	14.8±15.5	10.8±10.1	**0.04**

Table 3.3: Demographic and sleep data of the included cohort. Statistical comparison of the fraction of men was made with chi-square test; all other statistical comparisons were made with Kruskal-Wallis test. p-Values <0.05 are highlighted in bold. PLMS: periodic limb movement (PLM) during sleep; nPLM: limb movements not included in PLM series; AHI: apnea/hypopnea index; hREM/hNREM/hsleep: hours of REM/NREM/sleep. *PLMS index, #nPLM/hsleep, #PLM/hREM, #nPLM/hREM, #PLM/hNREM, and #nPLM/hNREM were available only for 18 HCs, 12 PDnonRBD, 8 PD+RBD, 26 iRBD, and 32 PLMD patients.

The PSG recordings were manually scored for sleep staging, respiratory events and limb movements according to AASM criteria [13]. The recordings included chin and anterior tibialis left (TIBL) and right (TIBR) EMG signals, sampled at 256 Hz. Different hardware systems were used for the recordings, of which one had an anti-aliasing low-pass frequency cut-off at 70 Hz. To make the analysis homogeneous, all the EMG signals were filtered with a 3 dB 4^{th} order zero-phase Butterworth band-pass filter between 10 and 70 Hz and with a 3 dB 4^{th} order zero-phase Butterworth notch filter at 50 Hz (cut-off frequencies at 48 and 52 Hz). For each EMG signal, saturation, detachment and pop-up electrode artifacts were identified as areas exceeding 4 mV. The 5 s preceding and following these areas were also excluded from analysis. Because of the absence of REM sleep in 5 PDnonRBD, 2 PD+RBD, 1 iRBD, and 1 PLMD patients, they were excluded from the analysis.

Automated methods for RSWA detection

The following six automated MA indices were compared in this study (see Table 3.2 for more details):

- REM atonia index (RAI) [18, 19], which measures the percentage of 1-s REM sleep mini-epochs with atonia in the chin EMG signal;
- Supra-threshold REM activity metric (STREAM) [20], which measures the percentage of 3-s REM sleep mini-epochs with RSWA in the chin signal;
- Short and long muscle activity index (sMAI and lMAI) [21, 22], which are measures of the number of short and long muscular activities per hour of REM sleep in the chin EMG signal, respectively;
- Frandsen index (FRI) [23], which measures the percentage of 3-s REM sleep mini-epochs with RSWA in the chin EMG signal;
- Kempfner index (KEI) [24, 25], which is a measure of the percentage of 3-s REM sleep mini-epochs with RSWA obtained by analyzing the chin and tibialis muscles.

All of them were implemented by following accurately the descriptions provided in the relevant scientific papers. It was not possible to implement the computerized version of the SINBAR method, due to lack of important implementation aspects in its description [26].

Evaluation method

To evaluate the influence of MA related to apneas and arousals on the different automated methods, they were applied in different *configurations*:

- *Configuration 1*: Without excluding any MA related to apneas and arousals.
- *Configuration 2*: Excluding from analysis areas located from 3 s before to 12 s after an arousal onset (Fig. 3.2a).
- *Configuration 3*: Excluding from analysis areas from 5 s before an apnea onset to 5 s after the same apnea end (Fig. 3.2b).
- *Configuration 4*: Excluding from analysis areas located 5s before an apnea event and 5s after its end (Fig. 3.2c).

- *Configuration 5*: Combination of configurations 2 and 3.
- *Configuration 6*: Combination of configurations 2 and 4.

The application of the automated methods resulted in having the values of each MA index (i.e. RAI, STREAM, FRI, sMAI, lMAI and KEI) in the different six configurations for all participants. For each configuration, it was evaluated how well each index was performing in the following *comparisons*:

- *Comparison 1:* (HC, PDnonRBD, PLMD) vs (iRBD, PD+RBD).
- *Comparison 2:* HC vs iRBD.
- *Comparison 3:* PLMD vs iRBD.
- *Comparison 4:* HC vs PDnonRBD.
- *Comparison 5:* HC vs PLMD.
- *Comparison 6:* PDnonRBD vs PD+RBD.
- *Comparison 7:* HC vs PLMD vs iRBD.

Logistic regression was employed to evaluate the capability of each index in differentiating the participants for each comparison and configuration. Logistic regression was chosen because it was hypothesized that the data were linearly separable, because of independence of the data and because of its easy interpretability [102].

Logistic Regression An overview of logistic regression is now presented [102]. At the basis of logistic regression, there is the concept of *logistic sigmoid* function $\sigma(\cdot)$, defined as:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{3.1}$$

In the context of $C = 2$ class classification (classes c_1 and c_2), the posterior probability of class c_1 can be written as:

$$p(c_1|\boldsymbol{\phi}) = y(\boldsymbol{\phi}) = \sigma(\mathbf{w}^T\boldsymbol{\phi}) \tag{3.2}$$

where the logistic sigmoid function acts on a linear combination (with weights $\mathbf{w}$) on a feature vector $\boldsymbol{\phi}$ of dimension M. The model in Eq. 3.2 is known as *logistic regression* model. By consequence, the posterior probability for class c_2 is $p(c_2|\boldsymbol{\phi}) = 1 - p(c_1|\boldsymbol{\phi})$.

To optimize the M weights, a maximum likelihood approach is used. By considering a training dataset $\{\boldsymbol{\phi}_n, t_n\}$ where $t_n \in \{0, 1\}$ represents the *true* class for the n^{th} feature vector $\boldsymbol{\phi}_n$ (with $n = 1, ..., N$), the likelihood function is expressed as:

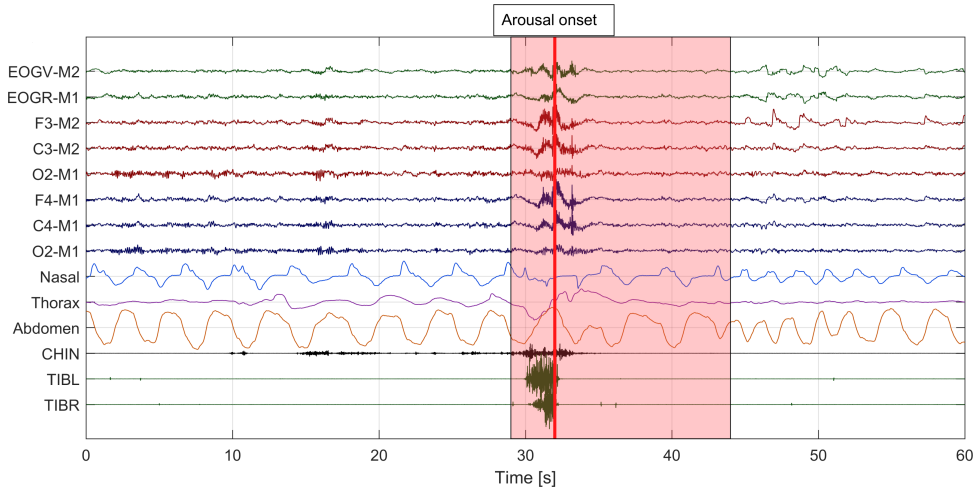
$$p(\mathbf{t}|\mathbf{w}) = \prod_{n=1}^{N} y_n^{t_n}(1 - y_n)^{1-t_n} \tag{3.3}$$

where $\mathbf{t} = (t_1, ..., t_N)^T$ and $y_n = \sigma(\mathbf{w}^T\boldsymbol{\phi}_n) = p(c_1|\boldsymbol{\phi}_n)$. The cross-entropy error can be defined as:

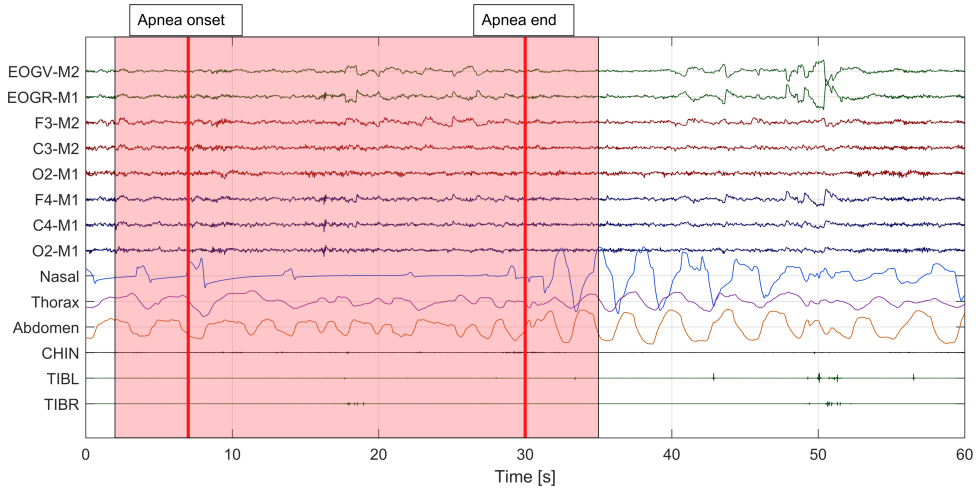
$$E(\mathbf{w}) = -\ln p(\mathbf{t}|\mathbf{w}) = -\sum_{n=1}^{N}(t_n \ln y_n + (1 - t_n)\ln(1 - y_n)) \tag{3.4}$$

The Newton-Raphson criterion is then used to optimize the weights as follows:

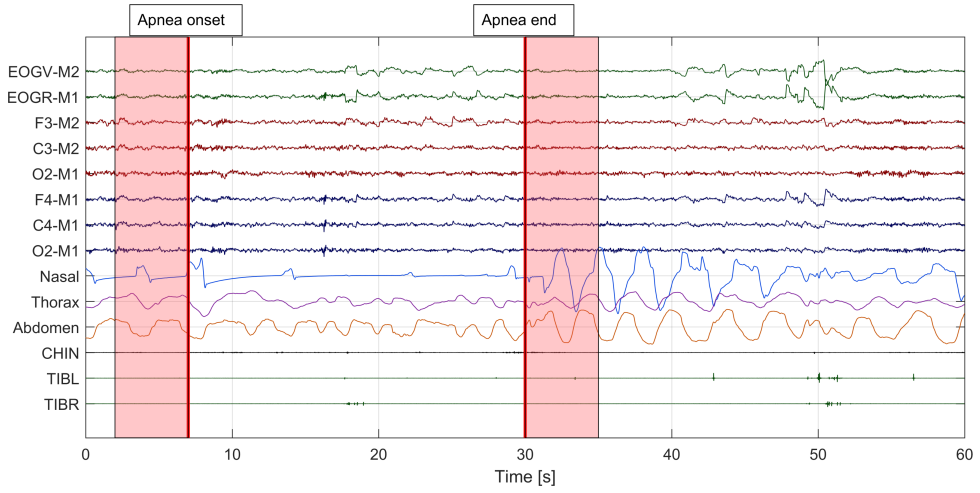
$$\mathbf{w}^{(new)} = \mathbf{w}^{(old)} - \mathbf{H}^{-1}\nabla E(\mathbf{w}) \tag{3.5}$$



(a) Segment removed related to an arousal.



(b) Segment removed related to an apnea in its whole length.



(c) Segments removed related to the onset and end points of an apnea.

Figure 3.2: Illustrations of the excluded segments (highlighted in red) of the PSG recording. (a) From 3 s before to 12 s after an arousal onset; (b) from 5 s before an apnea onset to 5 s after an apnea end event; (c) from 5 s before to an apnea onset and from an apnea end to 5 s later. The choice of the length of segments excluded for the analysis was based on empirical analysis of the signals.

where $\mathbf{H}$ is the Hessian matrix whose elements correspond to the second derivatives of $E(\mathbf{w})$ with respect to $\mathbf{w}$. For logistic regression, it can be shown that

$$\mathbf{w}^{(new)} = (\mathbf{\Phi}^T\mathbf{R}\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{R}\mathbf{z} \tag{3.6}$$

where $\mathbf{z} = \mathbf{\Phi}\mathbf{w}^{(old)} - \mathbf{R}^{-1}(\mathbf{y} - \mathbf{t})$, $\mathbf{\Phi}$ is a $N \times M$ whose n^{th} row is ϕ_n^T, $\mathbf{R}$ is a $N \times N$ diagonal matrix with $R_{nn} = y_n(1 - y_n)$, and $\mathbf{y} = (y_1, ..., y_N)^T$. From Eq. 3.6 it can be seen that the weights are continuously updated and the stop criterion can be set on the number of iterations (N_{it}) performed and/or on the value of $E(\mathbf{w}^{(new)}) - E(\mathbf{w}^{(old)})$ being lower than a certain threshold (ϵ).

To avoid overfitting, logistic regression can be modified with regularization. This means that the error function is

$$E(\mathbf{w}) = -\sum_{n=1}^{N}(t_n \ln y_n + (1 - t_n)\ln(1 - y_n)) + \frac{1}{2}\gamma||\mathbf{w}||^2 \tag{3.7}$$

Similarly, logistic regression can be adopted in a multi-class framework with $C > 2$ different classes. More details on this can be found in [102].

Application of logistic regression For generalization purposes, $R = 20$ runs of $K = 5$ fold cross-validation (CV) were implemented for each comparison, configuration and index (Fig. 3.3). In each run, the participants were divided into 5 random folds, of which 4 were used for training a logistic regression model and the remaining for validating the trained model (Fig. 3.4). This process was repeated for all the folds and runs, thus obtaining for each comparison, configuration and index 100 trained logistic regression models that were validated on the respective validation sets. During training, N_{it}, ϵ, and γ were set to 200, 10^{-4}, and 10^{-4} respectively. For comparisons 1-6, the validation accuracy (ACC), sensitivity ($SENS$), and specificity ($SPEC$) were calculated as:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad SENS = \frac{TP}{TP + FN} \quad SPEC = \frac{TN}{TN + FP} \tag{3.8}$$

where TP is the number of true positives (i.e. the model correctly predicts the positive class), TN the number of true negatives (i.e. the model correctly predicts the negative class), FP the number of false positives (i.e. the model incorrectly predicts the positive class), and FN the number of false negatives (i.e. the model incorrectly predicts the negative class). The values of sensitivity and specificity were calculated relatively to the following classes: (iRBD,PD+RBD) for comparison 1, iRBD for comparisons 2 and 3, PDnonRBD for comparison 4, PLMD for comparison 5, PD+RBD for comparison 6.

In case of the three-class problem of comparison 7, the performance measures were calculated as:

$$ACC = \frac{P_c}{N_{tot}} \quad SENS = \frac{TP_{iRBD}}{TP_{iRBD} + FN_{iRBD}} \quad SPEC_{iRBD} = \frac{TN_{iRBD}}{TN_{iRBD} + FP_{iRBD}} \tag{3.9}$$

where P_c is the number of participants classified correctly in each validation fold, N_{tot} the total number of participants included in each validation set, and TP_{iRBD}, TN_{iRBD}, FP_{iRBD}, FN_{iRBD} are the numbers of true positives, true negatives, false positives and false negatives for the iRBD class.

Statistical analysis To understand the influence of apnea and arousal-related MA on the different indices, it was evaluated whether the different configurations were leading to significantly

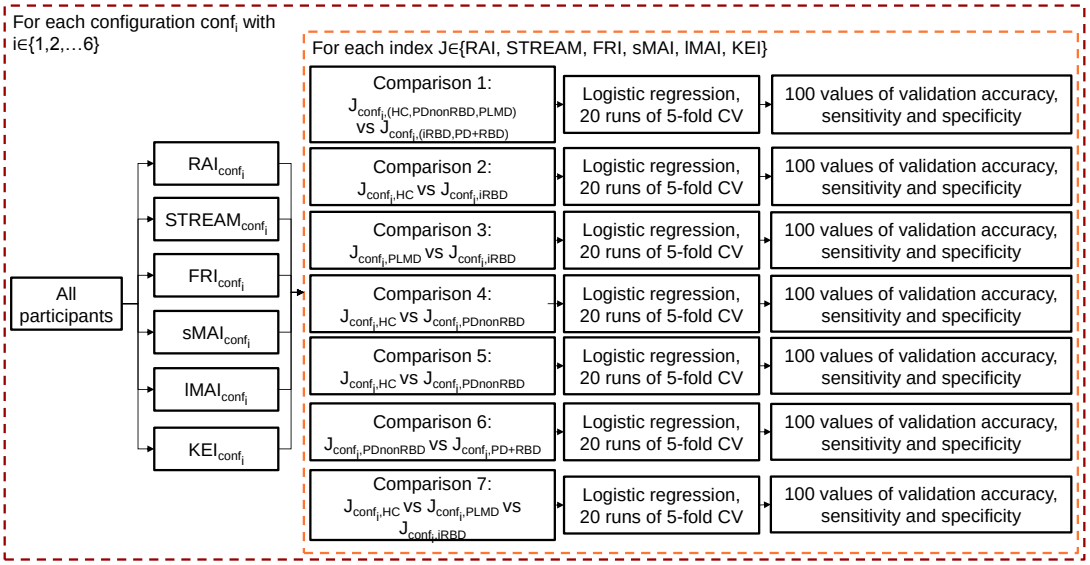


Figure 3.3: Schematic representation of the evaluation method implemented to compare the different automated methods. For each configuration, the values of all indices (RAI, STREAM, sMAI, lMAI and KEI) were calculated and used in seven comparisons. For each configuration, comparison and index, 20 runs of 5-fold CV with logistic regression models were implemented, thus obtaining 100 values of validation accuracy, sensitivity and specificity.

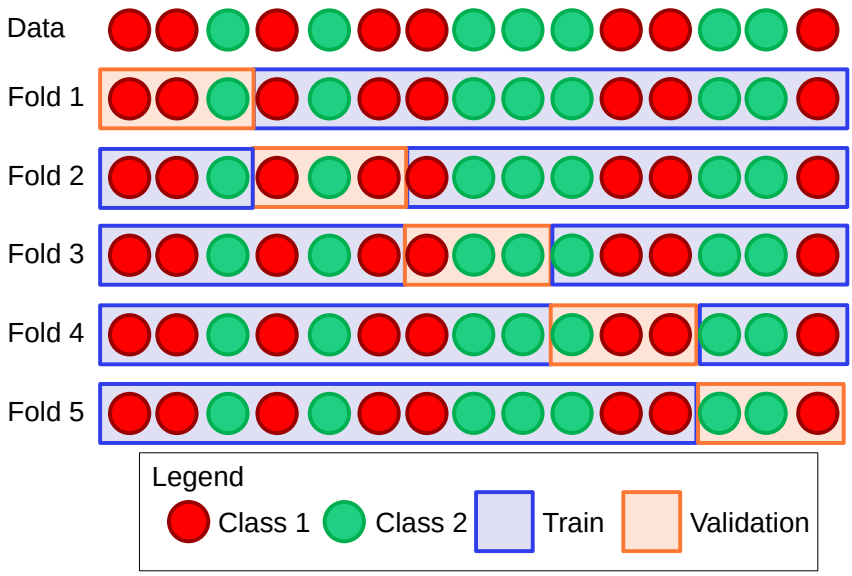


Figure 3.4: Graphical representation of 5-fold CV scheme. In each random fold, 4/5 of the data are used to train a model and the remaining 1/5 to validate it.

		Configuration					
		1	2	3	4	5	6
Comparison	1	$\overline{SENS}_{1,1}$	$\overline{SENS}_{1,2}$	$\overline{SENS}_{1,3}$	$\overline{SENS}_{1,4}$	$\overline{SENS}_{1,5}$	$\overline{SENS}_{1,6}$
	2	$\overline{SENS}_{2,1}$	$\overline{SENS}_{2,2}$	$\overline{SENS}_{2,3}$	$\overline{SENS}_{2,4}$	$\overline{SENS}_{2,5}$	$\overline{SENS}_{2,6}$
	3	$\overline{SENS}_{3,1}$	$\overline{SENS}_{3,2}$	$\overline{SENS}_{3,3}$	$\overline{SENS}_{3,4}$	$\overline{SENS}_{3,5}$	$\overline{SENS}_{3,6}$
	4	$\overline{SENS}_{4,1}$	$\overline{SENS}_{4,2}$	$\overline{SENS}_{4,3}$	$\overline{SENS}_{4,4}$	$\overline{SENS}_{4,5}$	$\overline{SENS}_{4,6}$
	5	$\overline{SENS}_{5,1}$	$\overline{SENS}_{5,2}$	$\overline{SENS}_{5,3}$	$\overline{SENS}_{5,4}$	$\overline{SENS}_{5,5}$	$\overline{SENS}_{5,6}$
	6	$\overline{SENS}_{6,1}$	$\overline{SENS}_{6,2}$	$\overline{SENS}_{6,3}$	$\overline{SENS}_{6,4}$	$\overline{SENS}_{6,5}$	$\overline{SENS}_{6,6}$
	7	$\overline{SENS}_{7,1}$	$\overline{SENS}_{7,2}$	$\overline{SENS}_{7,3}$	$\overline{SENS}_{7,4}$	$\overline{SENS}_{7,5}$	$\overline{SENS}_{7,6}$

Table 3.4: Table with average validation sensitivity values across configurations and comparisons for one index.

different values of validation sensitivity, specificity and accuracy. Considering for example one index (e.g. RAI) and one performance measure (e.g. validation sensitivity), the average performance measure across the 20 runs of 5-fold CV was calculated for each comparison and configuration, thus obtaining a 7×6 table as shown in Table 3.4. Friedman test was used to evaluate whether such average values were statistically different column-wise [103]. The same statistical analysis was repeated for all indices and performance measures, thus obtaining 18 final p-values, which were considered significant if lower than 0.0028 (i.e. 0.05/18, due to Bonferroni correction for multiple comparisons [104]).

3.3.2 Results

The automated indices were calculated for all configurations and Fig. 3.5 shows their distributions for configuration 1. The distributions for other configurations are shown in the Supplementary Figures in Appendix A.

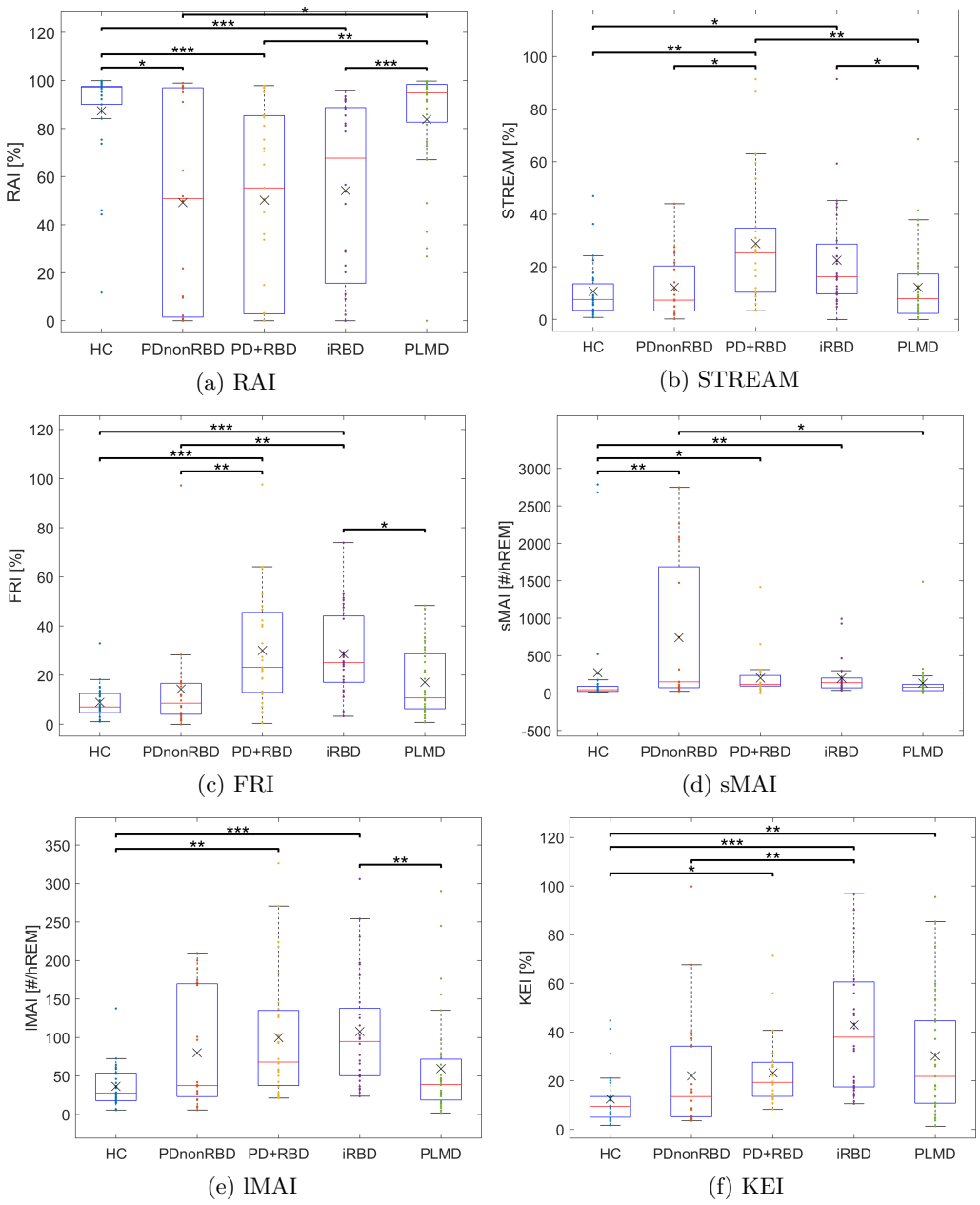
By analyzing the distributions of the indices across group participants and configurations, it was noticed that HCs were characterized by the lowest level of MA in REM sleep, whereas higher level was seen for iRBD and PD+RBD patients, as expected. The groups of PLMD and PDnonRBD patients were instead generally characterized by a level of MA in REM sleep intermediate between HC and (iRBD, PD+RBD).

Figs. 3.6 and 3.7 show the average values and 25^{th}-75^{th} percentiles of the validation sensitivity, specificity and accuracy across the logistic regression models for all indices, comparisons and configurations. The main outcomes are presented below:

- Comparison 1 (Figs. 3.6a and 3.7a): RAI, FRI and STREAM achieved the highest values of sensitivity, specificity and accuracy (average values around 70%) in distinguishing the patients with RBD from the ones without RBD.

- Comparison 2 (Figs. 3.6b and 3.7b): FRI outperformed the other indices in distinguishing HC from iRBD patients, with average values of sensitivity of around 90%, and specificity and accuracy of around 80%.

- Comparison 3 (Figs. 3.6c and 3.7c): RAI was the best index in differentiating iRBD from PLMD patients, with average values of sensitivity, specificity and accuracy of 70%.



(a) RAI (b) STREAM

(c) FRI (d) sMAI

(e) lMAI (f) KEI

Figure 3.5: Values of the automated indices calculated for configuration 1: (a) RAI, (b) STREAM, (c) FRI, (d) sMAI, (e) lMAI and (f) KEI. The values across the participant groups are shown as dots and with box plots with crosses denoting group means, and whiskers the 99$^{\text{th}}$ percentiles. Post-hoc statistical analysis performed with Mann-Whitney U-test with Tukey-Kramer correction *p<0.05, **p<0.01, ***p<0.001.

Index	Sensitivity	Specificity	Accuracy
RAI	0.7575	0.8032	0.3304
STREAM	0.2031	0.5882	0.8685
FRI	0.0681	0.0897	0.0897
sMAI	0.4844	0.9463	0.5297
lMAI	0.0150	0.0870	0.0155
KEI	0.1667	0.9189	0.3820

Table 3.5: p-Values obtained as outputs of the Friedman tests to evaluate the effect of the 6 different configurations across the seven comparisons on validation sensitivity, specificity, and accuracy.

- Comparison 4 (Figs. 3.6d and 3.7d): Theoretically, the indices should not be able to distinguish PDnonRBD from HC based on the levels of MA in REM sleep. STREAM presented the most coherent results with this (average sensitivity of around 0%, and average specificity and accuracy of around 50%).
- Comparison 5 (Figs. 3.6e and 3.7e): KEI achieved the highest average sensitivity, specificity and accuracy (range 60-80% excluding configuration 2) in distinguishing HC from PLMD patients.
- Comparison 6 (Figs. 3.6f and 3.7f): The highest average performances in distinguishing PDnonRBD from PD+RBD were achieved by STREAM, FRI (around 80% average sensitivity, 60% average specificity and 60-70% average accuracy) and sMAI (around 70% average sensitivity, 100% average specificity and 70% average accuracy).
- Comparison 7 (Figs. 3.6g and 3.7g): RAI outperformed the other indices in terms of average sensitivity (around 60-70%) and specificity (around 80%) in identifying iRBD patients in a cohort including also HCs and PLMD patients. Considering overall average accuracy, RAI and KEI presented comparable performances (around 50%).

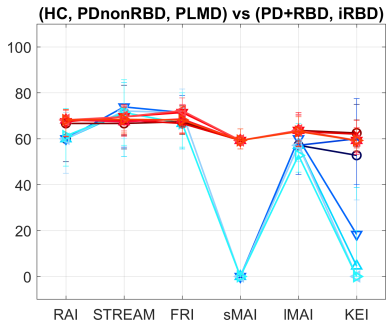
The results of the statistical analyses to evaluate the influence of MA related to apneas and arousals on the validation performances are shown in Table 3.5. None of the p-values was significant, thus indicating that the average values of sensitivity, specificity and accuracy did not change significantly across the different configurations in all comparisons. Only lMAI showed p-values close to significance for sensitivity and accuracy.

3.3.3 Discussion

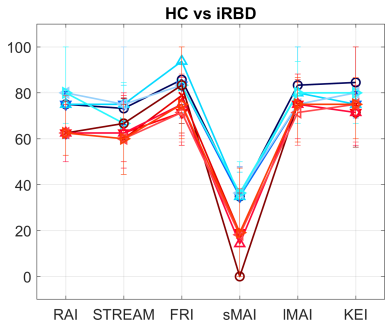
The presented analysis has two main outcomes: i) the performances of the methods vary depending on the comparison considered, thus none of the automated methods can be considered *the optimal* one for automatic RBD identification, and ii) the classification performances are not influenced by MA related to apneas and arousals.

Computation of the automated indices

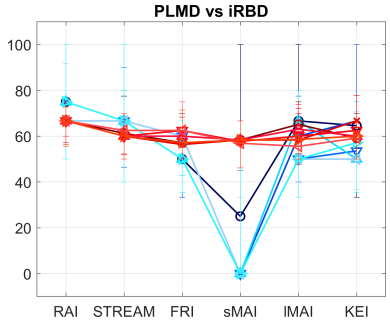
The automated indices were computed by implementing the algorithms according to the original descriptions (Table 3.2), with the only exception of filter settings. In fact, due to hardware setup, data analyzed in this work were filtered between 10 and 70 Hz, while RAI has been originally



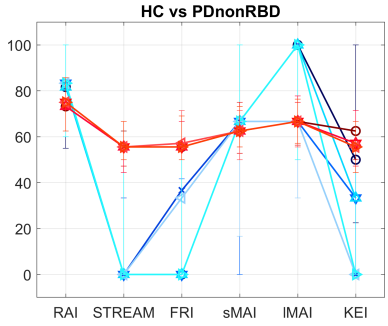
(a) Comparison 1



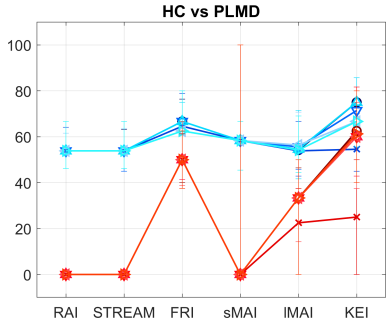
(b) Comparison 2



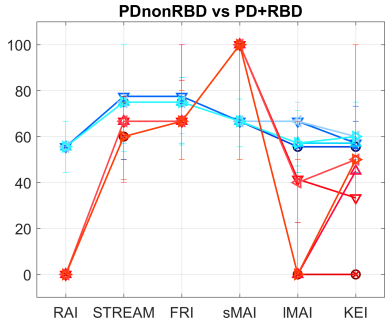
(c) Comparison 3



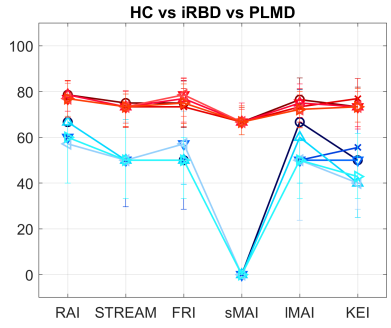
(d) Comparison 4



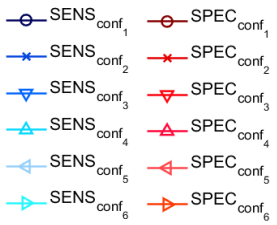
(e) Comparison 5



(f) Comparison 6

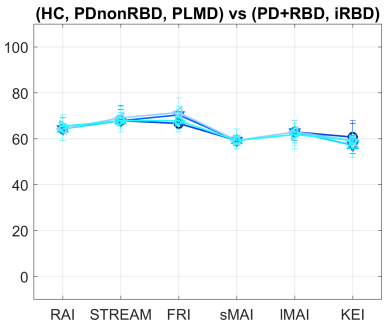


(g) Comparison 7

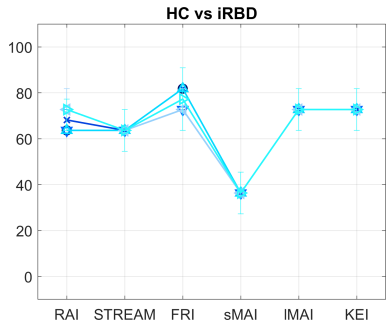


(h) Legend

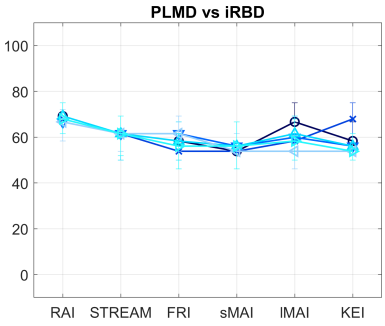
Figure 3.6: Sensitivity ($SENS$) and specificity ($SPEC$) values obtained during validation for each index and configuration for comparison 1-7 (a-g). $SENS$ and $SPEC$ values were calculated for the following classes: (a) (PD+RBD, iRBD), (b) iRBD, (c) iRBD, (d) PDnonRBD, (e) PLMD, (f) PD+RBD, and (g) iRBD. Values are shown as mean value with the whiskers indicating the 25^{th}-75^{th} percentiles through the 20 runs and 5 folds used in the classification scheme.



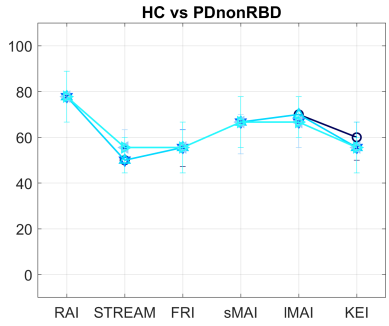
(a) Comparison 1



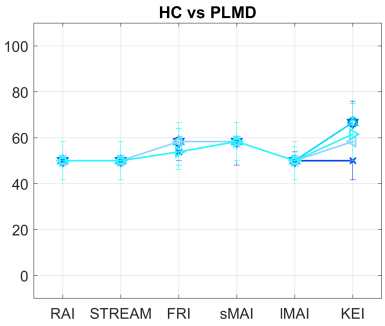
(b) Comparison 2



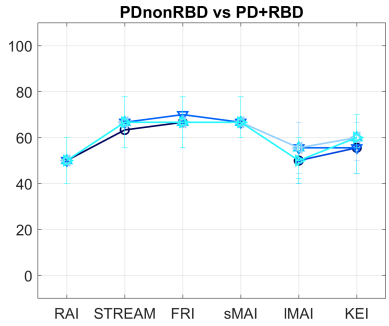
(c) Comparison 3



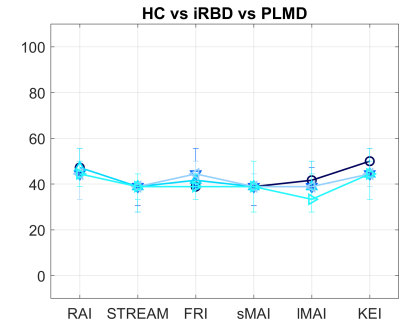
(d) Comparison 4



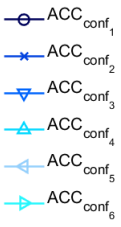
(e) Comparison 5



(f) Comparison 6



(g) Comparison 7



(h) Legend

Figure 3.7: Accuracy (ACC) values obtained during validation for each index and configuration for comparison 1-7 (a-g). Values are shown as mean value with the whiskers indicating the 25^{th}-75^{th} percentiles through the 20 runs and 5 folds used in the classification scheme.

implemented on signals filtered between 10 and 100 Hz [18, 19] and sMAI/lMAI on signals filtered between 10 and 120 Hz [21, 22]. However, it can be assumed that the different settings for filtering do not alter significantly the outcomes of the algorithms.

Another way for validating the implementation of the automated algorithms is the comparison between the values obtained in this cohort to the ones obtained in the original works in similar populations [18, 19, 20, 21, 23, 24, 25]. It was observed that the average values obtained in this cohort and the ones of the original publications were generally similar, with the only exception of sMAI (which had higher values in this cohort compared to the original study [21]). This could be caused by ECG artifacts in the chin EMG signal, which might be seen as short muscular activities. When the variance values were considered, higher variance values were generally obtained in this cohort compared to the original studies. Inter-clinical and inter-scorer variability, as well as artifacts such as snoring, could be the cause of this discrepancy. However, the generally high comparability of the indices computed in this cohort to the original studies is a good indication that the implementation was performed correctly.

Comparison of the automated methods

Despite the fact that it is not possible to identify the optimal automated method for RBD detection, FRI, KEI and RAI showed generally higher performances than the other methods. KEI and FRI have been developed in data recorded at DCSM, therefore a validation in other clinics is needed to further prove their robustness.

KEI had the highest performances in distinguishing PLMD patients from HC (comparison 5), but this might be simply the consequence of the inclusion of tibialis muscles in the computation of the index.

RAI is the most sensitive index for RBD detection, because of its high sensitivity in distinguishing iRBD patients from HC and PLMD patients (comparisons 2, 3 and 7). These results further confirm the usefulness of RAI in detecting RBD [19, 96, 105]. However, it should be noticed that RAI performed poorly in distinguishing PD+RBD from PDnonRBD (comparison 6) and could well distinguish HC from PDnonRBD patients (comparison 4), even if both these groups should not theoretically have RSWA. These results might be the consequence of RAI capturing MA related to RBEs, which are commonly seen in PD patients [14, 56]. This analysis gives the final picture that RAI is the most sensible method to detect RBD, but misclassifications could occur in presence of RBEs.

The average validation sensitivity and specificity reached by RAI, KEI and FRI were generally lower than 80% and varying across the comparison, thus suggesting that a more robust and generalized algorithm is needed.

Influence of arousal and apnea-related muscular activity

The p-values shown in Table 3.5 indicate that MA related to apneas and arousals does not influence the indices capability of distinguishing the participant groups and identifying patients with RBD. This result is particularly relevant considering that the groups differ significantly for the amount of apneas and arousals in REM sleep (Table 3.3).

This result seems to contradict the recommendations of removing arousal and respiratory related MA when applying an automated method [26, 95]. Only for the automated SINBAR method a systematic analysis of their influence has been performed, where it was shown that RBD could be detected with less accuracy when the MA related to apneas and arousals was not removed

[26]. However, the authors did not specify new cutoff criteria when apnea and arousal-related MA was not removed, but applied the same criteria as for signals where MA related to apneas and arousals was previously removed. In this study, different logistic regression models were trained for the different configurations. In this way, the logistic regression models could automatically learn patterns to best distinguish the participant groups, even in presence of MA related to apneas and arousals. This could explain the different outcome compared to [26].

Patients with severe sleep apnea were not included in this study. Severe sleep apnea is considered a confounder for RBD detection [106], therefore further studies should be performed to understand how the automated methods perform in patients with severe sleep apnea.

Limitations

This study has a number of limitations. First, the inclusion of the automated SINBAR method would have been valuable. Second, a larger sample size would have led to more solid results. Third, a preprocessing for snoring and ECG artifacts in the chin EMG signal has not been performed. Fourth, the length of the segments for exclusion of MA related to apneas and arousals has been chosen empirically, therefore some MA might still have been included and/or some relevant MA might have been excluded. Fifth, logistic regression was used, but it cannot be excluded that other classification techniques could have led to different results.

3.4 Conclusive remarks

The work and the results presented in Paper I [94] answer the research questions and address the objectives presented at the beginning of this chapter. Below, the research questions are reported again, together with the research outputs achieved in this chapter:

- ***Chapter research question 3.1:*** Which one(s) of the available automated methods can best identify patients with RBD?
  Chapter research output 3.1: None of the automated methods can be elected as *the optimal* one to identify patients with RBD. Among the methods investigated, FRI, KEI and RAI were generally the most accurate, with RAI having the highest sensitivity. However, the varying performances of these indices in different comparisons suggest that a more robust and generalized algorithm should be implemented.

- ***Chapter research question 3.2:*** To which extent does MA related to apneas and arousals influence the outcomes of the automated methods? Is removal of such MA necessary?
  Chapter research output 3.2: MA related to apneas and arousals does not influence the capability of the different automated methods in distinguishing between groups of participants and identifying patients with RBD. This is because the employed classification algorithm automatically learned MA patterns for best distinguishing participant groups, even in presence of arousal and apnea-related MA. This finding does not apply to patients suffering from severe sleep apnea, and more studies should investigate this matter.

From these research outcomes, it emerges the need of developing a more robust and generalized automated method to identify patients with RBD with higher performances. This will be the topic covered in the next two chapters.

Chapter 4

Data-driven identification of muscular activity

Chapter main objectives: The comparison of the currently available automated methods for RBD detection led to the conclusion that a new more robust and generalized algorithm for this purpose is needed. To achieve this goal, the first step consisted in the development of a data-driven method for identifying muscular activity from EMG signals.

This chapter aims at presenting the development and validation of a new data-driven method for muscular activity identification.

This chapter is based upon Paper II [115] and contributes in addressing thesis objective 2.

4.1 Research background

To develop a new automated method for detection of RBD from EMG signals, the first step was the development of an automated method for muscular activity (MA) detection. As a first proof of concept, this chapter presents the development and validation of this method on a cohort of HC and PLMD patients. PLMD patients have abnormal level of MA in the lower limbs in form of periodic limb movements (PLMs). According to AASM, limb movements (LMs) are defined as fast MAs lasting 0.5-10 s and a PLM series is defined when at least 4 LMs occur in intervals of 5-90 s [13]. Diagnosis of PLMD is established if more than 15 PLMs per hour of sleep occur (i.e. PLMS index $>$ 15), and if the patient complains of sleep onset and/or sleep maintenance problems [60]. Therefore, because of the known increased MA level in anterior tibialis muscles, a cohort of PLMD patients and HCs was considered as a valid choice to develop a new MA detector.

The starting point for its development was a critique investigation of the currently available automated methods for RBD identification [18, 19, 20, 21, 22, 23, 24, 25, 26], LM detection [107, 108, 109, 110, 111, 112], and detection of phasic activity in anterior tibialis muscles [113, 114]. Almost all of them are based on *traditional programming* techniques consisting of a preliminary data analysis, from which an algorithm is developed. Such an algorithm simply consists on a set of rules programmed explicitly. In other words, the final algorithm is built on human-based empirical observations of the data and decisions are made to maximize the performances in correctly identifying patients with RBD [18, 19, 20, 21, 22, 23, 26] and to have high correlation and agreement with manually scored LM and PLM series [107, 108, 109, 110, 111, 112].

Compared to traditional programming techniques, data-driven methods based on *machine learning* do not need explicit set of rules, but data patterns are automatically learned during the training of a machine learning model. Machine learning models retain therefore the potential to learn patterns that are hidden to human eyes and, because of this data-driven approach, are potentially superior to traditional programming methods [27].

Until now, only one automated machine learning method for RBD identification [24, 25] and two for EMG phasic activity recognition [113, 114] have been proposed. Out of these methods, the last two aimed at recognizing only phasic activity, which constitutes a significant limitation. The one proposed by Kempfner et al. (i.e. KEI) [24, 25] is a data-driven approach to identify MA and RBD patients, by applying a one-class support vector machine. This method, however, is computationally heavy and does not provide an easily interpretable identification of MA. In fact, chin and tibialis muscles are combined to get a final MA score (see Table 3.2 for more details). From this final score, it is not possible to deduce in which areas of the different EMG signals MA is identified. The difficult interpretability is a significant drawback of the method, as clinicians cannot easily visualize automatically identified MA.

From this analysis, it emerges that improvements in this field can be achieved by developing a method based on machine learning techniques and by making it interpretable, to ensure its efficient use in clinical environment.

4.2 Research questions and objectives of this chapter

Based on this research background, the following research question was formulated:

- ***Chapter research question 4.1***: Can an easily interpretable data-driven method automatically identify limb MA in PLMD patients?

This research questions led to the definition of the following objective:

- ***Chapter research objective 4.1***: To develop a data-driven method that can recognize limb MA in PLMD patients. The method should also be easily interpretable to facilitate its use in clinical environment.

4.3 Paper II: Probabilistic data-driven method for limb movement detection during sleep

The schematic overview of the methodology here implemented is shown in Fig. 4.1. The proposed data-driven method is not aiming at identifying LMs as defined by AASM, but more in general MA, which is expected to be increased in the anterior tibialis muscles of PLMD patients compared to HCs.

By using the assumption of atonia during REM sleep in HCs, a probabilistic model of atonia was built. Based on this trained probabilistic model, segments of tibialis EMG in unseen HC and PLMD patients having low likelihood of being atonia were identified as containing MA. Features describing the coverage of MA during REM and NREM sleep were then used to distinguish HCs and PLMD patients with a machine learning model. This model acts as a final decision stage which identifies MA patterns that can best separate HC and PLMD patients.

4.3.1 Methods: Development and validation of a new data-driven muscular activity detector

The 27 HC and 36 PLMD patients described in Table 3.3 were used in this work and their tibialis left (TIBL) and right (TIBR) EMG signals considered for the analysis. Fig. 4.2 shows a more detailed overview of the implemented algorithm, which is composed by the steps explained below.

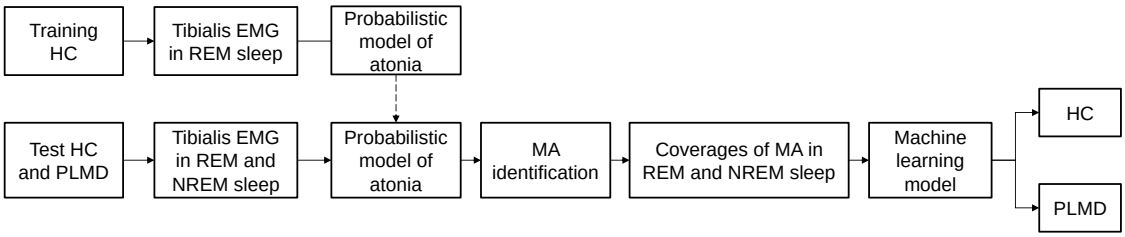
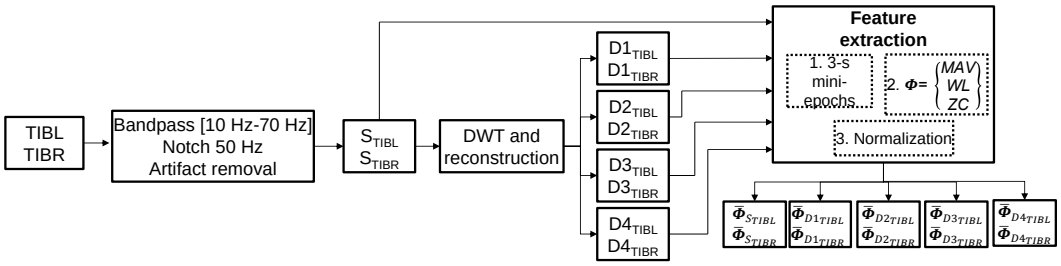
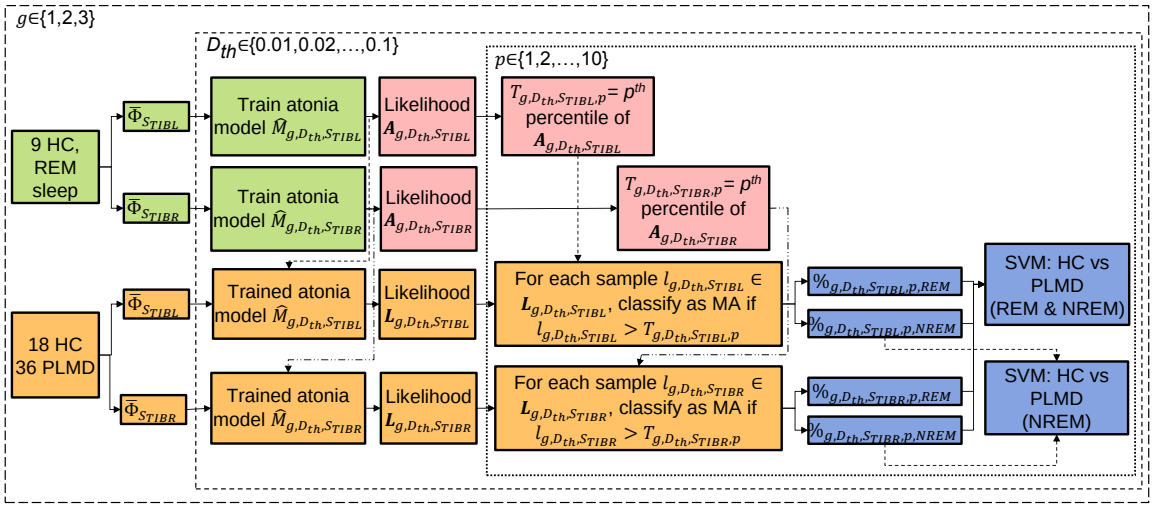


Figure 4.1: Overview of the methodology of Paper II [115]. Tibialis EMG signals in REM sleep from training HCs were used to build probabilistic models of atonia. By using such models, EMG areas in the tibialis muscles of test HCs and PLMD patients which had low likelihood of being atonia were labeled as muscular activity (MA). Based on the coverage of MA in REM and NREM sleep, HCs and PLMD patients were automatically distinguished with a machine learning classification algorithm.



(a) Feature extraction.



(b) Training and validation.

Figure 4.2: Schematic overview of the algorithm implemented: (a) pre-processing, discrete wavelet decomposition and reconstruction, and feature extraction; (b) framework used for classifying HC and PLMD groups based on the normalized features extracted from the signals S_{TIBL} and S_{TIBR} (the same approach was used for the other couples of signals): the green blocks show the training of the probabilistic models for defining atonia (one for each tibialis muscle); the red blocks show the calculation of the thresholds; the orange blocks show how the trained models were used to classify mini-epochs as either atonia or MA; the blue ones show the final classification of HC and PLMD based on the percentages of MA in REM and NREM sleep, or just in NREM sleep.

Pre-processing, discrete wavelet transform and feature extraction

The steps here described are shown in Fig. 4.2a. First, the TIBL and TIBR signals were pre-processed as described in Section 3.3.1, thus obtaining the signals S_{TIBL} and S_{TIBR}, respectively.

Biomedical signals are intrinsically non-stationary and, because of that, the wavelet transform (WT) is widely applied to them [116]. Moreover, WT is a popular technique for identification of muscular activity in EMG signals [114, 117]. Given a continuous signal $x(t)$, its WT is calculated as:

$$WT_{a,b}[x(t)] = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(\tau)\zeta\left(\frac{\tau - b}{a}\right) d\tau \tag{4.1}$$

where $\zeta(t)$ is the *mother wavelet*. The WT basically consists in the convolution of the signal $x(t)$ with a scaled (a) and translated (b) version of the mother wavelet. The WT achieves good time resolution at high frequencies and good frequency resolution at low frequencies [116].

The discrete version of the WT (DWT) is obtained by discretizing the scale and translation parameters. Usually, this is done with a dyadic sequence ($a = 2^k, b = 2^k l$ with k, l being positive integers). Mallat proposed a practical implementation of the DWT as a filter bank [118] (Fig. 4.3). More specifically, a signal $x(n)$ (with $n = \{1, 2, ..., N\}$) is passed through a series of high-pass ($h(n)$) and low-pass filters ($g(n)$) (derived from the corresponding mother wavelet and related with a quadrature mirror relationship), thus obtaining high and low frequency components, known as *details* (d) and *approximation* (a) coefficients, respectively. As at each filtration step half of the frequencies is removed, a downsampling can be performed according to Nyquist theorem. In DWT the detail coefficients are not further decomposed. By doing so, the DWT achieves good time resolution for high frequencies, which is a key factor to identify with precision muscular activity, as it is usually characterized by high frequency components in the range 5-450 Hz [119, 120].

It has to be mentioned that wavelet package analysis is a similar technique to DWT, where decomposition is performed also at detail level coefficients [121]. As the focus was not to identify the frequency bands characterizing MA with high precision, DWT was preferred to wavelet package analysis .

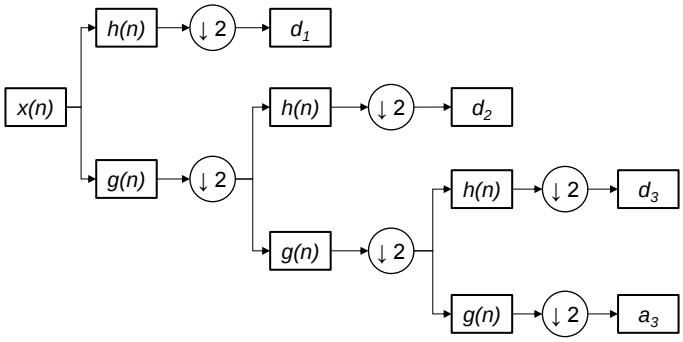


Figure 4.3: Implementation of the DWT as a filter bank. Three levels of decomposition are shown.

DWT was applied to the signals S_{TIBL} and S_{TIBR}, using the filters of the Daubechies 4 mother wavelet and the decomposition was made down to the 4^{th} level, thus obtaining level 1-4 detail coefficients (d_1-d_4) and level 4 approximation coefficients (a_4), corresponding to the frequency bands [64-128] Hz, [32-64] Hz, [16-32] Hz, [8-16] Hz and [0-8] Hz, respectively. The approximation coefficients a_4 were not considered for further analysis as their frequency content was previously removed with the bandpass filter between 10 and 70 Hz (Section 3.3.1). By considering one set

of detail coefficients at time, $D1$-$D4$ signals were obtained by applying inverse DWT. At the end of these steps, a total of 10 signals were available: S_{TIBL}, S_{TIBR}, $D1_{TIBL}$-$D4_{TIBL}$ and $D1_{TIBR}$-$D4_{TIBR}$.

Each signal was divided into 3-s mini-epochs, which is a time span often used in sleep analysis [24, 100], and the following features calculated for each mini-epoch $me(n)$, with $n = \{1, 2, ..., N\}$ and N representing the length of the mini-epoch:

- *Mean Amplitude Value* (MAV): the mean value of the absolute amplitude [122]:

$$MAV = \frac{1}{N}\sum_{i=1}^{N}|me(i)| \tag{4.2}$$

- *Waveform Length* (WL): a measure of complexity of a signal, defined as the sum of the signal absolute first derivative [122]:

$$WL = \sum_{i=1}^{N-1}|me(i+1) - me(i)| \tag{4.3}$$

- *Zero-crossings* (ZC): a measure of frequency information, defined as the number of times that the signal $me(n)$ crosses zero amplitude level [114, 122]:

$$ZC'(i) = \begin{cases} 1, & \text{if } me(i) \leq 0 \cap me(i+1) > 0 \\ 1, & \text{if } me(i) \geq 0 \cap me(i+1) < 0 \\ 0, & \text{otherwise} \end{cases} \quad ZC = \sum_{i=1}^{N-1} ZC'(i) \tag{4.4}$$

To compare the features across subjects, a normalization was performed. In particular, for each mini-epoch m (with $m = \{1, 2, ..., M\}$, and M representing the number of mini-epochs), the feature φ_m was normalized with respect to the previous and following 10 mini-epochs as:

$$\bar{\varphi}_m = \frac{\varphi_m}{\min(\varphi_{m-10}, \varphi_{m-9}, ..., \varphi_{m+9}, \varphi_{m+10}) + \epsilon} \tag{4.5}$$

where $\epsilon = 10^{-5}$ to ensure $\bar{\varphi}_m < \infty$. The normalization of the first and last mini-epoch was allowed by repeating them for 10 times at the beginning and at the end of the feature vectors, respectively. Overall, for each signal (e.g. S_{TIBL}), a normalized feature matrix (e.g. $\bar{\mathbf{\Phi}}_{S_{TIBL}}$) containing normalized values of MAV, WL and ZC for each mini-epoch was obtained.

Training of probabilistic models of atonia

The steps for training probabilistic models of atonia are shown in green in Fig. 4.2b. To build a model of atonia, the assumption that HCs show atonia during REM sleep was used and a non-parametric probabilistic approach was adopted, as a big training dataset was available and because of the easy interpretation of such an approach [123]. The 27 HCs were randomly partitioned into $G = 3$ subgroups of 9 subject each, and the features included in REM sleep of each subgroup were used for training. The normalized features extracted from each of the signals (e.g. $\bar{\mathbf{\Phi}}_{S_{TIBL}}$) were used separately for training different models of atonia.

A probability density was fit to the training data with the online Parzen window kernel density estimator (KDE) proposed by Kristan et al. [124]. Briefly, each new observation from the training data is modeled as a multivariate δ-Dirac distribution and the final distribution $p_{KDE}(\boldsymbol{\phi})$ is derived

as a mixture of multivariate Gaussian distributions and δ-Dirac distributions, which have different weights depending on their relative importance.

The overall algorithm can be divided into 3 steps (Fig. 4.4): (1) given a sample distribution $p_{Sample_{q-1}}(\boldsymbol{\phi})$, it is updated with the training observation $\boldsymbol{\phi}_q$; (2) the bandwidth for merging the distributions is calculated; and (3) the sample distribution $p_{Sample}(\boldsymbol{\phi})$ is finally obtained after refinition and compression by employing the calculated bandwidth. This compression consists in compressing together at maximum two components. In fact, it might turn out that a valid compression at one point in time could become invalid when a new training observation in considered. Therefore, to allow fast recovery from an early compression, this strategy is adopted. The final distribution $p_{KDE}(\boldsymbol{\phi})$ is obtained by convolution of the final distribution $p_{Sample}(\boldsymbol{\phi})$ with a Gaussian kernel having as covariance matrix the optimal bandwidth [124].

In other words, noting that a δ-Dirac distribution can be expressed as a Gaussian distribution with zero covariance, the final sample model on the feature space can be expressed as:

$$p_{Sample}(\boldsymbol{\phi}) = \sum_{i=1}^{N} \alpha_i \mathcal{N}_{\boldsymbol{\Sigma}_i}(\boldsymbol{\phi} - \boldsymbol{\phi}_i) \tag{4.6}$$

where

$$\mathcal{N}_{\boldsymbol{\Sigma}_i}(\boldsymbol{\phi} - \boldsymbol{\mu}) = (2\pi)^{-(d/2)} |\boldsymbol{\Sigma}|^{-(1/2)} \exp^{(-(1/2)((\boldsymbol{\phi}-\boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{\phi}-\boldsymbol{\mu}))} \tag{4.7}$$

is a Gaussian kernel centered at $\boldsymbol{\mu}$ with covariance matrix $\boldsymbol{\Sigma}$. Moreover, d is the dimension of the feature space, N the number of components and α_i is the weight of each Gaussian component i. The final distribution of the KDE is obtained as the convolution of the final distribution $p_{Sample}(\boldsymbol{\phi})$ with a kernel with covariance matrix (bandwidth) $\mathbf{H}$:

$$p_{KDE}(\boldsymbol{\phi}) = \mathcal{N}_H(\boldsymbol{\phi}) * p_{Sample}(\boldsymbol{\phi}) \tag{4.8}$$

In the algorithm, the bandwidth is automatically calculated, while the compression is regulated by the input parameter D_{th}: the higher D_{th}, the less compressed the model is.

The described algorithm was applied to all subgroups and all signals. In total, 300 models of atonia were trained $\hat{M}_{g,D_{th},s}$, were $g \in \{1,2,3\}$ represents the training subgroup, $D_{th} \in \{0.01, 0.02, ..., 0.10\}$ is the compression level and $s \in \{S_{TIBL}, S_{TIBR}, D1_{TIBL}, D1_{TIBR}, D2_{TIBL}, D2_{TIBR}, D3_{TIBL}, D3_{TIBR}, D4_{TIBL}, D4_{TIBR}\}$ is the signal considered. An example of a trained probabilistic model of atonia is shown in Fig. 4.5a.

Identification of muscular activity

To identify mini-epochs containing MA, the first step was the definition of a threshold from each trained probabilistic model of atonia (red blocks in in Fig. 4.2b). A likelihood vector $\mathbf{A}_{g,D_{th},s}$ was obtained as the likelihood of being atonia for the training data of the model $\hat{M}_{g,D_{th},s}$. The threshold $T_{g,D_{th},s,p}$ was defined as the p^{th} percentile of $\mathbf{A}_{g,D_{th},s}$. For each model, 10 percentile values ($p \in \{1,2,...,10\}$) were used.

For each test participant (i.e. 18 HC and 36 PLMD), the trained probabilistic atonia model $\hat{M}_{g,D_{th},s}$ was used to calculate the likelihood vector ($\mathbf{L}_{g,D_{th},s}$). Each sample $l_{g,D_{th},s} \in \mathbf{L}_{g,D_{th},s}$ was the likelihood of a 3-s mini-epoch of being atonia. For each percentile value p, each mini-epoch was classified as atonia if $l_{g,D_{th},s} \leq T_{g,D_{th},s,p}$ and as MA otherwise (Fig. 4.5b). These steps are shown in orange in Fig. 4.2b.

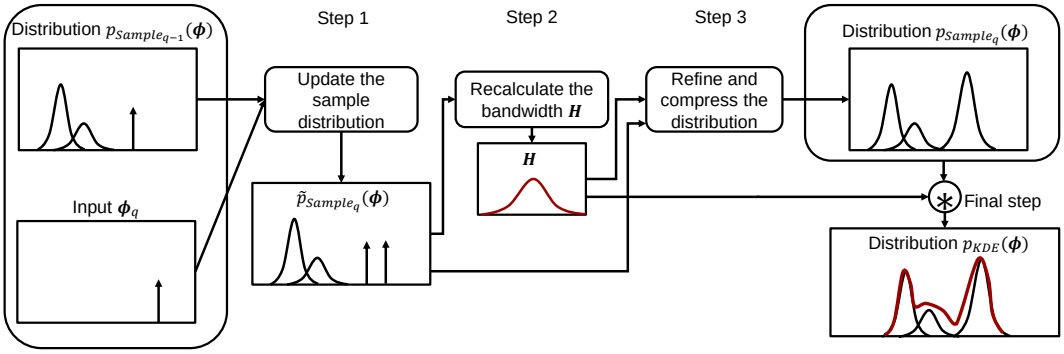
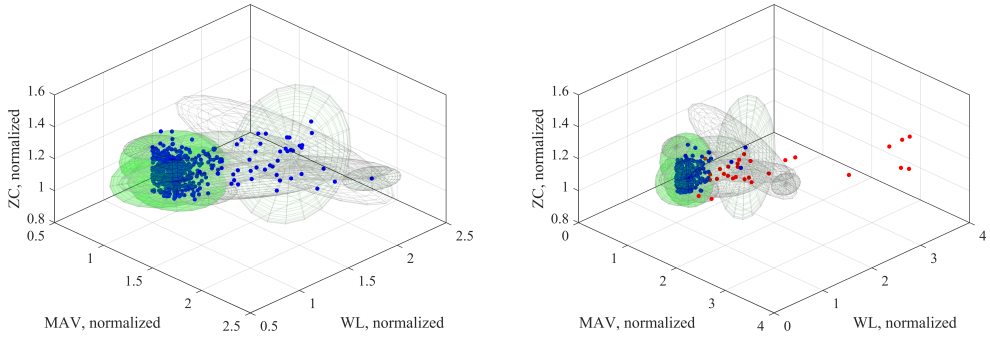


Figure 4.4: Three steps of the KDE algorithm, where the sample model $p_{Sample_{q-1}}(\phi)$ is updated with a new observation ϕ_q and finally compressed into the new sample model $p_{Sample_q}(\phi)$. The final model $p_{KDE}(\phi)$ is then obtained as convolution of the final sample model with a Gaussian kernel with bandwidth $\mathbf{H}$. Adapted from [124].



(a) Trained atonia model.

(b) Usage of the trained atonia model.

Figure 4.5: (a) Example of trained model of atonia. The blue points represent the training data and the Gaussian densities are indicated by drawing average value and 3 standard deviations. The green color of each Gaussian component is proportional to its weight in the model. (b) Example of how the trained model of atonia shown in Fig. 4.5a is used. The test mini-epochs that have likelihood lower than the defined threshold are classified as MA (red dots), while the others as atonia (blue dots).

Classification HC versus PLMD

The blue blocks in Fig. 4.2b show the classifications HC versus PLMD that were performed to test the validity of the MA detector. For each training group g, compression value D_{th}, signal s and percentile p, it was possible to calculate the percentages 3-s mini-epochs in REM and NREM sleep with MA for each test participant, $\%_{g,D_{th},s,p,REM}$ and $\%_{g,D_{th},s,p,NREM}$ respectively. For each combination of g, D_{th}, p, two classifications were performed as shown in Fig. 4.2b: one where both percentages of MA during REM and NREM sleep from left and right tibialis muscles were used (i.e. four features in total), and the other where only percentages of MA during NREM sleep from left and right tibialis muscles were used (i.e. two features in total). The second classification was made as it is known that the majority of limb movements in PLMD patients occur in NREM sleep [60]. For these classifications, support vector machine and Bayesian optimization were used as classifier and hyperparameter optimization method, respectively. An explanation of these methods is provided below and is followed by a final paragraph explaining how they were applied to distinguish HC and PLMD groups.

Support vector machine Among other binary classifiers, support vector machine (SVM) was chosen because it has proven to work well for small and moderately skewed training sets, because it is able to identify a global and unique solution to the classification problem and because it is robust to overfitting [102, 125, 126]. The following description of SVM is adapted from [102, 126].

Consider the following training set:

$$\{\boldsymbol{\phi}_i, t_i\}, \quad i = 1, ..., N, \quad t_i \in \{-1, 1\}, \quad \boldsymbol{\phi_i} \in \mathbb{R}^d \tag{4.9}$$

where N is the number of training samples, d the dimension of the feature space, $\boldsymbol{\phi}_i$ is the i^{th} feature vector and t_i the correspondent true class label, which can be either 1 if belonging to class c_1 and -1 if belonging to class c_2.

When the two classes are linearly separable and there is no overlap between the two class distributions (Fig. 4.6a), a hyperplane separating them can be identified as:

$$z(\boldsymbol{\phi}) = \mathbf{w}^T \boldsymbol{\phi} + w_0 \tag{4.10}$$

where $\mathbf{w}$ and w_0 can be found such that the hyperplane maximizes the *margin*, defined as the perpendicular distance between the decision boundary and the closest training data point (Fig. 4.6a). To find the best hyperplane, it should first be recalled that the distance between a point $\boldsymbol{\phi}$ and a hyperplane is given by $|z(\boldsymbol{\phi})|/||\boldsymbol{\phi}||$. Second, $\mathbf{w}$ and w_0 can be scaled so that the value of $z(\boldsymbol{\phi})$, at the nearest points in the two classes (defined as *support vectors* and circled in Fig. 4.6a), is equal to 1 for c_1 and to -1 for c_2. This means that the following conditions are required for getting the best hyperplane:

1. Having a margin of $\frac{1}{||\mathbf{w}||} + \frac{1}{||\mathbf{w}||} = \frac{2}{||\mathbf{w}||}$;
2. With: $(\mathbf{w}^T \boldsymbol{\phi} + w_0) \geq 1$, $\forall \boldsymbol{\phi} \in c_1$ and $(\mathbf{w}^T \boldsymbol{\phi} + w_0) \leq -1$, $\forall \boldsymbol{\phi} \in c_2$.

In other words, the best hyperplane can be found by minimizing the cost function:

$$J(\mathbf{w}, w_0) = \frac{1}{2}||\mathbf{w}||^2 \tag{4.11}$$

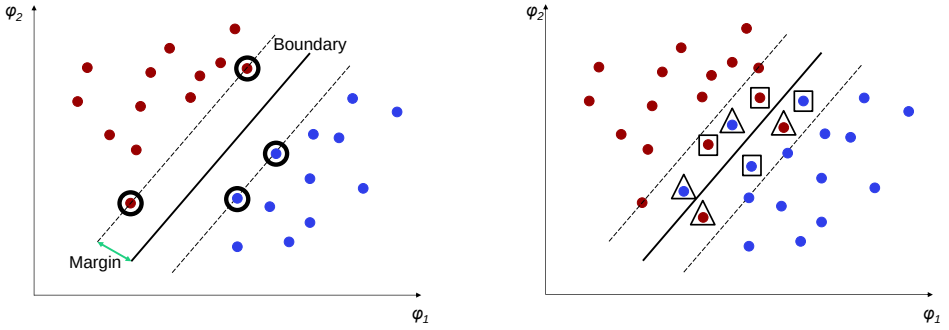
subject to

$$t_i(\mathbf{w}^T \boldsymbol{\phi}_i + w_0) \geq 1, \quad i = 1, ..., N \tag{4.12}$$

This is a quadratic programming problem, in which a quadratic cost function is minimized subject to a set of linearity constraints. These two conditions guarantee that any local minimum is also global and unique. By employing Lagrange multipliers and by using the Karush-Kuhn-Tucker conditions, this problem can be solved. When the trained SVM is evaluated on a test vector $\boldsymbol{\phi}'$, the sign of $z(\boldsymbol{\phi}')$ is evaluated and $\boldsymbol{\phi}'$ assigned to class c_1 if the sign is positive and to c_2 otherwise.

In the more general case in which the two classes have overlapping distributions (Fig. 4.6b), the training feature vectors belong to one of these three categories:

- Vectors that fall outside the bands and are correctly classified. These vectors satisfy $t_i(\mathbf{w}^T \boldsymbol{\phi}_i + w_0) \geq 1$;
- Vectors inside the bands that are correctly classified (placed in squares in Fig.4.6b). These vectors satisfy $0 \leq t_i(\mathbf{w}^T \boldsymbol{\phi}_i + w_0) < 1$;
- Vectors that are misclassified (placed in triangles in Fig. 4.6b). These vectors obey the inequality $t_i(\mathbf{w}^T \boldsymbol{\phi}_i + w_0) < 0$



(a) Non-overlapping classes. (b) Overlapping classes.

Figure 4.6: Schematic overview of how the SVM operates. (a) Linear boundary separating two non-overlapping classes. The support vectors are circled. (b) Linear boundary separating two overlapping classes. Squares represent vectors that are correctly classified but that lay inside the bands, triangles surround vectors that are misclassified. The figures show the case of a two-dimensional feature space (features φ_1 and φ_2). Adapted from [102] and [126].

By introducing the slack variables $\xi \geq 0$, the three inequalities before can be unified in:

$$t_i(\mathbf{w}^T \boldsymbol{\phi}_i + w_0) \geq 1 - \xi_i \tag{4.13}$$

where the first category of vectors correspond to $\xi_i = 0$, the second to $0 < \xi_i \leq 1$ and the third one to $\xi_i > 1$. In this context, the goal is to make the margin as large as possible, but also to reduced as much as possible the vectors with $\xi > 0$. From this, the separating hyperplane can be found by minimizing the cost function

$$J(\mathbf{w}, w_0, \boldsymbol{\xi}) = \frac{1}{2}||\mathbf{w}||^2 + \beta \sum_{i=1}^{N} \xi_i \tag{4.14}$$

subject to the condition in Eq. 4.13 and $\xi_i \geq 0$. The parameter β is a user-defined cost parameter (known as *box constraint*) indicating the penalty for misclassification. The problem is again convex and can be solved by using Lagrange multipliers, thus obtaining the values of $\mathbf{w}$ and w_0. After training, the SVM can be used on a test vector as described above.

Until now, the case in which a linear boundary was separating the classes was considered. However, sometimes a non-linear boundary is dividing the vectors of the two classes. In this case, a kernel is used to map the vectors into a space where they can be linearly separated. Among the kernels, the radial basis function (RBF) kernel $k(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j) = e^{\left(-\frac{||\boldsymbol{\phi}_i - \boldsymbol{\phi}_j||^2}{2\sigma^2}\right)}$ is widely used, where $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_j$ are two feature vectors in the training data.

Finally, SVM allows to obtain a classification score ($score(\boldsymbol{\phi})$) for each feature vector $\boldsymbol{\phi}$. This is calculated as:

$$score(\boldsymbol{\phi}) = \sum_j w_j < \hat{\boldsymbol{\phi}}_j, \boldsymbol{\phi} > + w_0 \tag{4.15}$$

where $w_j \in \mathbf{w}$ and w_0 are the estimated SVM parameters and $< \hat{\boldsymbol{\phi}}_j, \boldsymbol{\phi} >$ is the dot product between $\boldsymbol{\phi}$ and the support vectors $\hat{\boldsymbol{\phi}}_j$. The classification score ranges from $-\infty$ to $+\infty$ and can be converted to an estimation of the posterior probability for the positive class by applying the sigmoid function (Eq. 3.1). The posterior probability for the negative class is its complementary [127].

Bayesian optimization of hyperparameters When training a machine learning algorithm, the optimization of its hyperparameters is crucial to have a good and generalized classifier. A hyperparameter is defined as a classifier parameter that is set prior training. In the example of a SVM with RBF kernel, the two hyperparameters are the box constraint β and the standard deviation σ of the RBF function.

The problem of hyperparameter optimization can be formulated as follows [128]. Let $\mathcal{A}$ denote a machine learning algorithm with N hyperparameters and let Λ_n be the domain of the n^{th} hyperparameter. The overall hyperparameter space can be defined as $\mathbf{\Lambda} = \Lambda_1 \times \Lambda_2 \times ... \times \Lambda_N$. A vector of hyperparameters is $\boldsymbol{\lambda} \in \mathbf{\Lambda}$ and when the machine learning algorithm $\mathcal{A}$ is instantiated with $\boldsymbol{\lambda}$, it is denoted by $\mathcal{A}_{\boldsymbol{\lambda}}$. Given a finite dataset $\mathcal{D}$, the optimization consists in finding

$$\boldsymbol{\lambda}^* = \arg\min_{\boldsymbol{\lambda} \in \mathbf{\Lambda}} f(\mathcal{A}_{\boldsymbol{\lambda}}, \mathcal{D}_{train}, \mathcal{D}_{valid}) \tag{4.16}$$

where $f(\mathcal{A}_{\boldsymbol{\lambda}}, \mathcal{D}_{train}, \mathcal{D}_{valid})$ is the *objective function* measuring the validation loss of the machine learning algorithm $\mathcal{A}_\lambda$ trained on $\mathcal{D}_{train}$ and validated on $\mathcal{D}_{valid}$. In other words, the hyperparameters that minimize the validation loss are sought [128].

Because of the complexity of the problem, the objective function cannot be expressed analytically, but observations of this function are available. In this context, the optimization becomes a *black-box* problem. Grid search and random search of the best hyperparameter vector are commonly used, but have the limitation that they might be very time-consuming to find the optimal or near-to-optimal solution [128]. From a theoretical point of view, Bayesian optimization is now considered the state-of-the-art method for automatic tuning of hyperparameters for machine learning algorithms, but is main limitation is the time consumption that increases dramatically with high-dimensional problems [128]. As in the classifications here performed few hyperparameters had to be optimized, Bayesian optimization was selected.

For simplicity, the problem can be reformulated into maximizing the objective function $f(\mathbf{x})$. The core concept of Bayesian optimization is expressed in Algorithm 1 and in Fig. 4.7. More specifically, the dashed line in the three plots in Fig. 4.7 represent the *unknown* objective function $f(\mathbf{x})$. In the first plot, two observations of $f(\mathbf{x})$ are available, from which it is possible to estimate the posterior probability of $f(\mathbf{x})$ as a Gaussian process (GP), where the mean is represented by the solid line and the standard deviation by the coloured blue area around the mean. This estimation is used to calculate an acquisition function $u(\cdot)$, which is a trade-off between exploration (where the objective function is very uncertain) and exploitation (trying values where the objective function is expected to be high). The value maximizing the acquisition function is selected and the objective function sampled at that point. This observation is then used to update the GP (middle plot) and the process is repeated.

Algorithm 1: Bayesian optimization algorithm

Place a Gaussian process prior on $f(\mathbf{x})$;
Observe $f(\mathbf{x})$ at t_0 points according to an initial space-filling experimental design and set $t = t_0$;
while $t \leq T_0$ **do**
  By using Bayes' theorem, update the posterior distribution on $f(\mathbf{x})$ using all available data ;
  Let $\mathbf{x}_t$ be a maximizer of the acquisition function $u(\cdot)$, with the acquisition function computed by using the current posterior distribution ;
  Make the observation $f(\mathbf{x}_t)$;
  Increment t ;
end
Return as solution the point evaluated with the largest $f(\mathbf{x})$.

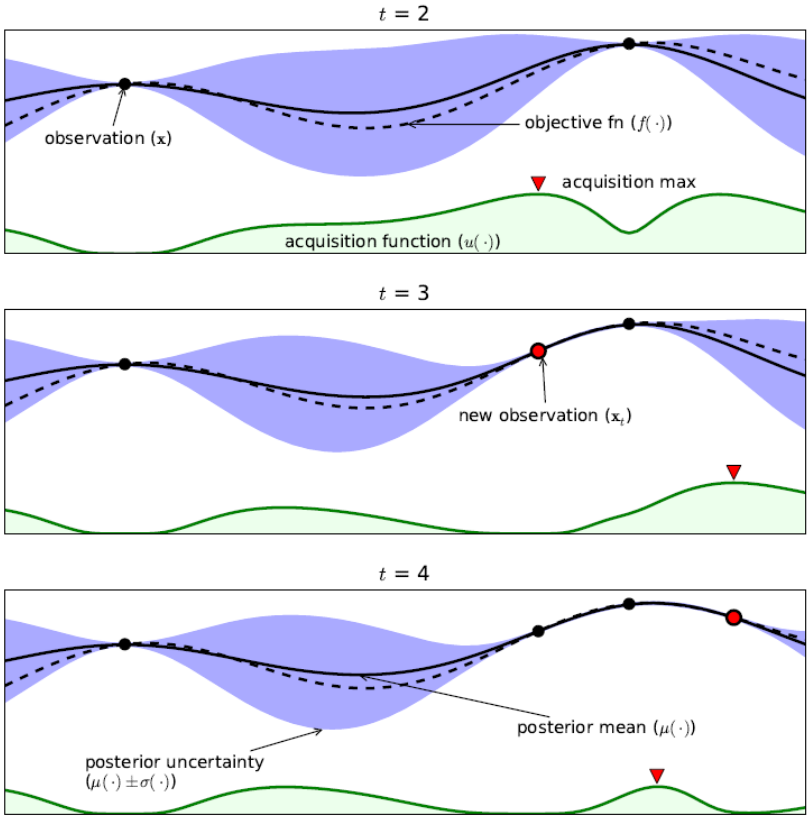


Figure 4.7: Schematic representation of Bayesian optimization algorithm. In the top plot, two observations of the objective function (dashed line) are known and they are used to estimate a GP approximation of the objective function with mean (solid black line) and standard deviation (blue area). The GP is used to estimate the acquisition function that is maximized to identify the new point where the objective function is sampled. The process is then repeated. Figure from [129].

To formalize mathematically this process, the core concept is the application of Bayes' theorem:

$$P(f(\mathbf{x})|\mathcal{O}_{1:t}) \propto P(\mathcal{O}_{1:t}|f(\mathbf{x}))P(f(\mathbf{x})) \tag{4.17}$$

meaning that the posterior probability of $f(\mathbf{x})$ given the observations of the objective function $\mathcal{O}_{1:t} = \{\mathbf{x}_{1:t}, f(\mathbf{x}_{1:t})\}$ is proportional to the likelihood of the observations $\mathcal{O}_{1:t}$ given the function $f(\mathbf{x})$ times the prior probability of $f(\mathbf{x})$.

More specifically, the prior of $f(\mathbf{x})$ can be specified as a GP:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \tag{4.18}$$

meaning that for a sample $\bar{\mathbf{x}}$, a normal distribution over the possible values of the objective function at $\bar{\mathbf{x}}$ is returned. Without loosing generalization, it can be assumed that $m(\mathbf{x}) = 0$ and several can be the choices for the kernel function. The one used here is the ARD Matérn 5/2 kernel, defined as [130]:

$$k(\mathbf{x}, \mathbf{x}') = \theta_0 \left(1 + \sqrt{5\theta_1^2(\mathbf{x}, \mathbf{x}')} + \frac{5}{3}\theta_1^2(\mathbf{x}, \mathbf{x}')\right) \exp(-\sqrt{5\theta_1^2(\mathbf{x}, \mathbf{x}')}) \tag{4.19}$$

The parameters θ_0 and θ_1 of the kernel function are also sought to be optimized [131].

By applying Bayes' theorem (Eq. 4.17) it is possible to show that the posterior probability of the objective function evaluated at sample $\mathbf{x}_{t+1}$ is:

$$P(f(\mathbf{x}_{t+1})|\mathcal{O}_{1:t}, \mathbf{x}_{t+1}) = \mathcal{N}(\mu_t(\mathbf{x}_{t+1}), \sigma_t^2(\mathbf{x}_{t+1})) \tag{4.20}$$

where $\mu_t(\mathbf{x}_{t+1}) = \mathbf{k}^T\mathbf{K}^{-1}f(\mathbf{x_{1:t}})$ and $\sigma_t^2(\mathbf{x}_{t+1}) = k(\mathbf{x}_{t+1}, \mathbf{x}_{t+1}) - \mathbf{k}^T\mathbf{K}^{-1}\mathbf{k}$, with $\mathbf{k} = [k(\mathbf{x}_{t+1}, \mathbf{x}_1)$, $..., k(\mathbf{x}_{t+1}, \mathbf{x}_t)]$ and $\mathbf{K} = \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_t) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_t, \mathbf{x}_1) & \cdots & k(\mathbf{x}_t, \mathbf{x}_t) \end{bmatrix}$. In other words, for each point $\mathbf{x}_{t+1}$, it is possible to obtain an estimation of the posterior probability of $f(\mathbf{x}_{t+1})$, based just on previous observations.

Thanks to this estimation, it is possible to formulate the acquisition function $u(\cdot)$ to identify the point to be evaluated next. Several acquisition functions have been proposed, but the one used in this work is the popular *expected improvement*. In particular, the improvement is defined as:

$$I(\mathbf{x}_{t+1}) = \max\{0, f(\mathbf{x}_{t+1}) - f(\mathbf{x}^+)\} \tag{4.21}$$

meaning that $I(\mathbf{x}_{t+1})$ is positive when the prediction at point $\mathbf{x}_{t+1}$ is higher than the best value known so far ($\mathbf{x}^+$), otherwise it is zero. The new query point is found by maximizing the expected improvement:

$$\hat{\mathbf{x}}_{t+1} = \arg\max_{\mathbf{x}_{t+1}} \mathbb{E}(I(\mathbf{x}_{t+1})) \tag{4.22}$$

and it can be shown that the expected improvement for $\mathbf{x}_{t+1}$ can be expressed as:

$$EI(\mathbf{x}_{t+1}) = \begin{cases} (\mu_t(\mathbf{x}_{t+1}) - f(\mathbf{x}^+))\Psi(Z) + \sigma_t(\mathbf{x}_{t+1})\psi(Z), & \text{if } \sigma_t(\mathbf{x}_{t+1}) > 0 \\ 0, & \text{if } \sigma_t(\mathbf{x}_{t+1}) = 0 \end{cases} \tag{4.23}$$

with $Z = \dfrac{\mu_t(\mathbf{x}_{t+1}) - f(\mathbf{x}^+)}{\sigma_t(\mathbf{x}_{t+1})}$, $\psi(\cdot)$ denoting the normal probability density function and $\Psi(\cdot)$ the normative cumulative density function. The observation maximizing EI is chosen as the next query point.

Application of SVM and Bayesian optimization SVM with RBF kernel was chosen to perform the classifications HC vs PLMD previously described. More specifically, for each set of 18 HC and 36 PLMD patients, an *external* 5-fold cross-validation (CV) scheme (Fig 3.4) was implemented. For each training fold, the hyperparameters of the SVM were optimized with Bayesian optimization in an *internal* 5-fold CV. The objective function to minimize was the internal CV loss, defined as the average classification loss across the internal folds. This strategy is known as nested or two-level cross validation and is schematically shown in Fig. 4.8. The box constraint and the standard deviation of the RBF kernel were optimized in the range of positive values log-scaled in $[1e-3, 1e3]$.

For each external fold, the performances of the trained classifier were evaluated by means of training and validation accuracy, sensitivity and specificity (Eq. 3.8), calculated with respect to the PLMD class.

4.3.2 Results

Fig. 4.9 shows an example of detected MA and an example of the calculated features $\%_{g,D_{th},s,p,REM}$ and $\%_{g,D_{th},s,p,NREM}$ for $g = 1$, $D_{th} = 1.0$, $s = \{S_{TIBL}, S_{TIBR}\}$ and $p = 2$ is

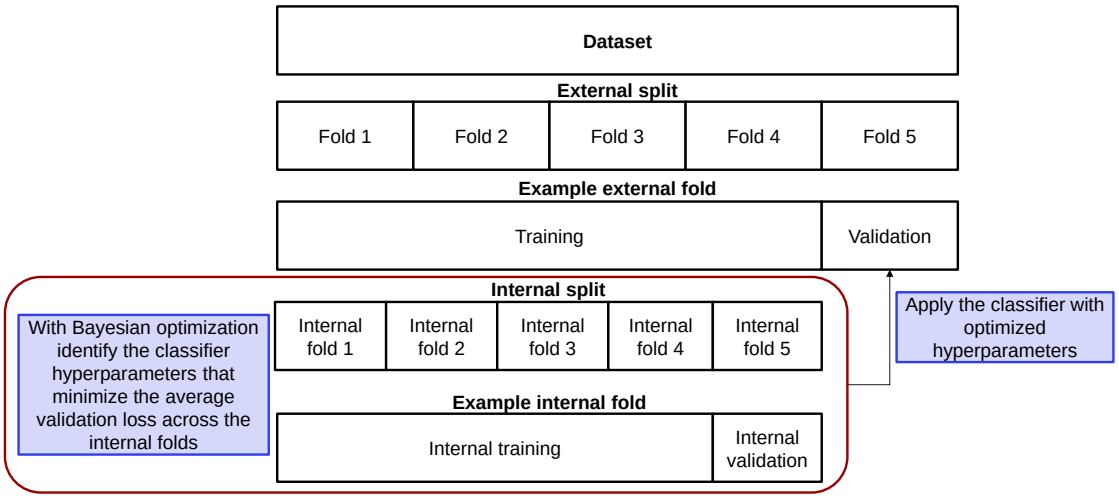


Figure 4.8: Schematic representation of a two-level 5-fold CV scheme. In the inner loop, the hyperparameters of the classifier are optimized with Bayesian optimization, where the objective function is defined as the cross-validation loss (i.e. the average validation loss across the folds).

shown in Fig. 4.10. PLMD patients presented increased percentages of MA in REM and NREM sleep compared to HC. Table 4.1 shows the combinations of D_{th} and p that led to the highest average validation accuracies, in both cases of considering or not REM sleep. The values of sensitivity, specificity and accuracy obtained during training and validation are shown as mean and standard deviation across the 5 folds of external CV and across the G training subgroups.

The highest validation accuracy (around 86%) was achieved for the couple of signals ($D1_{TIBL}$, $D1_{TIBR}$) when features calculated during both REM and NREM sleep were included. For the couples of signals (S_{TIBL}, S_{TIBR}), ($D1_{TIBL}$, $D1_{TIBR}$), ($D2_{TIBL}$, $D2_{TIBR}$) and ($D3_{TIBL}$, $D3_{TIBR}$) the average validation accuracy was always higher than 82%. The couple of signals ($D4_{TIBL}$,$D4_{TIBR}$) seems to be less informative, probably due to the intrinsic nature of high frequency of muscular activity [114]. The inclusion of REM sleep features improved only of about 2-4% the average accuracy values compared to including only NREM sleep features. This confirms that PLMD patients have increased muscular activity mainly during NREM sleep [60].

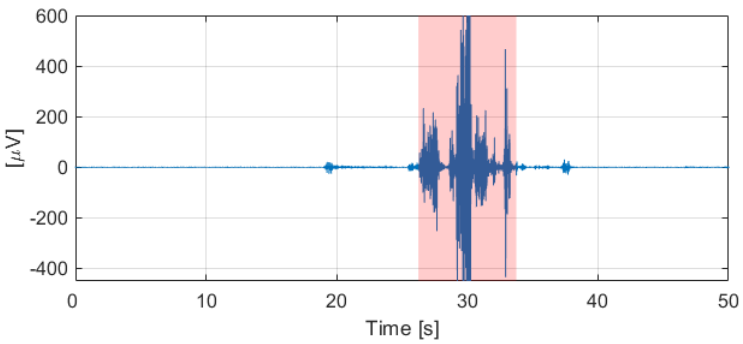


Figure 4.9: Example of detected MA. The parameters were $g = 1$, $D_{th} = 1.0$, $s = S_{TIBL}$ and $p = 2$.

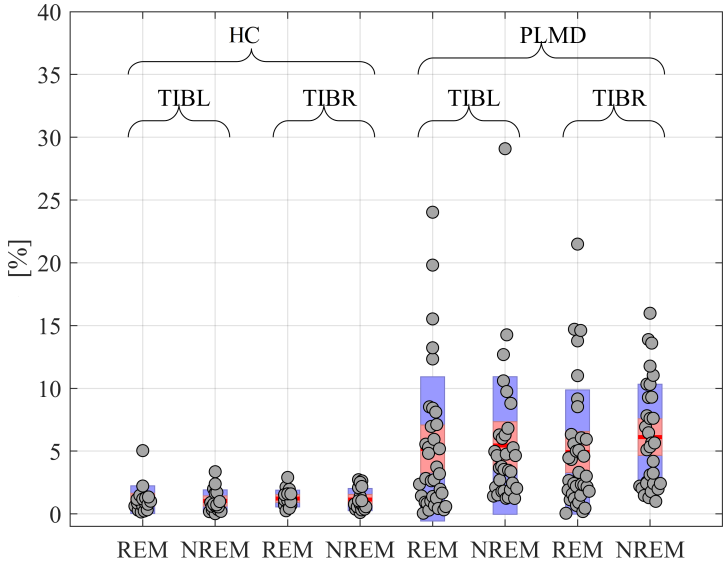


Figure 4.10: Percentages $\%_{1,1.0,S_{TIBL},2,REM}$, $\%_{1,1.0,S_{TIBR},2,REM}$, $\%_{1,1.0,S_{TIBL},2,NREM}$ and $\%_{1,1.0,S_{TIBR},2,NREM}$ obtained for the test HC and PLMD patients.

Signals	**Stage(s)**	**D_{th} p**	**Training**			**Validation**		
			ACC [-]	**SENS** [-]	**SPEC**[-]	**ACC** [-]	**SENS** [-]	**SPEC**[-]
S_{TIBL},S_{TIBR}	REM+NREM	0.04 2	0.93±0.00	0.94±0.01	0.92±0.02	0.84±0.06	0.86±0.11	0.80±0.03
$D1_{TIBL}$,$D1_{TIBR}$	REM+NREM	0.10 2	0.92±0.04	0.95±0.03	0.87±0.09	0.86±0.02	0.92±0.05	0.74±0.13
$D2_{TIBL}$,$D2_{TIBR}$	REM+NREM	0.04 1	0.95±0.02	0.95±0.00	0.95±0.05	0.84±0.03	0.86±0.05	0.80±0.07
$D3_{TIBL}$,$D3_{TIBR}$	REM+NREM	0.04 1	0.95±0.06	0.95±0.05	0.94±0.07	0.84±0.09	0.87±0.06	0.77±0.13
$D4_{TIBL}$,$D4_{TIBR}$	REM+NREM	0.04 2	0.93±0.02	0.93±0.03	0.93±0.04	0.80±0.06	0.85±0.08	0.71±0.11
S_{TIBL},S_{TIBR}	NREM	0.10 1	0.89±0.01	0.86±0.03	0.93±0.06	0.84±0.04	0.85±0.06	0.81±0.20
$D1_{TIBL}$,$D1_{TIBR}$	NREM	0.02 2	0.87±0.03	0.92±0.01	0.80±0.06	0.82±0.04	0.88±0.05	0.71±0.08
$D2_{TIBL}$,$D2_{TIBR}$	NREM	0.09 1	0.92±0.05	0.93±0.05	0.93±0.05	0.82±0.03	0.89±0.07	0.70±0.13
$D3_{TIBL}$,$D3_{TIBR}$	NREM	0.08 6	0.88±0.03	0.93±0.03	0.78±0.02	0.82±0.04	0.90±0.03	0.68±0.06
$D4_{TIBL}$,$D4_{TIBR}$	NREM	0.10 6	0.89±0.04	0.92±0.08	0.83±0.04	0.78±0.01	0.84±0.05	0.70±0.12

Table 4.1: Values of training and validation accuracy (ACC), sensitivity ($SENS$) and specificity ($SPEC$) for the classification of HC and PLMD groups expressed as mean and standard deviation across the 5 external folds and G HC subgroups. For each couple of signals, the pair (D_{th},p) leading to the highest validation accuracy is shown.

4.3.3 Discussion

This work proposes a data-driven method to identify MA. The method consists on building a probabilistic model of atonia from EMG signals recorded during REM sleep of HCs. From this, MA is identified as EMG areas having low likelihood of being atonia. The proposed method was used to identify MA in tibialis muscles of PLMD patients and the results show that PLMD patients could be successfully distinguished from HC, based on features describing the coverage of MA in REM and NREM sleep. Therefore, these results prove the feasibility and effectiveness of the proposed data-driven method for MA detection.

The best performances were obtained when the decomposition level $D1$ was considered. However, only slightly lower performances were obtained when the pre-processed signals (i.e. S_{TIBL} and S_{TIBR}) were used. This suggests that the initial wavelet decomposition could be avoided to reduce computational load.

The proposed approach is characterized by two innovative aspects. First, the definition of MA is not biased by human definition, but it is defined via a semi-supervised approach. In fact, the only assumption that is used is that HCs show atonia during REM sleep. Second, a machine learning approach was used to tune the definition of MA to distinguish with the highest possible performances PLMD patients from HCs.

This method does not aim to reproduce the scoring of LMs as defined by AASM [13], but proposes an alternative data-driven definition of MA. The proposed method is also easily interpretable, as it provides an easy visualization of the detected MA (Fig. 4.9) and on the total coverage of MA (Fig. 4.10).

Limitations

This study has a number of limitations. First, only 3-s mini-epochs were investigated as time window for MA detection; future studies should investigate whether better detection could be achieved with other window sizes and with overlap between windows. Second, no post-processing on the detected MA was performed, which could lead to increased performances. Third, only SVM with RBF kernel was used as classifier, but it cannot be excluded that other classifiers would have led to different classification performances. Fourth, the online kernel density estimator was used to build probabilistic models of atonia, because of its speed and proved robustness [124]; in the future, other probabilistic estimators could be tested to evaluate whether they could perform better in this context.

4.4 Conclusive remarks

The work and the results presented in Paper II [115] answer the research question and address the objective presented at the beginning of this Chapter. Below, the research question is reported again, together with the research output achieved in this Chapter:

- ***Chapter research question 4.1***: Can an easily interpretable data-driven method automatically identify limb MA in PLMD patients?
  Chapter research output 4.1: A new data-driven method to achieve this goal was developed. The new method automatically classifies each 3-s mini-epoch of tibialis muscle signals as having or not MA. Thanks to this, the proposed method is easily interpretable in a clinical environment. Moreover, the method does not achieve MA identification via implementation of human-based rules, but a two step data-driven approach was implemented. This consists on a first definition of a probabilistic model of atonia, followed by the refinement of MA detection with a machine learning approach. The achieved classification performances confirm the feasibility of the proposed approach for MA detection.

As the proposed approach for defining MA has proven its feasibility in distinguishing PLMD patients from HCs, it was employed (in a modified and expanded version) to identify patients with RBD. This topic will be discussed in the next Chapter.

Chapter 5

Development and validation of a data-driven method for RBD detection

Chapter main objectives: The previous chapter presented the proof of concept that the proposed data-driven method can successfully identify muscular activity.

This chapter aims to develop and validate on different cohorts a robust and generalized data-driven method for identification and characterization of patients with RBD.

This chapter is based upon Paper III [132] and Paper IV [133] and addresses thesis objective 2.

5.1 Research background

This chapter focuses on the development of a data-driven method for RBD detection and the methodology here presented is the expansion of the one described in the previous Chapter. To develop an automated method which is more robust and generalized than the ones described and compared in Chapter 3, some initial considerations were made.

First, among all the previously developed methods, only KEI was developed on a cohort including iRBD patients, HCs and patients with PLMD [24, 25]. All the other methods have been developed in cohorts including only HCs and iRBD patients [18, 19, 20, 21, 26, 23]. In the development of a robust method for identification of RBD patients, it is important to include also patients that might play a confounder, such as PLMD patients. In fact, these patients have increased limb motor activity, which is however not related to any neurodegenerative process.

The second consideration was about muscular activity (MA) during NREM sleep. Some studies have shown that patients with iRBD are characterized not just by RSWA, but also by abnormally increased muscular tone in NREM sleep [134, 135, 136]. Until now, none of the automated methods for RBD detection has analyzed NREM sleep and whether its inclusion could contribute to a better detection of RBD.

Thirdly, MA related to apneas and arousals was considered. The results presented in Chapter 3 have shown that previously developed methods were not influenced by MA related to apneas and arousals. However, in the development of a new method, this should be reconsidered.

The final consideration was that most of the previously developed methods for RBD identification have been validated only on data recorded in the same clinic where the method was developed. Only RAI has been validated in data recorded in several clinics [95, 96, 97, 101, 137, 138]. An automated method developed and validated only on data recorded in one clinic might be biased towards technical settings and equipment, environmental aspects, and how manual scoring is performed in

that particular clinic. Therefore, the validation of a method in data recorded in different clinics is fundamental to prove its robustness and generalization.

5.2 Research questions and objectives of this chapter

Based on this research background, the following research questions were formulated:

- ***Chapter research question 5.1:*** Can a new method based on a data-driven identification of MA overcome in performances previously proposed algorithms for RBD detection?
- ***Chapter research question 5.2:*** Can RBD be identified better if MA in NREM sleep is included?
- ***Chapter research question 5.3:*** Which is the influence of MA related to apneas and arousals on the performances of the new method?
- ***Chapter research question 5.4:*** Is the new method generalized? Does it perform equally well when applied on data that were recorded in a different clinic from the one where it was developed? If differences exist, can this method be used to evaluate inter-clinical differences?

This research questions led to the definition of the following objectives:

- ***Chapter research objective 5.1:*** To develop an automated method that, based on a data-driven approach, can overcome RAI, KEI and FRI in performances for RBD detection. These three methods are used for comparison as they have been shown to be the three best available methods for this purpose (Chapter 3).
- ***Chapter research objective 5.2:*** To investigate whether the inclusion of MA in NREM sleep leads to higher performances for RBD detection.
- ***Chapter research objective 5.3:*** To investigate the influence of MA related to apneas and arousals on the proposed method.
- ***Chapter research objective 5.4:*** To calculate the performances of the algorithm for RBD detection on data recorded in another clinic and to analyze the influence of inter-clinical differences on the proposed automated method.

5.3 Paper III: Validation of a new data-driven automated algorithm for muscular activity detection in REM sleep behavior disorder

An overview of the methodology of Paper III [132] is presented in Fig. 5.1. The main idea behind the proposed methodology is similar to the one presented in Chapter 4. Probabilistic models of atonia were derived from the chin and tibialis muscles of HCs during REM sleep. When a test subject (i.e. HC, iRBD or PLMD) was considered, a profile of MA probability ($p(MA)$) was obtained and MA was identified as areas having high values of $p(MA)$. The cases of including or not MA related to apneas and arousals, as well as the inclusion of MA in NREM sleep were considered. Features describing coverage and frequency of MA in REM (and NREM) sleep were used as input to a machine learning model which could classify a participant as either HC, iRBD or PLMD. The evaluation of classification performances of the different combinations (i.e. including or not MA related to apneas and arousals, and including or not MA in NREM sleep) led to define the best model for RBD identification.

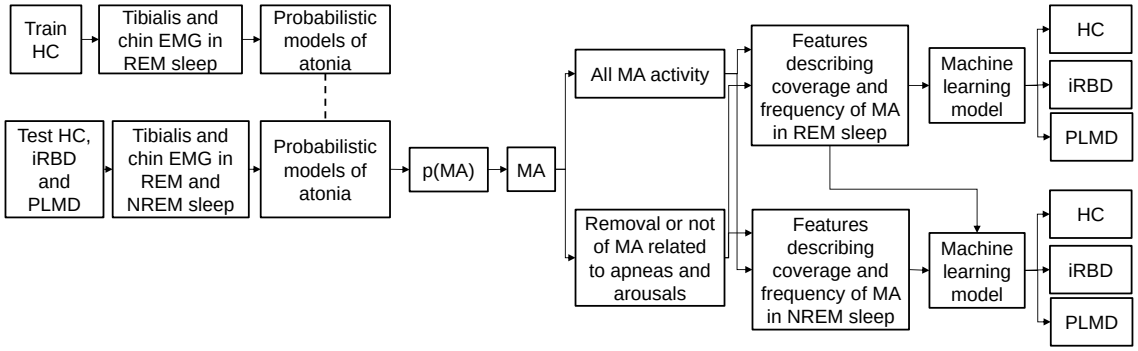


Figure 5.1: Overview of the methodology of Paper III [132]. Chin and tibialis EMG signals from REM sleep of HCs were used to define probabilistic models of atonia. When EMG signals from test HCs, iRBD patients and PLMD patients were given in input to such models, profiles of MA probability ($p(MA)$) were obtained. From such profiles, MA was then identified. By considering various scenarios of including or not MA related to apneas and arousals and including or not MA in NREM sleep, a machine learning approach was used to differentiate between HCs, iRBD patients and PLMD patients.

5.3.1 Methods: development and validation of the new data-driven algorithm to identify RBD patients

The participants included in this study were the cohort of 27 HC, 29 iRBD and 36 PLMD patients described in Table 3.3. The chin, tibialis left (TIBL) and right (TIBR) EMG signals were pre-processed with the same technique described in Section 3.3.1.

Feature extraction and normalization

After pre-processing, each EMG signal was divided into 1-s windows with 50% overlap and for each window the mean absolute amplitude value ($MAAV$) [122] was calculated as:

$$MAAV = \frac{1}{N}\sum_{i=1}^{N}|win(i)| \tag{5.1}$$

where $win(i)$ with $i = \{1, 2, ..., N\}$ are the EMG samples in each window.

Compared to the feature extraction reported in Section 4.3.1, there are three differences. First, the results presented in Section 4.3.2 showed that the application of the wavelet decomposition is not strictly necessary. Therefore, to reduce the computational load of the algorithm, only the pre-processed signal was used. Second, to further reduce the complexity of the algorithm, it was decided to implement only one feature. In this way, the time for training the probabilistic models of atonia could be significantly reduced. Finally, it was decided to calculate the features on 1-s windows with 50% overlap, and that is why a different nomenclature has been used compared to the previous Chapter. The window length of 1 s was chosen because it has been shown that phasic activity can be successfully identified in 1-s intervals [113], and because short bursts of phasic activity could be visually identified better with 1-s windows than with longer windows. Finally, 50% overlap of the windows ensured that short MA located at the edge of the windows was not smoothed down.

Feature normalization was performed to make the method robust to inter- and intra-subject variability. In particular, for each window n, the feature $MAAV_n$ was normalized with respect to

the previous and following 900 windows (i.e. 7.5 minutes of recording), according to:

$$MAAV'_n = \frac{MAAV_n}{\min\{MAAV_{n-900}, MAAV_{n-899}, ..., MAAV_{n+899}, MAAV_{n+900}\} + \epsilon} \tag{5.2}$$

where $\epsilon = 10^{-5}\mu V$ to ensure $MAAV'_n < \infty$ [24]. The normalization of the features for the windows located at the beginning and end of recording was allowed by repeating 900 times the first and last window features in the two extremities of the feature vector. Fig. 5.2 shows an example of pre-processed EMG signal, $MAAV$ features and their normalized values.

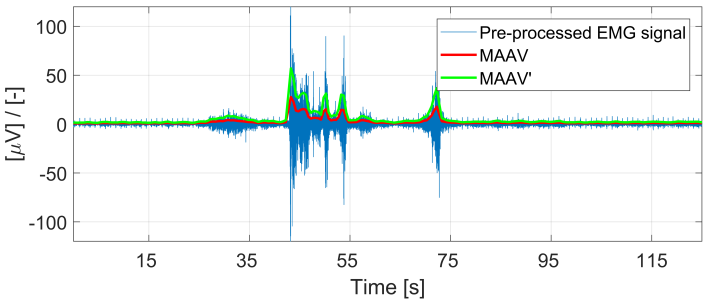


Figure 5.2: Example of pre-processed EMG signal, $MAAV$ features calculated for 1-s windows with 50% overlap (red) and respective normalized features (green).

Training of the probabilistic models of atonia

As in the previous study, the assumption that HCs are characterized by atonia during REM sleep was used to build probabilistic models of atonia. To make a more generalized method, the 27 HCs were divided into $G = 3$ random subgroups of 9 participants each. For each pre-processed signal (here referred as $s \in \{CHIN, TIBL, TIBR\}$) and for each subgroup $g \in \{1, 2, 3\}$, a probabilistic density was fit to the $MAAV'$ values in REM sleep by applying the online kernel density estimator introduced in Section 4.3.1. The levels of compression D_{th} ranged in $D_{th} \in \{0.01, 0.02, ..., 0.05\}$. Compared to Section 4.3.1, a smaller range of compression levels was used to have more compressed models (i.e. less Gaussian components). In this way, a total of 45 models of atonia $\hat{M}_{g,D_{th},s}$ (i.e. 15 for each muscle) were obtained, as shown schematically in Fig. 5.3. An example of an atonia model is shown in Fig. 5.4.

Muscular activity probability estimation

The aim of this step was to calculate the probability of MA ($p(MA)$) for each 1-s windows in the EMG signals of the test participants (i.e. the ones not included in the training phase of the probabilistic model of atonia). This is schematically explained in Fig. 5.5 for the chin signal, but the same approach was used also for the two tibialis EMG signals. More specifically, the training data used to build the model $\hat{M}_{g,D_{th},CHIN}$ were given as input to the model itself. A vector of likelihood $\mathbf{A}_{g,D_{th},CHIN}$ was therefore obtained, where the elements were the likelihood of the training data of being atonia. From this, ten values of threshold $T_{g,D_{th},CHIN,p}$ were derived as the p^{th} percentiles of the likelihood vector $\mathbf{A}_{g,D_{th},CHIN}$, with $p \in \{1, 2, ..., 10\}$. For each test participant, all the $MAAV'$ values derived from the chin signal were given as input to the model $\hat{M}_{g,D_{th},CHIN}$, thus obtaining a likelihood vector $\mathbf{L}_{g,D_{th},CHIN}$. For each sample i of the likelihood vector ($l^i_{g,D_{th},CHIN} \in \mathbf{L}_{g,D_{th},CHIN}$), and for each p, the probability of muscular activity

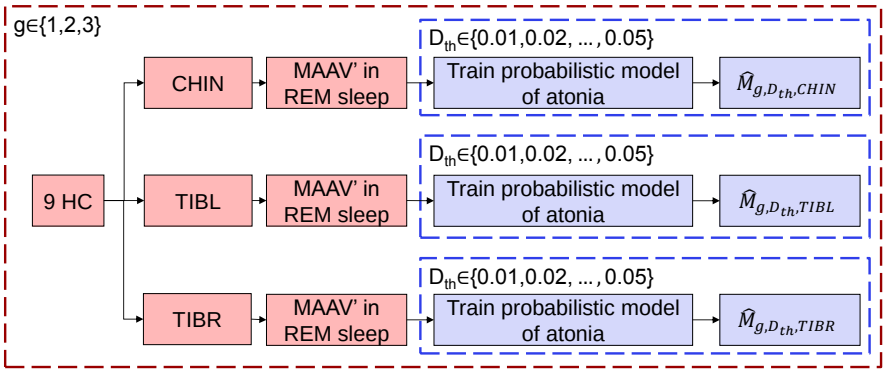


Figure 5.3: Schematic overview of how a total of 45 models of atonia were trained. For each training subgroup g, $MAAV'$ values in REM sleep of HCs were considered (red boxes). For each muscle, five models of atonia were then trained corresponding to five values of D_{th} (blue boxes).

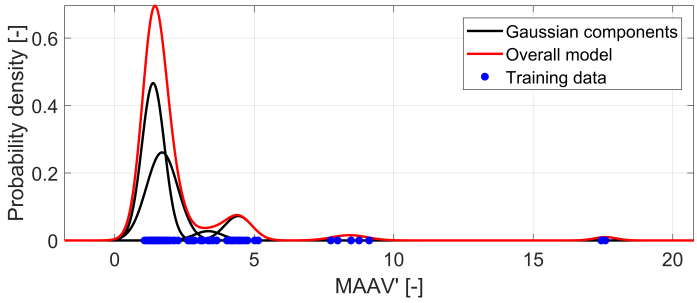


Figure 5.4: Example of trained probabilistic model of atonia from $MAAV'$ values obtained from REM sleep of HCs. The overall model (red) consists of a mixture of Gaussian distributions (black).

$p^i(MA)_{g,D_{th},CHIN,p}$ was calculated as:

$$p^i(MA)_{g,D_{th},CHIN,p} = \begin{cases} 0 & \text{if } (T_{g,D_{th},CHIN,p} - l^i_{g,D_{th},CHIN}) < 0 \\ 100 \cdot \dfrac{T_{g,D_{th},CHIN,p} - l^i_{g,D_{th},CHIN}}{T_{g,D_{th},CHIN,p}} & \text{otherwise} \end{cases} \tag{5.3}$$

Based on the trained model of atonia of Fig. 5.4, Fig. 5.6 shows the values of $p(MA)$ for nine $MAAV'$ test values. When $MAAV'$ test values are located in areas with high probability density, $p(MA)$ values are low and vice-versa. Fig. 5.7 shows a profile of probability of MA, corresponding to values of $p(MA)$ calculated for each 1-s window.

In summary, for each iRBD and PLMD patient, a total of 150 profiles of MA probability $p(MA)_{g,D_{th},CHIN,p}$ were obtained for the chin muscle, which were generated from the 15 models $\hat{M}_{g,D_{th},CHIN}$ and the ten values p. For the HCs, only 100 profiles were available, as each HC was used once in the training of 5 atonia models. Similar considerations apply for the two tibialis muscles.

Muscular activity identification and feature extraction

From each of the obtained profiles of MA probability $p(MA)_{g,D_{th},CHIN,p}$, three scenarios were considered to identify MA. These are illustrated in Fig. 5.8. MA was identified if values of $p(MA)$ were above H, with $H \in \{30\%, 60\%, 90\%\}$. Clusters of MA closer than 1 s were merged as previously done in literature [21].

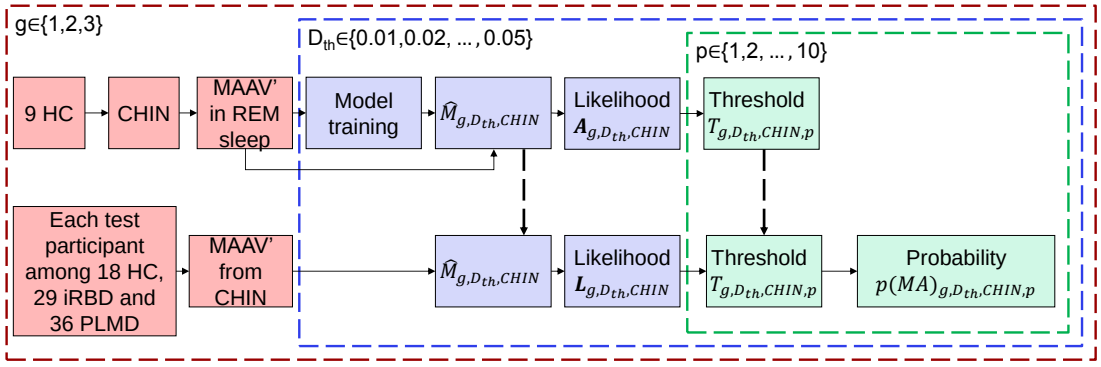


Figure 5.5: Schematic overview of how the probability of muscular activity was calculated for the test participants for the chin muscle. For each trained model of atonia $\hat{M}_{g,D_{th},CHIN}$ (see Fig. 5.4, a likelihood vector $\mathbf{L}_{g,D_{th},CHIN}$ was obtained by giving in input the training data of the model itself. The p^{th} percentile of such vector was used to define a threshold $T_{g,D_{th},CHIN,p}$ (green boxes). When $MAAV'$ values of a test subject were given in input to the atonia model $\hat{M}_{g,D_{th},CHIN}$, a profile of probability of muscular activity $p(MA)_{g,D_{th},CHIN,p}$ was obtained for each value of threshold $T_{g,D_{th},CHIN,p}$ (green boxes). The same was done for the two tibialis muscles.

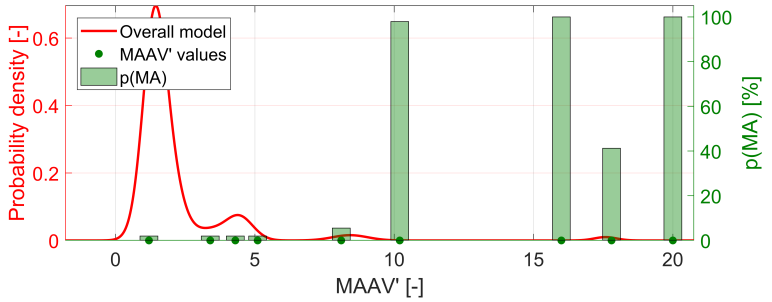


Figure 5.6: Illustration of nine $MAAV'$ and their muscular activity probability values ($p(MA)$) computed using the trained model of Fig. 5.4 (red curve).

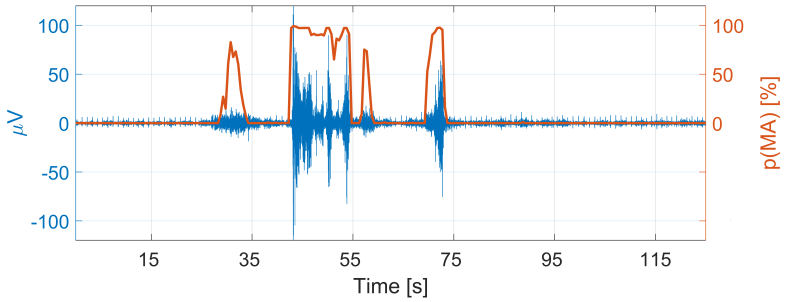
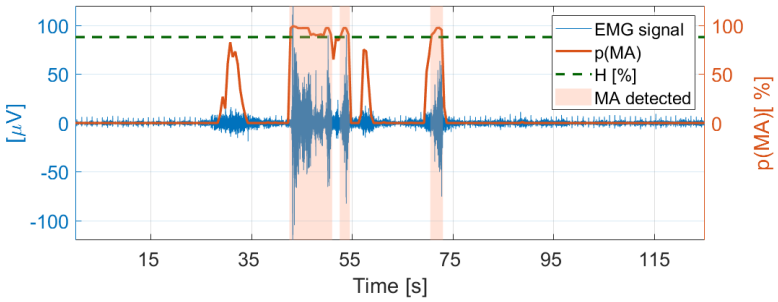
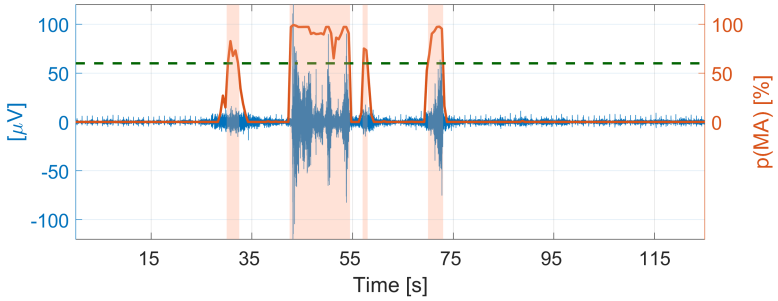


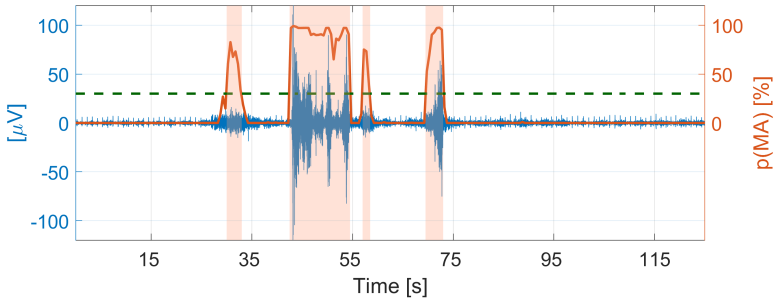
Figure 5.7: Example of an EMG signal (blue) and correspondent profile of MA probability ($p(MA)$ in red), spanning in the range 0-100%.



(a) $H = 90\%$



(b) $H = 60\%$



(c) $H = 30\%$

Figure 5.8: Identification of MA by applying different thresholds H. MA was identified when $p(MA)$ was exceeding the threshold and clusters of muscular activity closer than 1 s were merged.

After identification of MA, the following MA features (MAFs) were calculated (the same applies also to the two tibialis muscles):

- The percentage of 1-s windows during REM sleep containing MA ($\%_{g,D_{th},CHIN,p,H,REM}$);
- The percentage of 1-s windows during NREM sleep containing MA ($\%_{g,D_{th},CHIN,p,H,NREM}$);
- The median distance between onsets of MA during REM sleep ($D_{g,D_{th},CHIN,p,H,REM}$, measured in seconds);
- The median distance between onsets of MA during NREM sleep ($D_{g,D_{th},CHIN,p,H,NREM}$, measured in seconds).

Overall, for each iRBD and PLMD patient and for each combination of the parameters D_{th}, p and H, three *final* values of each MAF were obtained, corresponding to the $G = 3$ training subgroups. These values were averaged, so that a final value for each MAF was obtained in

correspondence to each combination of D_{th}, p and H. This is shown in Fig. 5.9 and the same applies also for the two tibialis muscles. For the HCs, a similar approach was used for extracting the *final* values of MAFs, but only two values were averaged for each combination of D_{th}, p and H, because each subject was used once for training of atonia models.

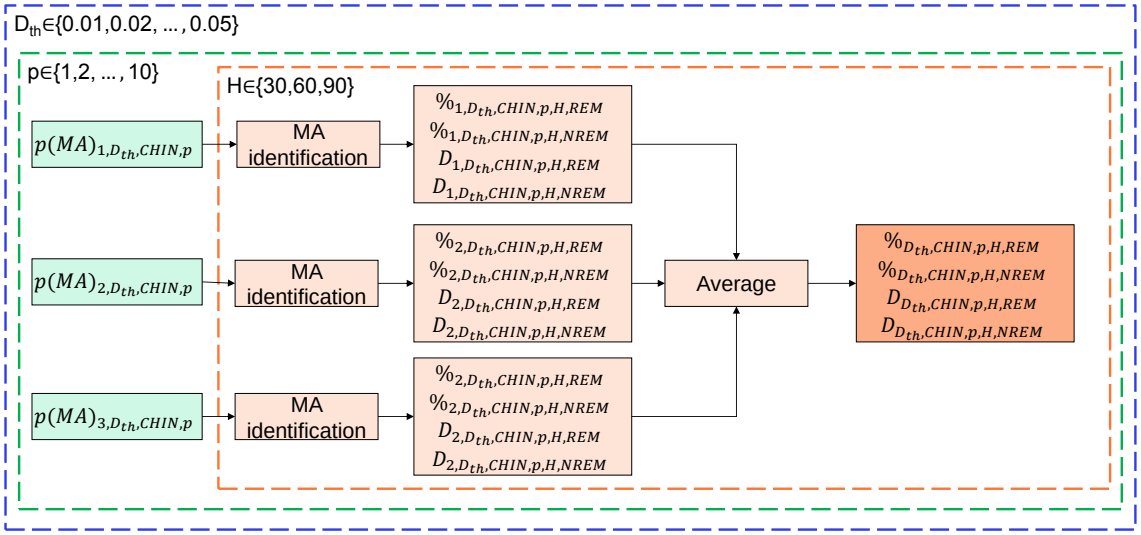


Figure 5.9: Schematic overview of the steps implemented for calculating the final MAF values for each combination of the parameters D_{th}, p and H and for the chin muscle, as the average values across the $G = 3$ training subgroups. This applies to iRBD and PLMD patients. For HCs, only two values were averaged for each combination of D_{th}, p and H, as each HC was used once in a training subgroup. The same approach was used also for the tibialis muscles.

Classification

In summary, for each combination of D_{th}, p and H, 12 *final* MAFs were obtained for each participant: $\%_{D_{th},CHIN,p,H,REM}$, $\%_{D_{th},CHIN,p,H,NREM}$, $D_{D_{th},CHIN,p,H,REM}$, $D_{D_{th},CHIN,p,H,NREM}$, $\%_{D_{th},TIBL,p,H,REM}$, $\%_{D_{th},TIBL,p,H,NREM}$, $D_{D_{th},TIBL,p,H,REM}$, $D_{D_{th},TIBL,p,H,NREM}$, $\%_{D_{th},TIBR,p,H,REM}$, $\%_{D_{th},TIBR,p,H,NREM}$, $D_{D_{th},TIBR,p,H,REM}$ and $D_{D_{th},TIBR,p,H,NREM}$. Fig. 5.10 shows how these features were used to build a machine learning system aiming to differentiate HC, iRBD and PLMD groups. For each combination of D_{th}, p and H, a 5-fold cross-validation (CV) scheme was applied for two different classifications: one where REM and NREM features were employed, and the second where only REM features were used. In both scenarios, each training fold was used to train one support vector machine (SVM) with linear kernel, one SVM with radial basis function (RBF) kernel and one random forest (RF) classifier. For the validation fold, the three trained classifiers were merged with Dempster-Shafer (DS) fusion method, thus obtaining a final classification. Before giving more details on the classification, an overview of the classifiers and the DS method is provided.

Multiclass SVM SVM is a binary classifier and to make SVM solve multi-class classification problems a number of SVMs needs to be combined. In this work, the combination was implemented with a *one vs one* approach. With this approach, if C classes are available, a total of $Q = C(C-1)/2$ SVM classifiers are trained, where in each of them one class is positive, another one is negative and

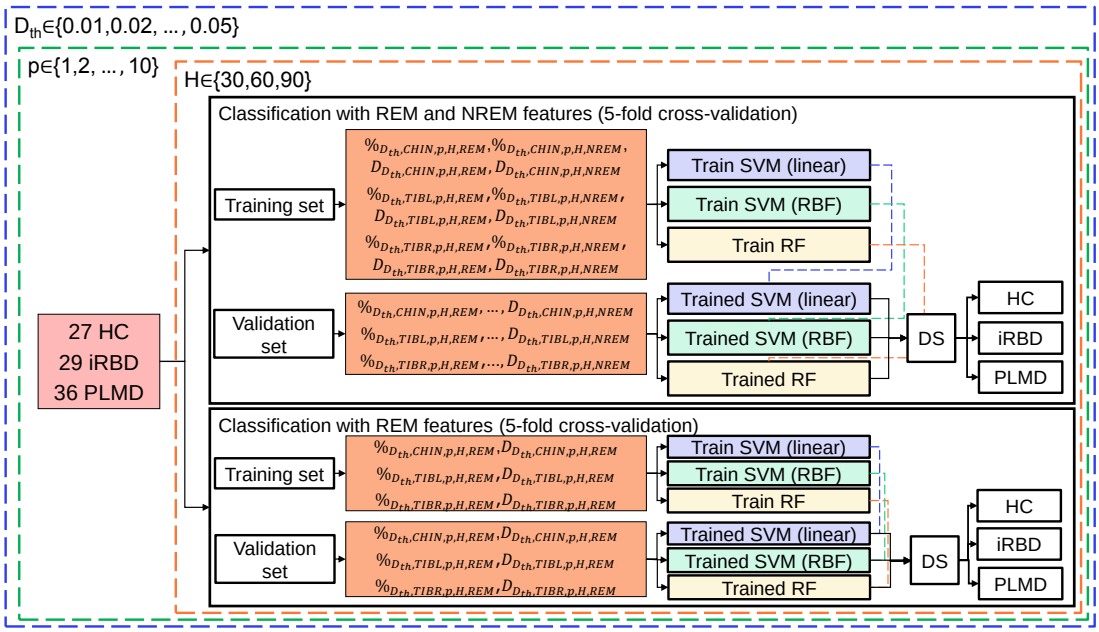


Figure 5.10: Schematic overview of the classification methodology adopted. For each combination of D_{th}, p and H, two classifications were performed, consisting in using only REM features and both REM and NREM features. In each of these two cases, 5-fold CV was employed. For each fold, three classifiers were trained using the training set and then applied to the validation set. The outputs from the classifiers prediction in the validation set were merged with DS technique, thus obtaining a final classification.

the rest is ignored. Considering $C = 3$ classes, the coding design CD is defined as

$$CD = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & -1 \end{bmatrix}$$

where the rows represent the three classes (c_1, c_2 and c_3) and the columns are $Q = 3$ SVMs. This means that the first SVM (first column) is trained with positive class c_1 and negative class c_2 etc. The class $\hat{c}$ that is chosen is obtained as:

$$\hat{c} = \arg\min_i \frac{\sum_{q=1}^{Q} |CD_{iq}| \hat{g}(CD_{iq}, score_q)}{\sum_{q=1}^{Q} |CD_{iq}|} \tag{5.4}$$

where CD_{iq} are the elements of the matrix CD (with i representing the i^{th} class and q the trained SVM), $score_q$ is the classification score of the q^{th} classifier (Eq. 4.15, mapped in the range [-1,1] with sigmoid function, Eq. 3.1) and $\hat{g}(\theta_0, \theta_1)$ is the quadratic binary loss, defined as:

$$\hat{g}(\theta_0, \theta_1) = \frac{1 - \theta_0(2\theta_1 - 1)^2}{2} \tag{5.5}$$

In this way, the trained SVMs are combined and the final prediction is obtained [139].

Random forest Random forest (RF) is a supervised machine learning ensemble method that is based on decision trees (DTs) [140]. The basic idea of DTs is to partition the feature space by recursively applying decision rules on the features. In particular, in the training phase of a DT, all the samples are considered in the root node of the tree, which is then split into sub-nodes by

applying a decision rule. This division is repeated until a node is *pure* (i.e. contains only one class), or when a specific stopping criterion is met. Such a node is called leaf node. A criterion that is often used is the *purity gain* Δ, defined as [141]:

$$\Delta = \Omega(root) - \sum_{b=1}^{B} \frac{\#(b)}{\#(root)} \Omega(b) \tag{5.6}$$

where $\Omega(root)$ is the impurity at the root node, $\Omega(b)$ is the impurity at the branch b (with $b = \{1, 2, ..., B\}$ and B representing the total number of branches), $\#(root)$ is the number of samples at the root and $\#(b)$ the number of samples at the branch b. Different impurity functions $\Omega(b)$ have been defined and a popular one is the Gini index, defined as:

$$Gini(b) = 1 - \sum_{i=1}^{C} p(c_i|b)^2 \tag{5.7}$$

where c_i with $i = 1, ..., C$ are the classes and $p(c_i|b)$ is the relative size of the class c_i at the branch b.

The purity gain is also used to identify the best split when growing a DT. This recursive way of growing a DT is known as Hunt's algorithm [142] and described in Algorithm 2.

Algorithm 2: Hunt's algorithm

```
Data: All samples contained in the root node
if stop criterion is met then
    Add a leaf node to the tree;
else
    Try different splits on the current branch and choose the split with the highest purity gain;
    Call the method recursively on the other branches;
end
```

RF is an ensemble method. This means that multiple DTs are built and their outputs are used to make a prediction. Because of this, ensemble methods are known to have increased performances compared to single classifiers [143]. In the context of RF, the final output is decided based on majority voting (i.e. the class that has been selected by most of the DTs is the final prediction) [141].

Another concept at the basis of RF is *bagging*. Bagging consists in considering a dataset $\mathcal{D}$ of size N, and then randomly select V new datasets $\mathcal{D}_1, ..., \mathcal{D}_V$ of size $N' \leq N$ by randomly subsampling $\mathcal{D}$ with replacement. This means that the same datapoint may occur multiple times in each $\mathcal{D}_v$ and that some other datapoints might be omitted (known also as *bootstrap*) [141].

In RF, bagging is applied to multiple DTs. Bagging produces V datasets and on each of them a DT is built, from which the final classification is obtained by majority voting. However, when this is done, a problem that may occur is that the DTs will often select the same splits, thus creating very correlated trees. This problem is overcome by considering only m of the M features available

($m < M$) to find the best decision rule to split a node [141]. This is also explained in Algorithm 3.

Algorithm 3: RF algorithm for classification [144].

Data: Dataset $\mathcal{D}$ of size N
for $v = 1 : V$ **do**
1. Take a sub-sampled data-strap $\mathcal{D}_v$ of $\mathcal{D}$ of size $N' < N$;
2. Grow a tree on $\mathcal{D}_v$
Iteratively go through the following steps until a stopping criterion is met
(a) Select m features randomly of the M features available
(b) Find the best split with the highest purity gain among m features
(c) Split the node into two daughter nodes

end
Take the majority vote of the ensemble of trees grown to make a prediction

An important property of RF is that it can provide a measure of the relevance of each feature in a trained model. This concept is known as *feature importance*. Gini feature importance is a popular method to calculate it [141]. For each node in a DT, the importance of node j for the feature i (FI_{ij}) is calculated as the difference between the decrease in impurity [145]:

$$FI_{ij} = w_j \Omega_j - (w_{left(j)} \Omega_{left(j)} + w_{right(j)} \Omega_{right(j)}) \tag{5.8}$$

where w_j is the weighted number of samples reaching node j, $left(j)$ and $right(j)$ are the the left and right child nodes coming from node j respectively, and Ω is the Gini impurity index (Eq. 5.7). From this, the importance of each feature i (FI_i) in a decision tree is calculated by summing FI_{ij} for each node j. Such a measure is then normalized to have the sum of the importance of all features equal to 1. This normalized value is then averaged through all the DTs of a RF, thus obtaining a final value of feature importance for each feature i.

As all classifiers, several hyperparameters should be optimized when training a RF. These include the number of trees, their depth, and how many features are selected randomly. Bayesian optimization (Section 4.3.1) can be used for choosing their optimal combination. In general, it is seen that by increasing the number of trees, the classification error stabilizes. However, a too high number of trees only increases the computational load. Additionally, when few features are randomly selected, the correlation between DTs is reduced and the performances of the RF usually increase. However, a too small number of features should not be chosen because of the risk of randomly selecting only irrelevant features, thus worsening the performances. Finally, too deep trees usually tend to overfit [144].

After training, a RF can predict the class of unseen data and can also return the posterior probability of each class. For each tree, the posterior probability of a class is calculated as the fraction of the observations of that class in the tree leafs. The final posterior probabilities are obtained by averaging these values over all the DTs [146].

In general, RF is a powerful classification method that is robust to outliers and noise, and it has shown good performances in literature [144].

Dempster-Shafer fusion method Dempster-Shafer (DS) method consists on a *fusion* of the outputs of different classifiers, to obtain a final prediction [143]. In the previous descriptions of SVM (Section 4.3.1) and RF, it has been pointed out that both classifiers, given in input a data sample for which the prediction is sought, can give in output the posterior probabilities of each class. In particular, given C classes and Q classifiers, the so-called *decision profile* for a feature

vector $\boldsymbol{\phi}$ can be derived as:

$$DP(\boldsymbol{\phi}) = \begin{bmatrix} p_{11}(\boldsymbol{\phi}) & \cdots & p_{1C}(\boldsymbol{\phi}) \\ \vdots & \ddots & \vdots \\ p_{Q1}(\boldsymbol{\phi}) & \cdots & p_{QC}(\boldsymbol{\phi}) \end{bmatrix} \tag{5.9}$$

where $p_{qj}(\boldsymbol{\phi})$ is the posterior probability of the classifier q for the class c_j for the input feature vector $\boldsymbol{\phi}$.

When the *training* data are considered, it is possible to calculate a decision profile for each training sample and to obtain a *decision template* for each class c_i, defined as:

$$DT_{c_i} = \frac{1}{N_{c_i}} \sum_{\boldsymbol{\phi}_{train} \in c_i} DP(\boldsymbol{\phi}_{train}) \tag{5.10}$$

This means that the decision template is simply is the average decision profile over the N_{c_i} training samples belonging to class c_i.

When a validation/test feature vector $\boldsymbol{\phi}_{test}$ is considered, its decision template $DP(\boldsymbol{\phi}_{test})$ can be calculated. Let now $DT^q_{c_i}$ be the q^{th} row of the decision template DT_{c_i} and $DP_q(\boldsymbol{\phi}_{test})$ the q^{th} row of the decision profile $DP(\boldsymbol{\phi}_{test})$. From this, it is possible to calculate the *proximity* $PR_{c_i,q}(\boldsymbol{\phi}_{test})$ of $DT^q_{c_i}$ to $DP_q(\boldsymbol{\phi})$:

$$PR_{c_i,q}(\boldsymbol{\phi}_{test}) = \frac{\left(1 + ||DT^q_{c_i} - DP_q(\boldsymbol{\phi}_{test})||^2\right)^{-1}}{\sum_{j=1}^{C} \left(1 + ||DT^q_{c_j} - DP_q(\boldsymbol{\phi}_{test})||^2\right)^{-1}} \tag{5.11}$$

From this, it is possible to compute the *belief* that the q^{th} classifier is correctly identifying $\boldsymbol{\phi}_{test}$ as belonging to class c_i:

$$bel_{c_i}(DP_q(\boldsymbol{\phi}_{test})) = \frac{PR_{c_i,q}(\boldsymbol{\phi}_{test}) \prod_{j \neq i} (1 - PR_{c_j,q}(\boldsymbol{\phi}_{test}))}{1 - PR_{c_i,q}(\boldsymbol{\phi}_{test}) \left[1 - \prod_{j \neq i} (1 - PR_{c_j,q}(\boldsymbol{\phi}_{test}))\right]} \tag{5.12}$$

After the belief values are calculated for each classifier q, they can be combined through Dempster's rule of combination:

$$\eta_{c_i}(\boldsymbol{\phi}_{test}) = \prod_{q=1}^{Q} bel_{c_i}(DP_q(\boldsymbol{\phi}_{test})) \tag{5.13}$$

This value is then normalized to ensure that the total support across the classes (i.e. final classification probability) is 1. The final prediction corresponds to the class c_i with higher value of η_{c_i} [143].

Application of the classifiers and Dempster-Shafer method As previously specified, classification was performed with a 5-fold CV scheme (Figs. 3.4 and 5.10). For each training fold, one linear SVM, one SVM with RBF kernel and one RF were trained and their hyperparameters optimized with Bayesian optimization (Section 4.3.1) in an inner 5-fold cross-validation loop (Fig. 4.8). For the linear SVM, the box constraint was the hyperparameter to be optimized and the search range consisted in positive values log-scaled in the range $[1e-3, 1e3]$. For the SVM with RBF kernel, the hyperparameters optimized were the box constraint and the standard deviation of the kernel (positive values log-scaled in the range $[1e-3, 1e3]$). For the RF, the hyperparameters optimized were the number of trees (search range being the positive integers, log-scaled in the range [10, 500]), and their depth (which consisted in the optimization of the maximum number of splits and the minimum leaf size, which were searched among the integers log-scaled in the

range $[1, max(2, N-1)]$ and $[1, max(2, \lfloor N/2 \rfloor))]$, respectively, with N being the number of training samples). For RF, the number of features to draw randomly was set to the square root of the total number of features. Moreover, for the two SVMs, the training features were standardized in the range 0-1. The parameters for standardization were employed in the validation set.

During validation, after standardizing the features, the trained classifiers were employed and merged with DS technique. From this, a final classification probability mixture for each participant was obtained (i.e. $p(HC)$, $p(iRBD)$ and $p(PLMD)$, corresponding to the probability of being HC, iRBD and PLMD patient respectively). A participant was classified according to the class with the highest probability. For each fold, the overall validation accuracy was computed as:

$$ACC_{overall} = \frac{P_c}{N_{tot}} \tag{5.14}$$

where P_c is the number of correctly classified participants and N_{tot} the total number of participants in the validation fold. Moreover, for each class i with $i = \{HC, iRBD, PLMD\}$, the validation accuracy, sensitivity and specificity were calculated as:

$$ACC_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \quad SENS_i = \frac{TP_i}{TP_i + FN_i} \quad SPEC_i = \frac{TN_i}{TN_i + FP_i} \tag{5.15}$$

where TP_i, TN_i, FP_i and FN_i are the number of true positives, true negatives, false positives and false negatives for each class i.

For each of the two classifications shown in Fig. 5.10, the combination of D_{th}, p and H leading to the highest average overall validation accuracy across the folds was selected as the *optimal* one.

Evaluation of the influence of apneas and arousals

To evaluate the influence of MA related to apneas and arousals, it was decided to train again probabilistic models of atonia and perform classification (as previously explained) after having removed MA related to manually labeled apneas and arousals. In particular, the segments removed were from 5s before to 5s after an apnea and from 3s before to 12s after an arousal onset, corresponding to configuration 5 in Section 3.3.1 (Fig. 3.2).

Comparison with other automated methods

To prove the robustness of the proposed method, its classification performances were compared to the ones achieved by RAI, FRI and KEI (which were found to be the automated indices achieving the highest performances for identifying RBD, as shown in Chapter 3). For each participant, the values of RAI, FRI and KEI were computed as described in Table 3.2. For each of these three indices, the same method as in Fig. 5.10 was applied, where the index values were features. The same was repeated after removing MA related to apneas and arousals.

5.3.2 Results

Table 5.1 shows the optimal combinations of D_{th}, p and H leading to the highest overall validation accuracy across the 5 folds in the 4 configurations here investigated:

1. Considering all MA during REM and NREM sleep;
2. Considering all MA during REM sleep;
3. Considering MA during REM and NREM sleep, after removal of MA related to apneas and arousals;

Sleep stages	Muscular activity	Optimal D_{th}	Optimal p	Optimal H
REM+NREM	All MA	0.01	3	90
REM	All MA	0.01	8	60
REM+NREM	Removed apnea and arousal-related MA	0.03	1	30
REM	Removed apnea and arousal-related MA	0.03	8	30

Table 5.1: Combination of the parameters D_{th}, p and H leading to the highest average validation overall accuracy across the 5 folds used in the 4 combinations analyzed.

4. Considering MA during REM sleep, after removal of MA related to apneas and arousals.

Fig. 5.11 shows the distribution across HC, iRBD and PLMD groups of the 12 extracted MAFs obtained for $D_{th} = 0.01$, $p = 3$ and $H = 90$ when all MA was included (i.e. first row of Table 5.1). The corresponding values are shown also in Table 5.2. As the median inter-muscular activity distance can be interpreted as the inverse of MA frequency, the values of Fig. 5.11 and Table 5.2 show that: i) iRBD patients have increased and more frequent MA in the chin and tibialis muscles than HC in REM sleep; ii) PLMD patients have increased and more frequent MA in NREM sleep in the tibialis muscles than HCs; iii) iRBD differ from PLMD patients because of the MA in chin during REM sleep. Similar feature distributions were obtained also for the other optimal combinations of D_{th}, p and H as shown in the Supplemental Material in Appendix C.

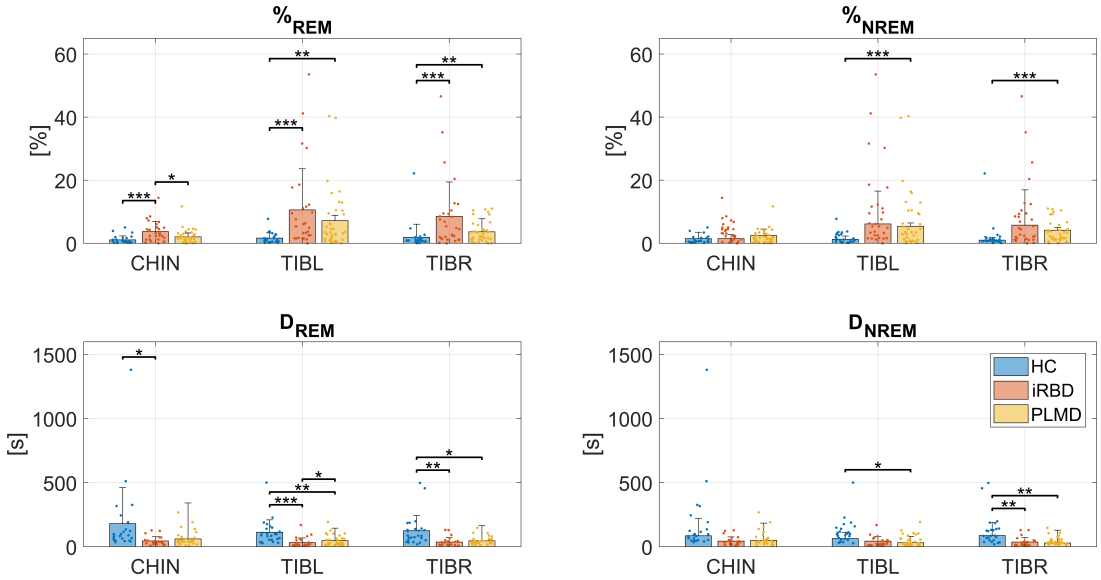


Figure 5.11: Distribution of the 12 MAFs obtained for $D_{th} = 0.01$, $p = 3$ and $H = 90$ when all MA was included. Each feature value from a subject is represented as a dot, together with bars representing the average group values and whiskers representing one standard deviation value for each group. Kruskal-Wallis tests were used to analyze group differences. In case of p-value<0.05, Wilcoxon rank sum tests were used for pairwise comparisons and corrected with Tukey-Kramer procedure. *: p-value<0.05; **: p-value<0.01; ***: p-value<0.001.

Table 5.3 presents the distributions of RAI, FRI and KEI across the three participant groups in the two cases of including or not MA related to apneas and arousals. The values of RAI and FRI showed significantly increased muscular tone in iRBD patients compared to HCs and PLMD patients. KEI showed significant increment in tone of iRBD and PLMD in comparison to HCs,

Feature	HC	iRBD	PLMD	Group	HC vs iRBD	HC vs PLMD	iRBD vs PLMD
$\%_{REM,CHIN}$	1.12±1.25	3.75±3.14	2.04±2.23	**<0.001**	**<0.001**	0.20	**0.03**
$\%_{REM,TIBL}$	1.68±1.69	10.61±13.03	7.21±9.59	**<0.001**	**<0.001**	**<0.01**	0.31
$\%_{REM,TIBR}$	1.93±4.15	8.53±10.91	3.66±3.39	**<0.001**	**<0.001**	**<0.01**	0.24
$\%_{NREM,CHIN}$	1.48±2.06	1.50±1.27	2.47±3.17	0.37	-	-	-
$\%_{NREM,TIBL}$	1.29±1.03	6.17±10.41	5.42±6.56	**<0.01**	0.08	**<0.001**	0.30
$\%_{NREM,TIBR}$	1.02±0.90	5.76±11.19	4.19±3.09	**<0.001**	0.06	**<0.001**	0.19
$D_{REM,CHIN}$ [s]	180.06±280.53	45.59±32.96	61.65±60.08	**0.01**	**0.01**	0.08	0.69
$D_{REM,TIBL}$ [s]	113.67±96.52	32.27±36.21	48.61±43.10	**<0.001**	**<0.001**	**<0.01**	0.05
$D_{REM,TIBR}$ [s]	124.66±117.90	36.17±35.71	45.98±30.87	**<0.01**	**<0.01**	**0.02**	0.71
$D_{NREM,CHIN}$ [s]	85.30±133.83	44.52±32.94	50.34±45.14	0.65	-	-	-
$D_{NREM,TIBL}$ [s]	64.79±48.75	44.38±35.74	31.05±14.74	**0.04**	0.17	**0.03**	0.81
$D_{NREM,TIBR}$ [s]	88.19±100.09	36.74±33.76	28.85±13.15	**<0.01**	**<0.01**	**<0.01**	0.99

Table 5.2: Values of the MAFs obtained for $D_{th} = 0.01$, $p = 3$ and $H = 90$ considering all MA activity. The values are shown as mean and standard deviation. Kruskal-Wallis tests were used to analyze group difference. In case of p-value<0.05, Wilcoxon rank sum test were used for pairwise comparisons and corrected with Tukey-Kramer procedure. Significant p-values are highlighted in bold.

MA	Index	HC	iRBD	PLMD	Group	HC vs iRBD	HC vs PLMD	iRBD vs PLMD
All MA	RAI [%]	87.39±21.04	54.12±37.26	83.76±24.95	**<0.001**	**<0.001**	0.84	**<0.001**
	FRI [%]	8.96±6.75	28.69±16.93	17.09±13.64	**<0.001**	**<0.001**	0.06	**0.01**
	KEI [%]	12.54±11.05	42.94±28.02	30.23±24.09	**<0.001**	**<0.001**	**<0.01**	0.12
No all MA	RAI [%]	89.27±19.39	55.99±38.12	83.79±27.52	**<0.001**	**<0.001**	0.97	**<0.001**
	FRI [%]	6.86±6.48	22.62±16.72	12.07±9.33	**<0.001**	**<0.001**	0.09	**0.02**
	KEI [%]	28.65±20.61	54.78±26.93	41.32±23.31	**<0.001**	**<0.001**	0.06	0.17

Table 5.3: Values of RAI, FRI and KEI shown as mean and standard deviation in the two cases of including all MA and removing MA related to apneas and arousals. Statistical analyses were performed as in Table 5.2.

when all MA was included, and only significantly increased tone in iRBD patients compared to HCs when apnea and arousal-related movements were removed.

Tables 5.4 and 5.5 show the classification performances of the proposed method, RAI, FRI and KEI (in both cases of including and excluding NREM features) in the two scenarios of considering all MA (Table 5.4) and removing apnea and arousal-related MA (Table 5.5). Generally, the proposed method achieved either similar or better performances than previously published methods in all the four combinations. The inclusion of NREM features led to generally similar or higher performances than in the case of their exclusion (Tables 5.4 and 5.5). Moreover, when comparing Tables 5.4 and 5.5, it can be seen that the inclusion of all MA generally led to increased performances when both REM and NREM sleep features were used. On the other hand, when only REM sleep features were used, the performances did not differ much, with the only exception of the sensitivity for RBD, which showed a drop of around 10% in the case where apnea and arousal-related MA was excluded.

For completeness of the results, Supplementary Tables in Appendix C show the average and standard deviation of all the performances across the 5 folds during training and validation of each single classifier and the final validation results obtained by applying DS method.

Performance	Proposed method (REM+NREM)	Proposed method (REM)	RAI	FRI	KEI
$ACC_{overall}$ [%]	70.76±7.75	61.93±10.45	46.67±8.63	45.61±7.00	48.95±10.49
ACC_{HC} [%]	81.52±8.20	79.18±12.12	69.42±8.90	65.09±12.75	77.29±24.81
$SENS_{HC}$ [%]	69.43±20.40	78.29±34.85	40.43±27.01	75.71±34.11	77.29±24.81
$SPEC_{HC}$ [%]	85.64±8.57	79.44±9.80	81.49±17.83	62.54±13.86	75.89±14.11
ACC_{iRBD} [%]	80.58±8.61	77.25±7.78	69.65±7.68	72.92±5.99	65.38±10.36
$SENS_{iRBD}$ [%]	73.38±10.16	69.86±18.41	72.90±4.95	66.52±11.97	47.95±24.11
$SPEC_{iRBD}$ [%]	84.04±13.37	81.18±17.45	68.56±9.53	76.43±12.82	73.77±24.98
ACC_{PLMD} [%]	79.42±8.67	67.43±11.59	54.27±9.66	53.22±10.08	55.44±10.56
$SENS_{PLMD}$ [%]	72.06±16.67	46.76±15.01	30.03±36.10	14.22±26.04	29.75±16.69
$SPEC_{PLMD}$ [%]	85.60±9.82	81.91±18.08	70.31±12.33	81.92±16.83	73.93±19.40

Table 5.4: Classification performances of the proposed method and RAI, FRI and KEI when all MA was considered. For the proposed method, the results are shown in both cases of including and excluding NREM sleep features (with the optimal combinations of parameters shown in the first two rows of Table 5.1).

Performance	Proposed method (REM+NREM)	Proposed method (REM)	RAI	FRI	KEI
$ACC_{overall}$ [%]	64.21±13.47	59.82±7.95	41.17±14.95	35.85±8.23	39.12±4.36
ACC_{HC} [%]	76.08±9.26	78.13±10.33	56.37±11.39	62.87±9.76	67.19±11.60
$SENS_{HC}$ [%]	72.43±31.86	82.29±26.68	30.43±23.24	57.38±30.51	51.43±7.76
$SPEC_{HC}$ [%]	76.49±14.54	76.23±10.43	66.23±16.26	65.82±21.27	73.56±17.26
ACC_{iRBD} [%]	78.42±13.58	75.09±5.69	68.59±9.77	67.54±7.74	63.22±13.48
$SENS_{iRBD}$ [%]	61.33±28.08	59.19±16.51	66.29±20.82	46.62±23.75	45.10±21.22
$SPEC_{iRBD}$ [%]	87.44±16.02	83.11±11.70	68.47±11.99	77.98±14.42	72.92±13.03
ACC_{PLMD} [%]	73.92±9.56	66.43±8.28	57.37±14.80	41.28±12.03	47.84±9.57
$SENS_{PLMD}$ [%]	61.87±11.78	43.87±8.70	25.11±18.39	17.08±24.84	29.27±10.94
$SPEC_{PLMD}$ [%]	82.51±11.51	80.93±11.21	76.59±21.72	61.73±24.89	61.52±20.85

Table 5.5: Classification performances of the proposed method and RAI, FRI and KEI when apnea and arousal-related MA was removed. For the proposed method, the results are shown in both cases of including and excluding NREM sleep features (with the optimal combinations of parameters shown in the second last two rows of Table 5.1).

Finally, since RAI, FRI and KEI were developed with the main aim of identifying RBD, it is particularly relevant to notice that the proposed method (in all the four combinations here presented) achieved sensitivity for iRBD identification in the same range of the previous methods and overcame them in accuracy and specificity. These performances were especially higher when all MA and NREM features were included.

5.3.3 Discussion

The proposed new data-driven method could estimate MA probability in sleep chin and tibialis EMG signals. From such probabilities, MA was defined and features describing the coverage and frequency of MA in REM and NREM sleep were extracted and used to successfully distinguish HCs, iRBD and PLMD patients. Three are the main outcomes of this study: i) the proposed data-driven method achieved higher performances in the overall classification accuracy of HC, iRBD and PLMD groups compared to previously proposed methods and in particular it overcame them

in the accuracy and specificity for iRBD detection; ii) the inclusion of features describing coverage and frequency of MA in NREM sleep led to increased classification performances when compared to considering only MA in REM sleep; and iii) the proposed method performed best when MA related to apneas and arousals during REM and NREM sleep was included in the analysis.

Comparison with other previously developed automated methods

Compared to RAI, FRI and KEI, the proposed method has some important methodological novelties. First, compared to RAI and FRI, the method is based on machine learning and not on traditional programming techniques, based on human-defined rules. Thanks to this, the proposed method identifies automatically patterns that contribute best to the differentiation of the participant groups. Second, compared again to RAI and FRI, the proposed method included in the analysis also tibialis muscles. This was done because limb muscular activity can lead to increased performances in RBD detection [100], and because tibialis muscles are expected to help in the distinction of HC from PLMD patients. Third, KEI, as the proposed method, uses a data-driven approach to identify atonia. However, the method used by KEI is computationally expensive and its output consists only in one index from the two tibialis and chin muscles (Table 3.2). Because of this, its clinical interpretation is difficult. The proposed method overcomes this drawback by highlighting areas of the EMG signals where MA is identified (Fig. 4.10). Finally, an important methodological innovation of the proposed method is that it provides a MA probability profile for each EMG signal. A MA probability profile is more informative than just a MA detector, as it is potentially possible to distinguish between major and minor MA.

Concerning classification performances, a fair comparison of the proposed method with RAI, KEI and FRI can be carried out by considering the accuracy, sensitivity and specificity for iRBD identification when only REM sleep was considered, as these three methods were developed for RBD detection based on RSWA levels. The proposed method achieved always higher average specificity and accuracy, and sensitivity in the same range as the other methods. The only exception is RAI, which achieved average sensitivity about 7% higher when apnea and arousal-related MA was excluded. However, the proposed method is generally as sensitive as previous methods for iRBD identification, but it is more accurate and specific.

Classification method

Concerning the classification method, three different classifiers with expected complementary information were used and combined with DS technique. This approach was implemented for two reasons. First, if only one of the three classifiers was used, it could not be guaranteed that such a classifier could perform well for RAI, FRI and KEI, and it could therefore introduce a bias beneficial only for the proposed method. Second, the combination of the three classifiers generally increased the overall performance of the classification (Supplemental Material in Appendix C). However, it cannot be excluded that using another classification technique would have led to different results from the ones here presented.

Muscular activity during NREM sleep

The accuracy, sensitivity and specificity for identifying iRBD patients were higher when MAFs extracted during NREM sleep were included (Tables 5.4 and 5.5). This means that, even if some of the single features were not statistically different (Table 5.2), the MA patterns in NREM sleep

contributed to better identify iRBD patients. Previous studies showed that iRBD patients have abnormal muscular activity in NREM sleep [21, 135, 134], and these results are in line with them. Moreover, a recent study has shown that RBD analysis can be successfully made by using actigraphy [147]. As no distinction between REM and NREM sleep can be made from actigraphy, it can be concluded that MA abnormalities in NREM sleep are present in iRBD patients.

These findings, taken all together, may lead to revise the definition of RBD and the guidelines for its diagnosis.

Influence of apnea and arousal-related muscular activity

In Chapter 3, it was found that the performances of RAI, FRI and KEI were not changing significantly when MA related to apneas and arousals were included or excluded. Some changes in the performances (due to the different classification technique used) can be seen in Tables 5.4 and 5.5. However, as a statistical analysis is not comprehensive here and based on previous findings, it can be concluded that these changes are not significant.

Tables 5.4 and 5.5 show that the proposed method performed best when MA related to apneas and arousals were not excluded. Therefore MA, traditionally considered as noise, indeed hold information that could help in distinguishing the groups when using data-driven methods.

Limitations

This study has some limitations. First, other window lengths for calculating $MAAV$ were not investigated. Second, artifacts such as snoring and ECG interference were not removed from the EMG signals. Third, the proposed method is not fully automated as the manual annotations of REM and NREM sleep are still required. However, making a good REM sleep detector working in iRBD patients is a challenging task [148, 149] and further research is needed for developing a robust fully automated system. Fourth, no participants with severe sleep apnea were included, therefore these results should be considered valid only for patients with up to moderate sleep apnea. Finally, EMG from arm muscles were not included. The SINBAR group has shown that their inclusion leads to higher performances [100], therefore a future development of this method should include them.

5.4 Paper IV: External validation of a data-driven algorithm for muscular activity identification during sleep

As stated at the beginning of this Chapter, most of the methods previously proposed in literature for identifying RBD have been developed and validated only on data recorded in the same clinic. The work presented in this Section aims to validate the new data-driven method on sleep data recorded in another clinic, namely Paracelsus-Elena Klinik in Kassel, Germany. The methodology of this work is schematically shown in Fig. 5.12. A similar cohort to the one used in Section 5.3 was included, consisting of: i) participants without RBD and with PLMS index lower than 15 PLMS per hour of sleep (RBD(-)PLMS(-), analogous to the HC group in Section 5.3), ii) participants with RBD (RBD(+), analogous to the iRBD group in Section 5.3), and iii) participants without RBD but with PLMS index higher than 15 PLMS per hour of sleep (RBD(-)PLMS(+), analogous to the PLMD group in Section 5.3). Each participant was given in input to the proposed data-driven method and the classification performances obtained in the German cohort were compared to the ones achieved in the Danish cohort presented in Section 5.3.2.

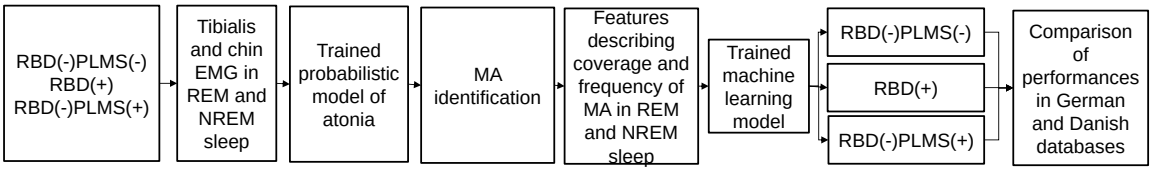


Figure 5.12: Schematic overview of the methodology of Paper IV [133]. The German participants were divided into three groups: RBD(-)PLMS(-) (similar to the the Danish HC group), RBD(+) (similar to the Danish iRBD group) and RBD(-)PLMS(+) (similar to the Danish PLMD group). The algorithm discussed in Section 5.3 was applied to them and the classification performances obtained in the German cohort were compared to the ones obtained in the original Danish cohort.

5.4.1 Methods: Validation of the new data-driven method on data recorded in another clinic and evaluation of inter-clinical differences

Subjects and recordings

A total of 240 elderly participants, part of the baseline evaluation of the "DeNoPa" cohort [14, 56, 150], were included in this study. The cohort included *de novo* PD patients, iRBD patients and sex-, age- and education-matched neurologically HCs. The study was approved by the local ethical committee (Landesärztekammer Hessen, Germany) in accordance with the Declaration of Helsinki and all participants signed informed consent for the scientific use of their data.

Two full-night v-PSGs were performed for each participant at Paracelsus-Elena Klinik, Kassel, Germany, and the second night was considered for analysis. In the rare cases of absence of recording or technical problems in the second night, the first one was considered for evaluation. Sleep experts manually scored sleep stages, respiratory events, RSWA and PLMS according to international standards [100, 151]. Sleep diagnoses were made according to the ICSD-2 international criteria [152]. All participants were drug-naïve for PD and RBD medications at the time of the PSG recording. The participants were grouped according to PD diagnosis (with PD [PD(+)] or without PD [PD(-)]), RBD diagnosis ([RBD(+)] or [RBD(-)]) and PLMS index ([PLMS(+)] for PLMS index $\geq$15 PLMS/hsleep and [PLMS(-)] for PLMS index <15 PLMS/hsleep). Using this grouping, a participant with no diagnosis of PD, no diagnosis of RBD and with PLMS index <15 PLMS/hsleep was thereby allocated to the group PD(-)RBD(-)PLMS(-). Table 5.6 shows the demographic and sleep information of the participants included in this study.

The PSG montage was set up according to standards [151] and included chin, TIBL and TIBR EMG signals, which were analyzed at a sampling frequency of 256 Hz. Because of different hardware used in this cohort compared to the one in Section 5.3, the strategy for detection of artifacts due to electrode pop-ups and detachments was modified. For each EMG signal, these artifacts were identified as areas where the EMG signal was above 1e4 μV or constantly 0 μV for at least 10 samples. After that, the same band-pass filters (with cut-offs at 10 and 70 Hz) were applied to the EMG signals as in Section 3.3.1.

Application of the data-driven method and group classification

The method developed in Section 5.3 was applied to each participant of the German cohort according to Fig. 5.13. The method was applied without removing any MA related to apneas and arousals and including both REM and NREM MAFs, as this was found to be the optimal configuration (Section 5.3.2). More specifically, $MAAV'$ values were extracted in 1-s windows

	PD(-) RBD(-) PLMS(-)	**PD(-) RBD(-) PLMS(+)**	**PD(-) RBD(+) PLMS(-)**	**PD(-) RBD(+) PLMS(+)**	**PD(+) RBD(-) PLMS(-)**	**PD(+) RBD(-) PLMS(+)**	**PD(+) RBD(+) PLMS(-)**	**PD(+) RBD(+) PLMS(+)**	**p - value**
Count	49	45	14	17	48	37	9	21	-
Gender (M/F)	21/28	36/9	9/5	12/5	27/21	26/11	5/4	15/6	**0.02**
Age (years, $\mu \pm \sigma$)	65.24 ± 7.00	67.96 ± 6.56	66.00 ± 7.39	64.65 ± 13.56	63.33 ± 9.02	66.86 ± 10.79	63.67 ± 9.31	68.86 ± 7.07	0.14
PLMS index (# PLMS/hsleep, $\mu \pm \sigma$)	4.37 ± 4.47	51.61 ± 28.82	6.23 ± 5.40	49.81 ± 34.97	3.32 ± 4.06	50.70 ± 34.13	7.54 ± 5.13	67.69 ± 45.03	**<0.001**
AHI (# apneas/hsleep, $\mu \pm \sigma$)	3.13 ± 5.98	2.3 ± 3.04	0.69 ± 1.84	2.06 ± 2.12	2.57 ± 4.71	5.18 ± 7.2	0.71 ± 0.7	2.01 ± 2.87	0.02
W (% TIB, $\mu \pm \sigma$)	23.46 ± 11.01	25.15 ± 10.66	21.66 ± 7.00	25.94 ± 12.06	23.18 ± 10.19	24.5 ± 11.41	23.44 ± 8.35	25.39 ± 9.47	0.91
REM (% TST, $\mu \pm \sigma$)	17.28 ± 5.82	17.46 ± 6.17	25.83 ± 5.63	21.65 ± 7.73	19.36 ± 6.63	19.23 ± 7.34	19.47 ± 5.75	20.00 ± 6.91	**<0.01**
N1 (% TST, $\mu \pm \sigma$)	25.97 ± 12.65	28.33 ± 9.86	20.74 ± 8.09	23.60 ± 7.10	22.22 ± 7.77	22.92 ± 11.41	25.41 ± 6.60	24.32 ± 8.38	**0.04**
N2 (% TST, $\mu \pm \sigma$)	47.56 ± 11.06	47.61 ± 11.04	44.49 ± 11.10	48.70 ± 11.39	48.80 ± 9.83	50.33 ± 9.62	44.83 ± 7.19	50.02 ± 10.1	0.65
N3 (% TST, $\mu \pm \sigma$)	9.19 ± 6.56	6.60 ± 7.23	8.94 ± 9.96	6.05 ± 5.96	9.62 ± 9.00	7.51 ± 7.27	10.29 ± 10.86	5.66 ± 7.01	0.29

Table 5.6: Demographics and sleep information of the DeNoPa cohort. Statistical analyses for group comparisons of continuous variables were performed with Kruskal-Wallis tests and for gender distribution with chi-squared test. $p < 0.05$ was considered significant and is shown in bold font. TIB: time in bed; TST: total sleep time; AHI: apnea/hypopnea index.

with 50% overlap according to Eqs. 5.1 and 5.2 from chin and tibialis muscles. For each muscle, such values were given in input to their three respective models of atonia (with optimal parameter D_{th}=0.01), and by using the thresholds defined for $p = 3$, three MA probability profiles were obtained ($p(MA)$). By applying the threshold $H = 90\%$, MA was identified from each MA probability profiles. From the identified MA, the MAFs $\%_{REM}$, $\%_{NREM}$, D_{REM} and D_{NREM} were finally derived. Such features were averaged across the three atonia models, thus obtaining four final MAFs from each muscle. In Section 5.3, a 5-fold CV technique was used, meaning that five optimal machine learning systems were available (each of them composed by the DS combination of one linear SVM, one SVM with RBF kernel and one RF). The 12 final MAFs were used as input for each of the five trained machine learning systems and each of them returned in output three probabilities (summing to 1):

- $p[RBD(-)PLMS(-)]$: the probability that the input participant showed neither RBD-like muscular activity nor increased PLMS index. This corresponds to the probability of being HC ($p(HC)$) in Section 5.3.1, as Danish HCs had neither RBD nor increased PLMS index;

- $p[RBD(+)]$: the probability that the input participant had RBD-like muscular activity. This corresponds to the probability of being an iRBD patient ($p(iRBD)$) in Section 5.3.1, without any further specification on PLMS index, as 12 out of the 29 Danish iRBD patients had increased PLMS index (Table 3.3).

- $p[RBD(-)][PLMS(+)]$: the probability that the input participant had increased PLMS index, but no RBD-like MA. This corresponds to the probability of being a PLMD patient ($p(PLMD)$) in Section 5.3.1, as Danish PLMD patients had increased PLMS index but not RBD.

This interpretation of the output probabilities did not include any consideration on PD, as it was assumed that the proposed algorithm could identify RBD and increased PLMS index without RBD independently from PD diagnosis. Lastly, the final output probability values $P[RBD(-)PLMS(-)]$, $P[RBD(+)]$ and $P[RBD(-)PLMS(+)]$ were calculated as the median across the five classifiers' outputs.

Statistical analysis of the features

From each participant, 12 final MAFs were extracted: $\%_{REM}$, $\%_{NREM}$, D_{REM} and D_{NREM} from chin, TIBL and TIBR EMG signals. For the following statistical analysis, MAFs values coming from the two tibialis muscles were averaged, to have a global overview of the limb activity. For each of the so-obtained eight MAFs, a multivariate linear regression analysis was performed, where the feature was the variable and the following were the factors: the presence of PD (categorical factor), the presence of RBD (categorical factor) and increased PLMS index (categorical factor, set to positive when the PLMS index was above 15). To fulfill the normality assumption, a *log*-transformation was applied to all the features, after adding 0.1 to the percentages of MA and 1 to the median distances to avoid infinite variable values. A correction for age and gender was performed, and no interaction effect was considered. Multiple comparison correction was performed with Bonferroni-Holm procedure [104] and a visual check of the residuals ensured that the normality assumption was met.

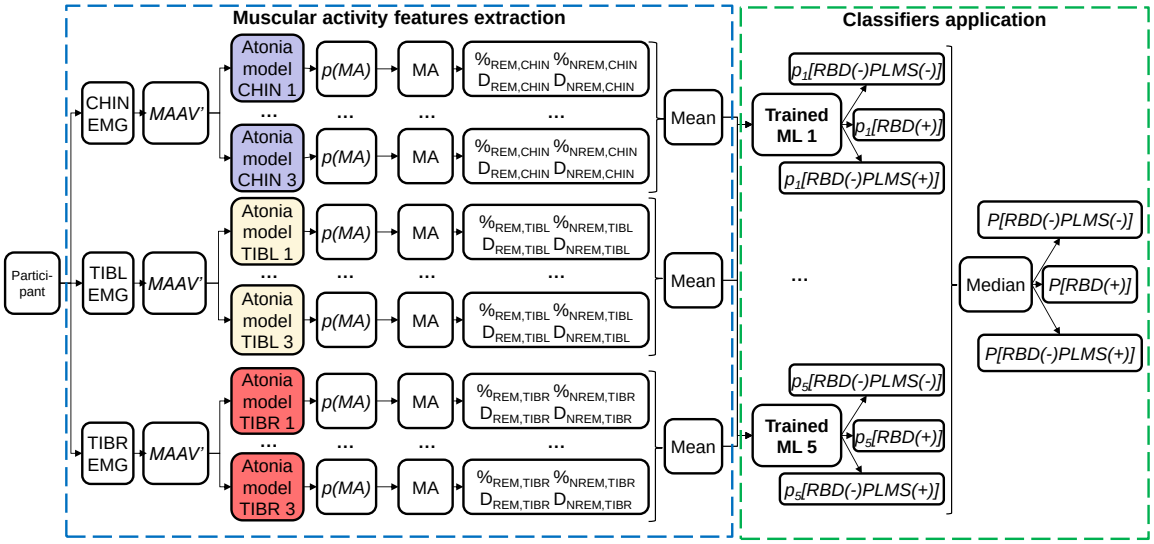


Figure 5.13: Schematic visualization of how the proposed data-driven method was applied to each participant of the German cohort. From each muscle, $MAAV'$ values were calculated and given in input to three atonia models which returned profiles of probability of MA ($p(MA)$). From these, MA was defined and MAFs derived. MAFs were then averaged across the three atonia model for each muscle, thus obtaining the final set of 12 MAFs. The final features were given as input to the five trained machine learning (ML) systems of the 5-fold CV implemented in Section 5.3.1. Their output probabilities were averaged to obtain the final probabilities, which were used for final classification.

Comparison of the classification performances in the Danish and German cohorts

The aim of this study was to analyze the capability of the proposed data-driven method in distinguishing German groups and to compare the performances achieved in the German cohort to the ones in the Danish cohort (Section 5.3). For this purpose, the following German groups were selected: (a) RBD(-)PLMS(-), analogous to the Danish HCs because of the absence of RBD diagnosis and normal PLMS index; (b) RBD(+), including both RBD(+)PLMS(-) and RBD(+)PLMS(+), similar to the Danish iRBD patient group; and (c) RBD(-)PLMS(+), comparable to the Danish PLMD group for the increased PLMS index. In the three groups both participants with and without PD were included, as it was assumed that RBD diagnosis and identification of increased PLMS index were independent from PD diagnosis. Each participant was automatically classified as either RBD(-)PLMS(-) or RBD(+) or RBD(-)PLMS(+) based on the highest value between $P([RBD(-)PLMS(-)])$, $P([RBD(+)])$ and $P([RBD(-)PLMS(+)])$ (Fig. 5.13). The classification performances were evaluated by means of overall accuracy (Eq. 5.14) and class-specific accuracy, sensitivity and specificity (Eq. 5.15).

To understand whether inter-clinical variability affected the classification performance, the same classification was performed after normalizing each MAF φ of each participant (with $\varphi \in \{\%_{REM,CHIN}, \%_{NREM,CHIN}, D_{REM,CHIN}, D_{NREM,CHIN}, \%_{REM,TIBL}, ..., D_{NREM,TIBL}, \%_{REM,TIBR}, ..., D_{NREM,TIBR}\}$) by multiplying it by the normalizing factor N_φ defined as:

$$N_\varphi = \frac{\bar{\varphi}_{HC}}{\bar{\varphi}_{PD(-)RBD(-)PLMS(-)}} \tag{5.16}$$

where $\bar{\varphi}_{HC}$ is the average value of the MAF φ across the Danish HCs, and $\bar{\varphi}_{PD(-)RBD(-)PLMS(-)}$ the average value of φ across the German group PD(-)RBD(-)PLMS(-). The normalizing factors

were designed to make the German group PD(-)RBD(-)PLMS(-) have similar MAF values to the Danish HCs, to overcome possible inter-clinical differences in the way in which these analogous control groups were identified in the two centres.

To compare the classification performances obtained in the German cohort of RBD(-)PLMS(-), RBD(+) and RBD(-)PLMS(+) to the ones obtained in the Danish cohort of HCs, iRBD and PLMD, it was evaluated whether the former were in the range of one standard deviation from the average performances obtained in the 5-fold CV achieved for the Danish cohort (Table 5.4, first column).

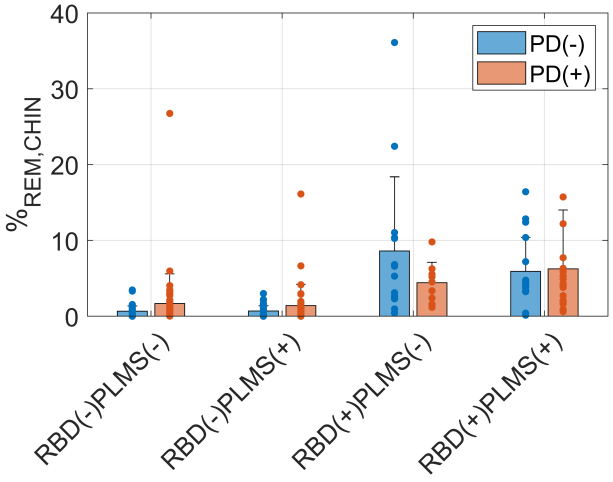


Figure 5.14: Distribution of the percentage of 1-s windows with detected MA in the chin muscle during REM sleep ($\%_{REM,CHIN}$) across the participant groups. Bar heights represent average values in the groups, whiskers represent one standard deviation, and dots represent singular participant values.

5.4.2 Results

Statistical analysis of the extracted features

Fig. 5.14 shows the distribution of the feature $\%_{REM,CHIN}$ as an example of MAF distribution across the groups (the remaining ones are shown in the Supplemental Material in Appendix D). Table 5.10 reports the values of all eight MAFs across the groups and Table 5.7 shows the results of the multiple linear regression analysis. RBD was related to a significant increase in the percentage of MA in chin and tibialis muscles, not only during REM sleep but also during NREM sleep, thus confirming the finding of abnormal MA in NREM sleep. Moreover, a significant effect of RBD was also seen in $D_{REM,CHIN}$, meaning that MA in chin during REM sleep is not just increased in percentage coverage, but is also more frequent. An increased PLMS index was found to be related to significantly increased and more frequent muscular activity during NREM sleep and increased tone in REM sleep in the tibialis muscles. The presence of PD did not have any significant influence on the analyzed MA features.

Comparison of the classification performances in Danish and German cohorts

The classification performances obtained in the Danish cohorts of HCs, iRBD and PLMD groups and the ones obtained in the German cohort of RBD(-)PLMS(-), RBD(+) and RBD(-)PLMS(+) are shown in Figure 5.15. For the German participants, the results are shown for normalized (Fig.

MA feature	PD		RBD		PLMS	
	t-stat	p-value	t-stat	p-value	t-stat	p-value
$\%_{REM,CHIN}$	0.638	0.524	13.344	**<0.001**	-0.525	0.600
$\%_{NREM,CHIN}$	-1.902	0.058	3.451	**<0.001**	0.449	0.654
$D_{REM,CHIN}$	0.424	0.672	-6.392	**<0.001**	0.073	0.942
$D_{NREM,CHIN}$	1.837	0.068	-0.011	0.992	-0.368	0.714
$\%_{REM,TIB}$	-0.477	0.634	8.150	**<0.001**	4.621	**<0.001**
$\%_{NREM,TIB}$	-1.894	0.060	4.56	**<0.001**	13.83	**<0.001**
$D_{REM,TIB}$	-0.113	0.910	-1.675	0.096	-2.459	0.015
$D_{NREM,TIB}$	-1.871	0.063	-0.960	0.339	-9.607	**<0.001**

Table 5.7: Results of the multivariate regression statistical analyses. For each analysis, the Student's t-test statistic (t-stat) and corresponding p-value are shown for each factor. p-values that remained significant after Bonferroni-Holm correction are in bold font.

		Predicted		
		RBD(-)PLMS(-)	**RBD(+)**	**RBD(-)PLMS(-)**
Actual	**RBD(-)PLMS(-)**	87 (46/41)	6 (1/5)	4 (2/2)
	RBD(+)	5 (3/2)	36 (16/20)	20 (12/8)
	RBD(-)PLMS(+)	46 (25/21)	7 (2/5)	29 (18/11)

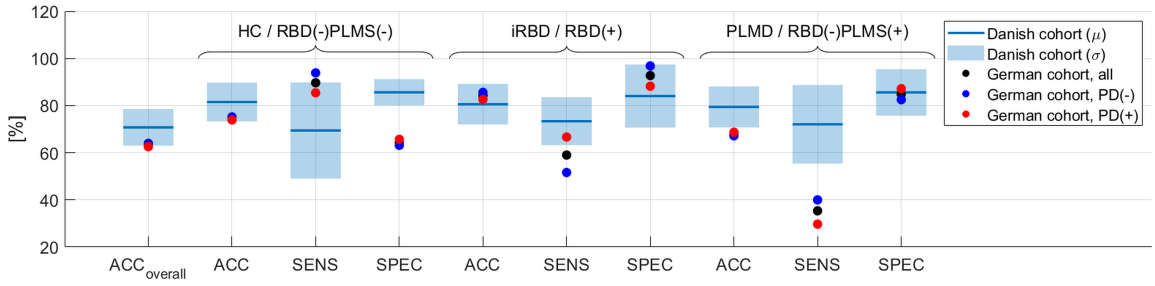
Table 5.8: Confusion matrix for classifying RBD(-)PLMS(-), RBD(+) and RBD(-)PLMS(+) participants when not normalized MA features were used. Results are shown for all participants and for PD(-)/PD(+) participants in parentheses.

		Predicted		
		RBD(-)PLMS(-)	**RBD(+)**	**RBD(-)PLMS(-)**
Actual	**RBD(-)PLMS(-)**	71 (39/32)	16 (4/12)	10 (6/4)
	RBD(+)	2 (0/2)	43 (21/22)	16 (10/6))
	RBD(-)PLMS(+)	15 (6/9)	8 (2/6)	59 (37/22)

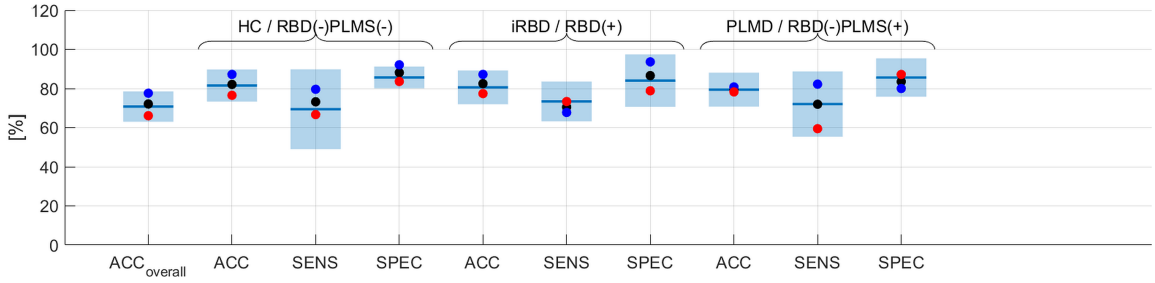
Table 5.9: Confusion matrix for classifying RBD(-)PLMS(-), RBD(+) and RBD(-)PLMS(+) participants when normalized MA features were used. Results are shown for all participants and for PD(-)/PD(+) participants in parentheses.

5.15b) and not normalized (Fig. 5.15a) MAFs, and including all participants (black dots) and considering only the ones with and without PD (red and blue dots, respectively). The numeric values of the classification performances are shown in Table 5.11. Tables 5.8 and 5.9 show the confusion matrices in the cases of using not normalized and normalized MAFs, respectively. Table 5.12 reports the normalizing factors N_φ.

The overall accuracies in the German cohort were always in the range of the ones obtained for the Danish cohort, with the only exception being the one achieved with not-normalized MAFs for PD(+) participants. When not-normalized features were used, the algorithm performed in the same range as in the Danish cohort for RBD(-)PLMS(-) accuracy, RBD(+) accuracy and specificity, and RBD(-)PLMS(+) specificity. Most of the errors were caused by RBD(-)PLMS(+) participants, which were misclassified as RBD(-)PLMS(-). When normalized features were used, the classification performances were always in the ranges obtained for the Danish cohort or even higher. Finally, the algorithm generally performed worse for PD(+) compared to PD(-) participants.



(a) Not normalized MA features.



(b) Normalized MA features.

Figure 5.15: Comparison between the classification performances obtained for the Danish cohort (HCs, iRBD and PLMD patients) and the ones achieved in the German cohort (RBD(-)PLMS(-), RBD(+) and RBD(-)PLMS(+)) in the case of using (a) not normalized and (b) normalized MAFs.

5.4.3 Discussion

The proposed data-driven method for RBD detection, originally developed on a Danish cohort, can successfully identify patients with RBD in PSG data recorded in another clinic. Three are the main outocomes of this study: i) when the algorithm was applied without normalizing MAFs, patients with RBD could be identified with the same performances as in the Danish database, but a tendency to misclassify patients with increased PLMS index without RBD was observed; ii) when the algorithm was applied with normalized features, the classification performances were similar to the ones achieved in the Danish database; and iii) the classification performances were lower for patients suffering from PD than for participants without PD.

Comparison with other automated methods

When considering other automated methods for RBD detection, only the automated SINBAR has been applied to a big dataset including PSGs recorded in several clinics. However, a formal validation of this method in different clinics is lacking [55]. Only RAI has been extensively validated on PSG data recorded in different clinics [95, 96, 97, 101, 137, 138]. When validated on similar cohorts to the one here investigated, RAI could distinguish iRBD patients from HCs with 82.6% accuracy, 84% sensitivity and 81% specificity [97], and RBD could be identified in PD patients with 85.5% accuracy, 96.6% sensitivity and 72.0% specificity [96]. The proposed method could identify RBD with accuracy comparable to RAI (Table 5.11 with not normalized MAFs for PD(-) and PD(+)), but with higher specificity and lower sensitivity, meaning that the proposed algorithm identified relatively fewer patients with RBD, but with a higher certainty than RAI.

MA feature	**RBD(-)PLMS(-)**		**RBD(-)PLMS(+)**		**RBD(+)PLMS(-)**		**RBD(+)PLMS(+)**	
	PD(-)	**PD(+)**	**PD(-)**	**PD(+)**	**PD(-)**	**PD(+)**	**PD(-)**	**PD(+)**
$\%_{REM,CHIN}$ [%]	0.68±0.69	1.70±3.91	0.70±0.69	1.41±2.82	8.62±9.79	4.44±2.69	5.92±4.48	6.26±7.75
$\%_{NREM,CHIN}$ [%]	1.37±1.21	1.84±2.48	1.60±1.84	1.87±2.72	2.69±2.61	3.03±4.30	4.45±3.78	1.70±1.97
$D_{REM,CHIN}$ [s]	124.34±165.35	100.03±91.38	125.88±139.02	180.70±251.17	41.23±41.69	37.83±26.60	32.27±37.30	38.83±37.91
$D_{NREM,CHIN}$ [s]	49.77±49.55	98.18±166.11	53.03±91.19	107.84±145.41	44.38±24.11	42.35±37.44	31.72±16.11	76.94±86.94
$\%_{REM,TIB}$ [%]	0.78±0.79	0.87±1.31	1.38±1.31	1.86±2.95	2.96±3.96	1.27±1.24	4.23±3.91	4.61±4.24
$\%_{NREM,TIB}$ [%]	0.54±0.66	0.52±0.67	2.49±1.84	2.44±2.05	2.27±2.63	0.45±0.31	5.00±3.73	3.48±3.55
$D_{REM,TIB}$ [s]	70.04±82.96	87.22±56.08	62.18±40.31	66.25±50.03	63.20±26.55	92.18±67.16	60.77±56.99	38.95±22.62
$D_{NREM,TIB}$ [s]	110.57±103.77	123.14±130.16	35.84±17.47	37.19±18.02	105.50±67.86	177.93±166.68	32.08±11.26	54.00±58.27

Table 5.10: Distribution of the 8 MAFs used for statistical analysis across the participants as mean and one standard deviation.

			HC/RBD(-)PLMS(-)			**iRBD/RBD(+)**			**PLMD/RBD(-)PLMS(+)**		
		$ACC_{overall}$ [%]	**ACC** [%]	**SENS** [%]	**SPEC** [%]	**ACC** [%]	**SENS** [%]	**SPEC** [%]	**ACC** [%]	**SENS** [%]	**SPEC** [%]
Danish cohort	μ	70.76	81.52	69.43	85.64	80.58	73.38	84.04	79.42	72.06	85.60
	$\mu - \sigma$	63.01	73.32	49.03	80.07	71.97	63.22	70.67	70.75	55.39	75.78
	$\mu + \sigma$	78.51	89.72	89.83	91.21	89.19	83.54	97.41	88.09	88.73	95.42
	All	**63.33**	**74.58**	**89.69**	64.33	**84.17**	59.02	**92.73**	67.92	35.37	**84.81**
German cohort (1)	PD(-)	**64.00**	**75.20**	93.88	63.16	**85.60**	51.61	**96.81**	67.20	40.00	**82.50**
	PD(+)	62.61	**73.91**	**85.42**	65.67	**82.61**	**66.67**	**88.24**	68.70	29.73	**87.18**
	All	**72.08**	**82.08**	**73.20**	**88.11**	**82.50**	**70.49**	**86.59**	**79.58**	**71.95**	**83.54**
German cohort (2)	PD(-)	**77.60**	**87.20**	**79.59**	92.11	**87.20**	**67.74**	**93.62**	**80.80**	**82.22**	**80.00**
	PD(+)	**66.09**	**76.52**	**66.67**	**83.58**	**77.39**	**73.33**	**78.82**	**78.26**	**59.46**	**87.18**

Table 5.11: Classification performances in the Danish and German cohorts, with (1) not normalized MAFs and (2) with normalized MAFs. The performances in the Danish cohort are shown as average value (μ) and in the range of one standard deviation ($[\mu - \sigma, \mu + \sigma]$) across the 5-fold CV scheme (Section 5.3). Performances achieved in the German cohort that lay in the range of the Danish ones are shown in bold font.

	$\%_{\text{REM,CHIN}}$	$\%_{\text{NREM,CHIN}}$	$D_{\text{REM,CHIN}}$	$D_{\text{NREM,CHIN}}$	$\%_{\text{REM,TIBL}}$	$\%_{\text{NREM,TIBL}}$	$D_{\text{REM,TIBL}}$	$D_{\text{NREM,TIBL}}$	$\%_{\text{REM,TIBR}}$	$\%_{\text{NREM,TIBR}}$	$D_{\text{REM,TIBR}}$	$D_{\text{NREM,TIBR}}$
N[-]	1.65	1.08	1.45	1.71	2.02	2.23	1.23	0.67	2.62	2.04	1.54	0.68

Table 5.12: Normalizing factors (N) calculated for all the MAFs.

Inter-clinical variability

While the proposed algorithm without MAF normalization achieved acceptable classification performances for RBD detection, it tended to misclassify RBD(-)PLMS(+) as RBD(-)PLMS(-) (Fig. 5.15a, Tables 5.8 and 5.11). However, when feature normalization was applied, the classification performances in the German data improved and became similar to the ones achieved in the Danish data (Fig. 5.15b, Tables 5.9 and 5.11).

This might be caused by the differences in how the control groups (i.e. the German PD(-)RBD(-)PLMS(-) group and the Danish HCs) were defined. The normalizing factors (Table 5.12) can help to understand such differences. The normalizing factors for $\%_{REM,TIBL}$, $\%_{REM,TIBR}$, $\%_{NREM,TIBL}$ and $\%_{NREM,TIBR}$ were higher than 2. This indicates that Danish HCs showed on average increased limb muscular tone in REM and NREM sleep compared with the German PD(-)RBD(-)PLMS(-) participants. The normalizing factors for $\%_{REM,CHIN}$, $\%_{NREM,CHIN}$,$D_{REM,CHIN}$, $D_{REM,TIBL}$, $D_{REM,TIBR}$ and $D_{NREM,CHIN}$ were all in the range 1.08-1.71, indicating similar values for the two control groups. The normalizing factors for $D_{NREM,TIBL}$ and $D_{NREM,TIBR}$ were lower than 1. This suggests that the Danish HC group had on average more frequent muscular activity during NREM in the tibialis muscles compared to the German PD(-)RBD(-)PLMS(-) group.

From this, it can be concluded that the participants manually categorized as controls in Denmark showed (on average) increased and more frequent muscular tone in the tibialis muscles compared to the German control group. This difference cannot be attributed to respiratory events, as AHI distributions were similar in the two groups (Tables 3.3 and 5.6 and no statistical difference between the distributions). This dissimilarity might, however, be caused by inter-clinical differences in the way in which limb movements were scored (under-scored in Denmark or over-scored in Germany) and/or in the way in which controls were defined based on the scored muscular events. In the future, the use of an objective automated method has the potential to overcome such differences and make scoring and diagnosis more uniform across clinics worldwide.

Influence of PD diagnosis on the classification performances

Despite the presence of PD not showing a significant influence on MAFs (Table 5.7), the classification performances in Fig. 5.15 and Table 5.11 were generally lower for PD(+) participants compared to PD(-) ones. This means that, in a multivariate approach, the presence of PD led to more uncertainty in discriminating groups. It is interesting to notice, however, that RBD could be identified with higher sensitivity in PD(+) participants compared to the PD(-) ones, but with lower specificity. This might have been caused by the high number of PD(+) participants with REM behavior events (RBEs, Section 2.3.4) [56]. Therefore, RBEs might play as confounder for the automated method, as MA related to RBEs could increase the probability of RBD. At the same time, the algorithm made more errors in distinguishing RBD(-)PLMS(-) from RBD(-)PLMS(+) in PD(+) compared to PD(-) participants. This could be caused by the variety of abnormal MA during sleep, such as excessive fragmentary myoclonus, which is frequently seen in PD patients [153].

Limitations

The main limitation of this study is that it has not been investigated how many controls are needed to make the normalization efficient. Nevertheless, the proposed method still achieved acceptable performances when identifying patients with RBD also without normalization.

5.5 Conclusive remarks

The results presented in Paper III [132] and Paper IV [133] answer the research questions and address the research objectives outlined at the beginning of this Chapter. Below, the research questions are reported again, together with the research outputs achieved in this Chapter:

- ***Chapter research question 5.1***: Can a new method based on a data-driven identification of MA overcome in performances previously proposed algorithms for RBD detection?
  Chapter research output 5.1: The results presented in Paper III [132] show that the proposed data-driven method overcomes previously developed automated methods in identifying patients with iRBD in a cohort including also HCs and PLMD patients.

- ***Chapter research question 5.2:*** Can RBD be identified better if MA in NREM sleep is included?
  Chapter research output 5.2: The inclusion of features describing the coverage and frequency of MA in NREM sleep led to increased performances for correct identification of patients with RBD in Paper III [132]. Moreover, a significant influence of MA in NREM sleep in connection to RBD has been shown in Paper IV [133]. This further confirms the hypothesis that RBD is not a disease related to alterations during REM sleep only.

- ***Chapter research question 5.3:*** Which is the influence of MA related to apneas and arousals on the performances of the new method?
  Chapter research output 5.3: The results presented in Paper III [132] further confirm the findings of Paper I [94]. More specifically, it has been found that the proposed method achieved higher performances when MA related to apneas and arousals was included. This means that MA traditionally considered as noise can actually improve the classification performances. It has to be further remarked that these results can be considered valid for participants showing up to moderate sleep apnea and that they should be further validated for patients with severe sleep apnea.

- ***Chapter research question 5.4:*** Is the new method generalized? Does it perform equally well when applied on data that were recorded in a different clinic from the one where it was developed? If differences exist, can this method be used to evaluate inter-clinical differences?
  Chapter research output 5.4: When the proposed method was applied to data recorded in another clinic, it could identify patients with RBD with similar classification performances as the ones achieved in the cohort where it was developed. This proves that it is generalized for RBD detection. However, the application of the method to a second cohort revealed also inter-clinical differences in the detection of limb movements and/or how healthy controls are defined. Such differences could be overcome with a normalization technique.

Chapters 3, 4 and this Chapter focused on currently available methods and a new data-driven automated method for identifying RBD based on EMG signals. All these methods are actually semi-automated, as they still require manual identification of REM sleep. The next Chapter will propose a fully automated method aiming to identify RBD and prodromal RBD. Because of the known absence of abnormality in EMG signals in prodromal RBD patients, the method will use only EEG and EOG signal.

Chapter 6

Detection and characterization of RBD and prodromal RBD in Parkinson's disease

Chapter main objectives: Recent findings have shown that, in PD patients, development of RBD is a continuous process and that definite RBD might be preceded by a prodromal phase. It is hypothesized that the same development might occur also in absence of PD. The identification and electrophysiological characterization of prodromal RBD is therefore of utmost importance to intercept the ongoing neurodegeneration in its early stages.

This Chapter aims to present the development and validation of a fully automated method that can identify RBD and prodromal RBD in PD patients.

This Chapter is built on Paper V [154] and Paper VI [155] and addresses thesis objective 3.

6.1 Research background

Recent studies have found that, in PD patients, RBD is not a disease with a clear and definite starting point, but it might be preceded by a prior stage known as prodromal RBD [14]. Prodromal RBD is identified when at least two REM behavioral events (RBEs, defined as vocalization or behaviors with purposeful components in REM sleep) occur in one night sleep, without presence of REM sleep without atonia (RSWA) (Section 2.3.4). Despite the evolution from prodromal to definite RBD has been shown in PD patients only, a similar pattern is also expected in subjects without overt PD. The identification of prodromal RBD has therefore the potential to intercept neurodegeneration in its very early stages (Fig. 2.12).

As prodromal RBD is not characterized by RSWA, an automated method for its detection should use different modalities from EMG. In Section 2.3.3 a series of sleep EEG and EOG alterations in iRBD patients have been described and they have been interpreted as biomarkers of early neurodegeneration [64]. From this, it can be hypothesized that prodromal RBD patients, despite not having RSWA, might present some of these sleep electrophysiological abnormalities. Currently, data on prodromal RBD are available only for PD patients. Therefore, this Chapter focuses on the development of a fully automated method, based only on EEG and EOG signals, for prodromal RBD identification in patients with PD. Such a method is developed with the perspective of applying it to prodromal RBD patients without PD, when data will be available.

To develop such a method, the starting point was a literature search on the known differences between PD patients with and without RBD. About 33-46% of PD patients have RBD [57, 56] and, besides the known presence of RSWA, PD patients with RBD are characterized by more

severe motor and clinical features when compared to PD patients without RBD and the same disease duration [156]. In PD patients with RBD, more pronounced morphological changes in brain imaging (including more severe nigrostriatal dopaminergic impairment at caudate level [157], reduced thalamic volume [158], extensive cortical abnormalities and reduced volume in the putamen [40]) have been also found when compared to PD patients without RBD. These findings support the idea that PD patients with RBD have a more severe neurodegeneration. Few studies have investigated other electrophysiological changes in PD patients with RBD compared to PD patients without RBD, besides RSWA. No differences in sleep spindle density [74] or sleep macro-structure have been found [88]. Only one study based on PSG recordings has found EEG slowing during wakefulness in PD patients with RBD when compared to PD patients without RBD [159].

From this overview, it is therefore clear that more investigations are needed to evaluate electrophysiological abnormalities related to RBD in PD patients and to understand whether such abnormalities also characterize prodromal RBD.

6.2 Research questions and objectives of this chapter

From this background overview, the following research questions were defined:

- ***Chapter research question 6.1:*** Which electrophysiological changes, besides RSWA, characterize RBD in PD? Can a fully automated method use them to correctly identify RBD in PD patients?
- ***Chapter research question 6.2:*** Do PD patients with prodromal RBD have the same electrophysiological changes as PD patients with RBD? Can PD patients with prodromal RBD be identified with the same automated methods developed for identification of PD patients with RBD?

These research questions led to the definition of the following objectives for this Chapter:

- ***Chapter research objective 6.1:*** To develop a fully automated system that, based on sleep EEG and EOG signals, performs i) sleep staging, ii) feature extraction, and iii) classification to identify RBD in PD patients. Such a system should highlight the electrophysiological abnormalities related to RBD making the classification possible.
- ***Chapter research objective 6.2:*** To apply the developed system to evaluate whether it can identify PD patients with prodromal RBD. This will make it clear whether PD patients with prodromal RBD have similar electrophysiological abnormalities as PD patients with definite RBD.

6.3 Paper V: A clinically applicable interactive micro and macro-sleep staging algorithm for elderly and patients with neurodegeneration

With the aim of developing a fully automated system, the first objective was the development of a sleep staging algorithm working efficiently in patients with neurodegeneration. In literature, several sleep staging algorithms have been proposed [160], which use either single or multi-channel recordings in combination with traditional machine learning or new deep learning techniques [161, 162, 163, 164, 165]. However, most of them have been trained and validated only on small databases including only young healthy controls [162, 163, 164], or have not been tested in patients

with neurodegeneration [165]. Sleep staging algorithms developed and tested only on young healthy controls encounter problems when applied to elderly due to the electrophysiological changes that are connected with the aging process (Section 2.2.3). Sleep staging is even more challenging in patients with neurodegeneration due to the profound electrophysiological changes affecting them (Section 2.3.3). These scoring difficulties are reflected in a low inter-rater reliability, which has a median Cohen's kappa of 0.8 in HC [166], and only 0.6 in PD patients [167]. Few algorithms have tried to face sleep staging in patients with neurodegenerative diseases [149, 168, 169]. However, these algorithms either showed moderate performances [149, 169] or were tested on very small cohorts [168]. Therefore, there is the need of developing a new and efficient automated sleep staging algorithm for elderly and patients with neurodegeneration.

In the process of developing a new sleep staging algorithm, few preliminary considerations were made. First, AASM fixed the length of sleep epochs to 30 s [13]. This approach oversimplifies the sleep structure and such an epoch length is not explained by any physiological process [170]. To capture with more precision the physiological transitions between stages, a new sleep staging algorithm should be able to perform sleep staging not just in 30-s epochs (macro-sleep staging), but also at a smaller scale (micro-sleep staging). Secondly, almost all previously developed sleep staging algorithms have been developed as *black boxes*, from which the clinicians do not receive any feedback. The new algorithm should therefore be *interactive*, by indicating areas where the sleep stage predictions have been made with high and low confidence. In this way, clinicians could visually inspect the ones predicted with low confidence by the automated algorithm. Finally, it has been recently shown that scoring of the three NREM sleep stages is highly unreliable in clinical practice [171]. Therefore, the new sleep staging algorithm should perform only staging of wakefulness (W), REM and NREM sleep.

The methodology of the proposed sleep staging algorithm is shown in Fig. 6.1. Briefly, one EEG and two EOG channels are given as input to the system. After pre-processing, continuous wavelet transform is applied to the signals, and each signal is divided into 5-s mini-epochs. A convolutional neural network returns the sleep stage predictions for both 5-s mini-epochs, from which the sleep stage predictions in 30-s epochs are derived. The predictions for the 30-s epoch are further labeled as *certain* or *uncertain*, depending on the level of confidence of the prediction itself.

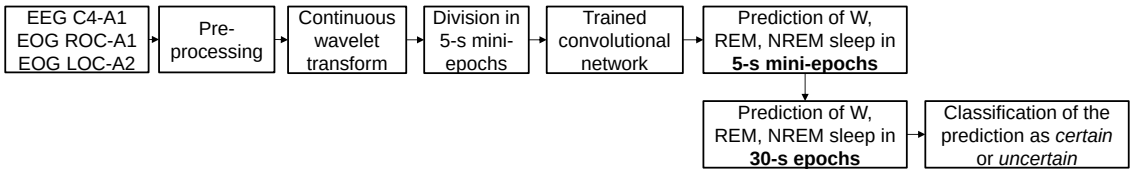


Figure 6.1: Overview of the methodology of Paper V [154]. One EEG and two EOG channels were the input of the algorithm. After pre-processing, the signals were divided into 5-s mini-epochs and given in input to a convolutional neural network, which returned the sleep stage predictions for 5-s mini-epochs (micro-sleep staging). From such predictions, the predictions in 30-s epochs were obtained (macro-sleep staging). Moreover, each epoch was labeled as either *certain* or *uncertain* based on the level of confidence of the macro-sleep prediction.

6.3.1 Methods: Development and validation of a convolutional neural network for macro- and micro-sleep staging

Subjects and recordings

The cohort included for the development of the new sleep staging algorithm consisted in the "DeNoPa" cohort at the baseline evaluation (Section 5.4.1). In this context, the groups were defined according to their neurological and RBD diagnosis, without taking into account the PLMS index. The groups therefore included 94 healthy controls (HCs), 85 PD patients without RBD (PDnonRBD), 30 PD patients with RBD (PDnonRBD) and 31 iRBD patients. In the development of the sleep staging algorithm, it was decided to exclude 13 HCs and 28 PDnonRBD patients, as they had RBEs, but not RSWA. Sleep experts manually scored each 30-s sleep epoch of the PSGs according to international criteria [151].

Sleep staging algorithm

As in the baseline study of the "DeNoPa cohort" only central EEG derivations were available, the C4-A1 EEG and LOC-A2 and ROC-A1 EOG channels were considered for developing the sleep staging algorithm, which is described in details below.

Pre-processing Saturation and pop-ups artifacts were identified in the three signals with the same methodology described in Section 5.4.1. Thirty-second sleep epochs containing such artifacts were removed from the analysis.

An adaptive filter was used to remove ECG and EOG interference from the EEG signal, and to remove ECG interference from the EOG signals [172]. Fig. 6.2 presents a schematic overview of the adaptive filter applied to C4-A1 signal. The input is the EEG signal $\alpha(n)$ which is modelled as a mixture of the *true* EEG signal $\beta(n)$ and a noise component $\gamma(n)$ (i.e. $\alpha(n) = \beta(n) + \gamma(n)$). The signals $r_{EOG}(n)$ and $r_{ECG}(n)$ are the two reference inputs, corresponding to the ROC-A1 and ECG signal, and it is assumed that they are correlated (in some unknown way) to the noise $\gamma(n)$. Finally, $\theta_{EOG}(m)$ and $\theta_{ECG}(m)$ are the finite impulse response (FIR) filters of length M (which can be different in the two filters). The desired output $e(n)$ is obtained as:

$$e(n) = \alpha(n) - \hat{r}_{EOG}(n) - \hat{r}_{ECG}(n) = \beta(n) + [\gamma(n) - \hat{r}_{EOG}(n) - \hat{r}_{ECG}(n)] \tag{6.1}$$

where

$$\hat{r}_i(n) = \sum_{m=1}^{M} \theta_i(m) r_i(n+1-m) \tag{6.2}$$

is the filtered referenced signal, with $i \in \{EOG, ECG\}$. By assuming that $\beta(n)$ is a zero-mean stationary random signal uncorrelated with $\gamma(n)$, r_{EOG} and r_{ECG}, the expected value of $e^2(n)$ is calculated as:

$$\begin{gathered} E[e^2(n)] = E[(\beta(n) + \gamma(n) - \hat{r}_{EOG}(n) - \hat{r}_{ECG}(n))^2] = \\ E[\beta^2(n)] + E[(\gamma(n) - \hat{r}_{EOG}(n) - \hat{r}_{ECG}(n))^2] \end{gathered} \tag{6.3}$$

The goal is to make the signal $e(n)$ as close as possible to the true signal $\beta(n)$, by adjusting the filter coefficients. Since $E[\beta(n)^2]$ is not affected by these coefficients, the minimization of $E[(\gamma(n) - \hat{r}_{EOG}(n) - \hat{r}_{ECG}(n))^2]$ is equivalent to the minimization of $E[e^2(n)]$. This minimization problem can be solved with a recursive-least-square algorithm (further details can be found in [172]). The algorithm works also with one reference signal.

The adaptive filtering was applied to C4-A1, LOC-A2 and ROC-A1 signals with the parameters $M = 3$ and forgetting factor set to 0.995 as suggested in [172].

After the adaptive filtering, the EEG and EOG signals were filtered with a 4^{th} order zero-phase band-pass Butterworth filter with cutoffs at 0.5 and 35 Hz. For computation reasons, each signal was downsampled to 100 Hz with an anti-aliasing FIR low-pass filter.

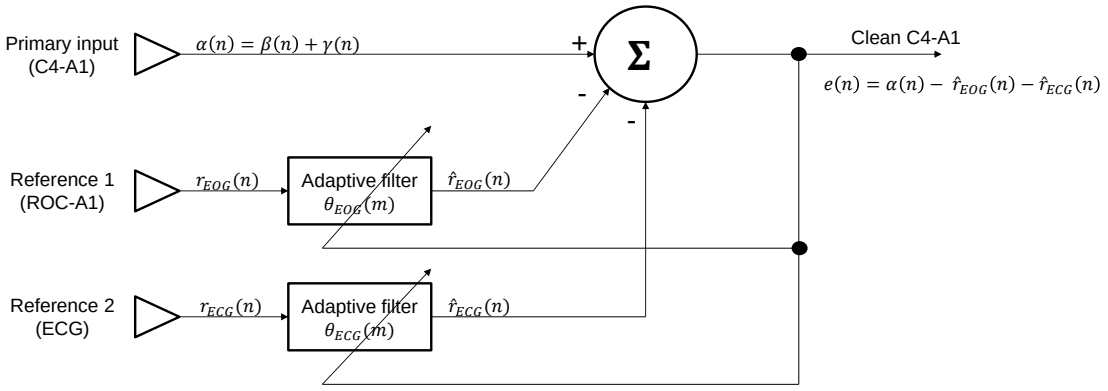


Figure 6.2: Schematic overview of the adaptive filtering applied to the C4-A1 signal to remove the interference from ROC-A1 EOG and ECG signals. Adapted from [172].

Scalogram computation Continuous wavelet transform (CWT, Eq. 4.1) was used to transform the signals into images. Compared to the discrete wavelet transform introduced in Section 4.3.1, CWT is not implemented by discretizing the scale and translation on a dyadic scale, but by allowing them to change for any integer value [173]. By varying scale and translation parameters, the output of the CWT is a 2D time-frequency representation of the signal analysed [174].

The Morse mother wavelet with time-bandwidth product $P = 60$, the symmetry parameter $\gamma = 3$ and 6 voices per octave [175] was used for calculating the CWT of the EEG and the two EOG signals. The absolute value was then calculated, thus obtaining a final *scalogram* (i.e. an image).

Due to the parameters used for implementing the CWT, the frequency band in the range 0-50 Hz was divided into 111 bins (with higher resolution at lower frequencies and lower resolution at higher frequencies). Therefore, from each signal of length L seconds, the final scalogram had dimensions of $111 \times (100 \cdot L$ pixels). Because of the band-pass filtration, only the frequencies 0.5-40 Hz were included and the scalogram was resampled in time to 4 Hz, thus reducing its size to 39 $\times$ $(4 \cdot L)$ pixels. Finally, each scalogram was divided into 5-s mini-epochs with 50% overlap, thus obtaining final images of size 39 $\times$ 20 pixels, which were normalized in the range 0-1 (Fig. 6.3).

Classification algorithm To perform sleep staging, a simple *convolutional neural network* (CNN) was built. An explanation of artificial neural networks, CNNs and the structure of the implemented CNN is provided here below.

Artificial neural networks In Fig. 6.4 the structure of a basic artificial neural network is shown. Given the input $x_1, ..., x_{\hat{D}}$, M linear combinations of the input can be obtained as:

$$\hat{a}_j = \sum_{i=1}^{\hat{D}} w_{ji}^{(1)} x_i + w_{j0}^{(1)} \tag{6.4}$$

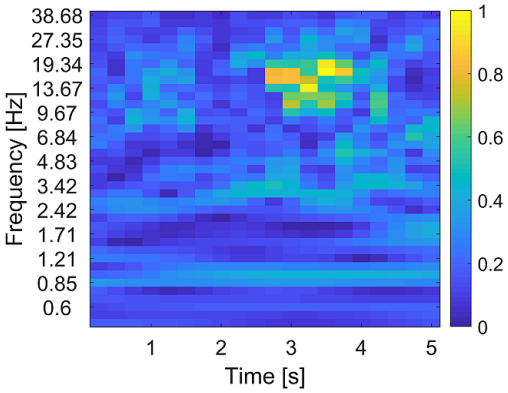


Figure 6.3: Final image obtained from the scalogram computed from C4-A1 signal for a 5-s mini-epoch. The normalized magnitude is shown by color. The area with pixels in yellow indicates that, between 3 and 4 s, the EEG signal has strong components in the range 10-25 Hz.

where $j = 1, ..., M$ and the superscript (1) indicates the parameters in the first layer of the network. Usually, the parameters $w_{ji}^{(1)}$ are referred as weights and the parameters $w_{j0}^{(1)}$ as biases. The values $\hat{a}_j$ are known as activations and each of them is the input of a nonlinear activation function $\hat{h}$. The ouput is:

$$\hat{z}_j = \hat{h}(\hat{a}_j) \tag{6.5}$$

These quantities are known as hidden units. Such values are again combined, thus obtaining:

$$\hat{a}_k = \sum_{j=1}^{M} w_{kj}^{(2)} \hat{z}_j + w_{k0}^{(2)} \tag{6.6}$$

where $k = 1, ..., \hat{K}$ and $\hat{K}$ is the total number of outputs. These activations are then given as input of an appropriate activation function to get the final output: $y_k = \hat{h}(\hat{a}_k)$. Usually, the final activation function is the sigmoid function for binary class problems (Eq. 3.1) and the softmax function multiclass problems for $t = 1, ..., C$ (where C is the total number of classes):

$$softmax(\mathbf{x})_t = \frac{e^{x_t}}{\sum_{k=1}^{C} e^{x_k}} \tag{6.7}$$

During training, it is sought to minimize the cross-entropy error function $E(\mathbf{w})$ (Eq. 3.4, where $\mathbf{w}$ is the vector containing weights and biases) and such a minimization is obtained by finding a good combination of weights and biases. These parameters can be optimized iteratively by applying different methods. A simple one is the stochastic gradient descent optimization, defined as:

$$\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} - \hat{\eta} \nabla E(\mathbf{w}^{(\tau)}) \tag{6.8}$$

where $\hat{\eta}$ is the learning rate and $\nabla E(\mathbf{w}^{(\tau)})$ is the gradient of the error function evaluated at the iteration τ. In neural networks, the gradient of the error function is estimated with the *backpropagation* algorithm. This algorithm consists simply in first propagating the input through the network, calculating the error and propagating it backwards to update the weights (additional details can be found in [102]). The weights are optimized with Eq. 6.8 until convergence or until a stopping criterion is met.

From both theoretical and practical points of view, artificial neural networks are particularly interesting as the *universal approximator theorem* applies to them. This means that an artificial neural network, with appropriate activation functions, can approximate *any* function [176].

A deep neural network is defined as an artificial neural network with multiple layers between the input and the output layers.

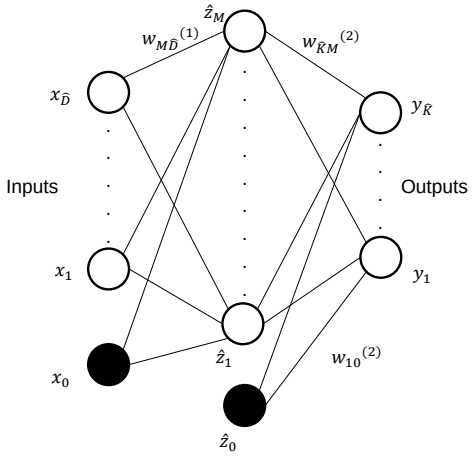


Figure 6.4: Schematic representation of a two-layer neural network, where the input, hidden, and output variables are represented by nodes, and the weights as links between the nodes. Adapted from [102].

Convolutional neural networks CNNs are simply neural networks where the matrix multiplications (e.g. in Eq. 6.4) are substituted with convolutions in at least one layer. Fig. 6.5 shows the convolution of an image of dimensions 3×4 with a 2-dimensional kernel of dimensions 2×2. When using a convolutional layer, the dimensions of the kernel can be chosen, as well as the number of the kernels, the stride (i.e. the step size for traversing the input vertically and horizontally) and whether the input should be padded and how.

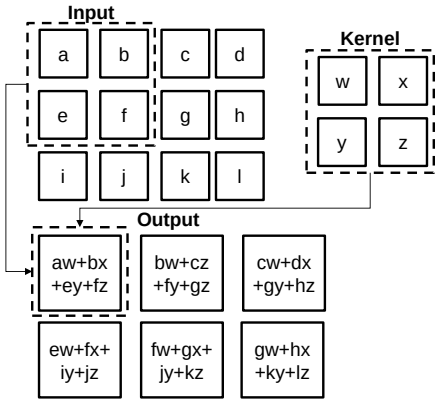


Figure 6.5: Example of 2-D convolution between an image and a kernel. No padding and stride (1,1) were used. Adapted from [177].

Structure of the implemented CNN The architecture of the CNN implemented to achieve sleep staging prediction is shown in Fig. 6.6. The input 3D matrix consisted of the stacked 5-s mini-epoch scalograms of C4-A1, ROC-A1 and LOC-A2. The input dimension was therefore $39 \times 20 \times 3$.

The input was connected with the first convolutional layer having 24 filters (i.e. kernels), each of them with dimensions 39×1. Because of the filters dimensions, this convolutional layer could be interpreted as a filter-bank performing filtering in the frequency domain. The obtained output

had dimensions $1 \times 20 \times 24$ and was followed by a rectified linear unit (ReLU) activation function, with expression $ReLU(x) = \max(0, x)$. A second convolutional layer (96 filters of dimensions 1×4) followed, and its output was a matrix with dimensions $1 \times 17 \times 96$. Again, a ReLU activation function was used after it. Because of the filters dimensions, the second convolutional layer could be interpreted as a time-domain filter bank. The outputs after the second activation function were the input to a fully connected layer of dimensions $1 \times 1 \times 64$. This means that all the values in output from the second activation function were connected to all 64 nodes of this layer. This was followed again by a ReLU activation function and a second fully connected layer with dimensions $1 \times 1 \times 3$. Finally, the output of the second fully connected layer were given as input to a softmax activation function (Eq. 6.7). The final output consisted of three probability values (summing to 1), corresponding to the probabilities that the input mini-epoch was wakefulness ($p_{ME}(W)$), REM sleep ($p_{ME}(REM)$) and NREM sleep ($p_{ME}(NREM)$).

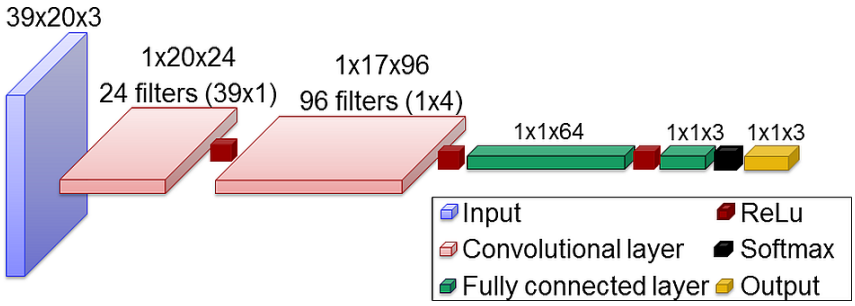


Figure 6.6: Schematic view of the implemented CNN. The type and size of each layer and the activation functions are specified.

Training and validation The CNN was trained and validated on 30 and 10 randomly selected HCs, respectively. To let the algorithm learn a clean and true representation of W, REM and NREM sleep, the training data included only 5-s mini-epochs contained in 30-s epochs not located at sleep stage transitions. Because of absence of manual scoring in 5-s mini-epochs, the sleep stage assigned to each 5-s mini-epoch was the manual sleep stage of the 30-s epoch where the mini-epoch was located. For each of the 30 HCs used for training, the amount of mini-epochs of the three sleep stages were balanced by randomly under-sampling the two most represented classes to the number of mini-epochs of the least represented one.

When training CNNs (or in general deep learning architectures), not all training data is used at once. The training data is instead divided into *mini-batches*, which consist of portions of the training data. In each iteration of network optimization, a different mini-batch is used to make the network robust and generalized. Once all the mini-batches have been used, a training *epoch* is completed. To increase the network performances, usually several training epochs are repeated.

The final structure of the CNN previously described was obtained by trial and error based on the evaluation and comparison of the cross-entropy error in the training and validation datasets, to obtain high performances and avoid overfitting. The parameters optimized with trial and error procedure were the number of layers, the dimensions of the filters, the learning rate, the number of training epochs and the mini-batch size. The final network was trained with mini-batches of 128 samples and with 30 training epochs. Moreover, instead of the traditional stochastic gradient descent optimization (Eq. 6.8), the Adam optimizer [178] was used. Compared to stochastic gradient, the main difference is that Adam optimizer computes individual adaptive learning rates

for different parameters from estimates of first and second moments of the gradients [178]. Moreover, it has been shown to be faster and generally superior to stochastic gradient descent [178].

Evaluation of the classification performances After training, the algorithm was applied to the remaining test set, consisting of 41 HCs, 31 iRBD, 57 PDnonRBD and 30 PD+RBD patients. For each 5-s mini-epoch, the trained algorithm returned the mixture of probabilities $p_{ME}(W)$, $p_{ME}(REM)$ and $p_{ME}(NREM)$ (Fig. 6.7a). As only manual annotations for 30-s epochs were available, it was decided to evaluate the classification performances only at macro-scale level. The following steps were applied to obtain the macro-scale sleep staging predictions:

- *Smoothing*: The probability values obtained in the mini-epochs were first smoothed with a Blackman window of length 38. An example of smoothed probabilities in mini-epochs (i.e. $p_{ME,S}(W)$, $p_{ME,S}(REM)$ and $p_{ME,S}(NREM)$) is shown in Fig. 6.7b.
- *Averaging*: For each 30-s epoch, the probabilities of W, REM and NREM sleep (i.e. $p_{E,S}(W)$, $p_{E,S}(REM)$ and $p_{E,S}(NREM)$, in Fig. 6.7c) were obtained by averaging the values $p_{ME,S}(W)$, $p_{ME,S}(REM)$ and $p_{ME,S}(NREM)$ across the mini-epochs included in that specific 30-s sleep epoch;
- *Prediction*: For each 30-s sleep epoch, the predicted sleep stage was the one with the highest probability between $p_{E,S}(W)$, $p_{E,S}(REM)$ and $p_{E,S}(NREM)$.

Sleep stage prediction in 30-s sleep epochs were obtained for the training, validation and test datasets. By comparing the predicted sleep stages with the manual annotations, it was possible to calculate the overall accuracy (Eq. 5.14) and the overall Cohen's kappa defined as [179]:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{6.9}$$

where p_o is the overall accuracy and p_e is the hypothetical probability of chance of agreement. In particular, p_e is calculated as [179]:

$$p_e = \frac{1}{N^2} \sum_{k=1}^{C} n_{c_k,manual} n_{c_k,automatic} \tag{6.10}$$

where C is the number of classes, N the total number of observations and $n_{c_k i}$ the number of times that the "scorer" (i.e. manual or automatic) predicted the class c_k [179]. Moreover, the stage-specific accuracy, sensitivity and specificity (Eq. 5.15) were calculated for training, validation and test datasets.

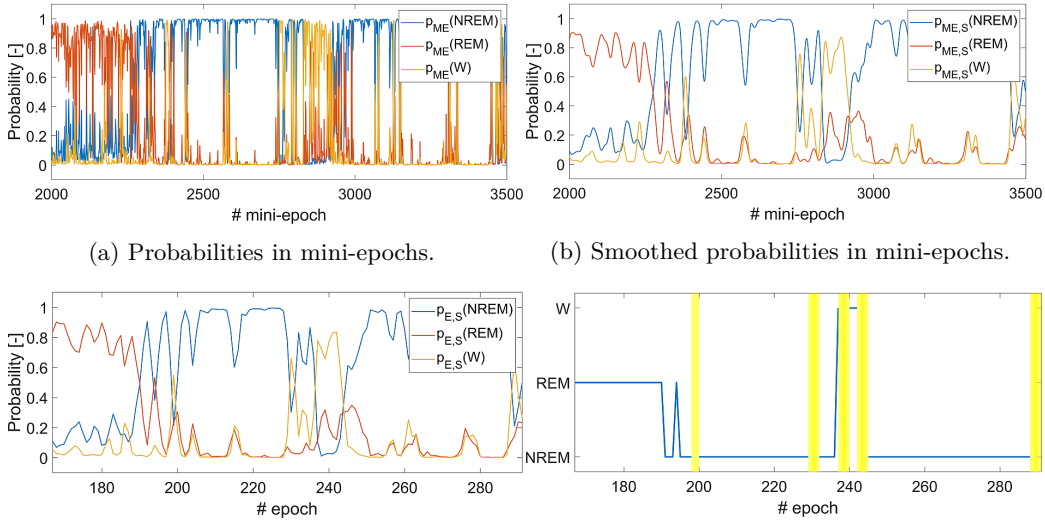
The length of the Blackman window for the smoothing procedure was chosen as the one in the range $[1, 2, ..., 150]$ achieving the highest training overall accuracy.

Identification of *uncertain* epochs

To make the proposed sleep staging *interactive*, each macro-sleep prediction was labeled as either *certain* or *uncertain*. The labeling was done to propose a framework where sleep technicians could manually check the epochs labeled as *uncertain* only and keep the automatic predictions for the *certain* ones.

The labeling was made on the test dataset only and the following steps were performed. The 30-s sleep epochs where manual and automatic scoring agreed in scoring wakefulness were labeled as $\bar{A}_W$, and the ones manually scored as wakefulness but automatically scored otherwise were labeled as

$\bar{D}_W$. For these epochs, their $p_{E,S}(W)$ values were considered. In a 10-fold cross-validation, receiver operating characteristic analysis was used to identify the threshold th_W (varying in the range $[0.01, 0.02, ..., 1.00]$) for $p_{E,S}(W)$ that was maximizing the area under the curve in distinguishing the epochs $\bar{A}_W$ from $\bar{D}_W$ in the training fold. In the validation fold, the epochs classified as W with $p_{E_S}(W) < th_W$ were labeled as *uncertain* and as *certain* otherwise. The same was repeated also for REM and NREM sleep.



(a) Probabilities in mini-epochs.

(b) Smoothed probabilities in mini-epochs.

(c) Probabilities in epochs.

(d) Predictions and *uncertain* epochs.

Figure 6.7: Output probabilities of the CNN and their post-processing. (a) Output probabilities of the trained CNN for a sleep segment of a test healthy control. For each mini-epoch the probabilities of W, REM and NREM sleep ($p_{ME}(W)$, $p_{ME}(REM)$ and $p_{ME}(NREM)$) were obtained; (b) smoothed mini-epoch probabilities with a Blackman window of length 38 ($p_{ME,S}(W)$, $p_{ME,S}(REM)$ and $p_{ME,S}(NREM)$); (c) probabilities estimated for each epoch by averaging the smoothed probabilities across the mini-epochs included in the respective 30-s epochs ($p_{E,S}(W)$, $p_{E,S}(REM)$ and $p_{E,S}(NREM)$); (d) final prediction and labeling of uncertain epochs in yellow.

6.3.2 Results

The classification performances on 30-s sleep epochs are shown in Tables 6.2 and 6.3 in the two cases of including all sleep epochs and only the *certain* ones, respectively. The percentages of epochs classified as *uncertain* were 9.95±4.15% in HC, 11.13±7.86% in PDnonRBD, 18.39±7.38% in iRBD and 18.90±8.00% in PD+RBD. In all 10 folds, the thresholds selected for labeling epochs as *uncertain* were $th_W = 0.73$, $th_{REM} = 0.46$ and $th_{NREM} = 0.48$.

The classification performances of the proposed CNN were compared to the ones obtained by previously developed methods [149, 169] in Table 6.1. These methods were trained and tested in different databases from the one used here and they aimed to also classify the three NREM sleep stages. The performances of Table 6.1 were obtained by adapting the original 5-class confusion matrices [149, 169] to a 3-class problem and by pulling all the subjects of the studied cohort together. The proposed method achieved performances that were superior or comparable to previously developed automated macro-sleep staging algorithms.

	HC			**iRBD**			**PD(+RBD)**		
	[149]	All epochs	*Certain* epochs	[149]	All epochs	*Certain* epochs	[169]	All epochs	*Certain* epochs
Subjects	53	41	41	53	31	31	77 (42% +RBD)	87 (34% +RBD)	87 (34% +RBD)
ACC [-]	0.92	0.88	0.91	0.82	0.76	0.85	0.81	0.83	0.90
κ [-]	0.84	0.78	0.83	0.63	0.60	0.75	0.64	0.70	0.81

Table 6.1: Comparison of the performances achieved by the proposed algorithm (when all epochs and only the *certain* ones were included) to the performances achieved by previous studies in similar cohorts.

6.3.3 Discussion

The proposed algorithm scores wakefulness, REM and NREM sleep in elderly and patients with neurodegeneration both at micro and macro-scale level with good performances. Two are the main innovations of the proposed method. First, it performs not only the traditional staging at a macro-scale level (i.e. 30-s epochs), but also at a micro-scale level (i.e. 5-s mini-epochs), with the objective of giving a better understanding of the sleep physiology and evolution. This is in line with recent studies showing the importance of overcoming the 30-s epoch AASM convention [165, 170]. Second, it is *interactive*, meaning that the 30-s epochs where a clear sleep pattern is not recognized are identified as *uncertain* and left to a sleep technician for manual revision.

The proposed CNN was trained on 5-s sleep mini-epochs of HC not located at sleep stage transitions. Moreover, the CNN did not contain layers introducing temporal dependency between mini-epochs (such as long short-term memory cells), making each 5-s mini-epoch prediction independent from each other. In this way, the algorithm could learn typical and clear time-frequency representation of wakefulness, REM and NREM sleep in 5-s mini-epochs in elderly HCs. Sleep epochs with unclear patterns could be present in transitional periods as well as in pathologies, and this is why it was decided to implement a strategy for identifying *certain* and *uncertain* epochs in macro-sleep staging. The higher number of *uncertain* epochs seen in patients with neurodegeneration reflect their altered sleep structure, which challenges both human scoring [171, 167] as well as the proposed system. It should be noticed that, even in patients with neurodegeneration, less than 20% of the sleep epochs (in average) were classified as *uncertain*. This means that, compared to traditional manual sleep staging, the proposed algorithm can still be considered an important tool for saving time in clinics.

The performances shown in Tables 6.2 and 6.3 are generally high. However, the main limitation is the low sensitivity for REM sleep in the presence of RBD. This was expected, as scoring REM sleep in iRBD and PD+RBD patients is challenging and often requires a close video analysis. However, the specificity for REM sleep is high, thus ensuring that the lower amount of epochs that were scored as REM sleep by the algorithm were actually true REM sleep.

A limitation of the proposed method is that manual annotations of sleep stages in 5-s mini-epochs were not available, therefore a validation of the algorithm on the micro-scale level could not be performed. However, as the macro-scale sleep stage predictions were obtained based on the micro-scale ones, the performances shown in Tables 6.2 and 6.3 can be used as a proof that also the micro-scale predictions are reliable.

Group	Overall		W			NREM			REM		
	ACC [-]	κ [-]	**ACC** [-]	**SENS** [-]	**SPEC** [-]	**ACC** [-]	**SENS** [-]	**SPEC** [-]	**ACC** [-]	**SENS** [-]	**SPEC** [-]
Train HC	0.86±0.04	0.75±0.08	0.94±0.03	0.85±0.09	0.96±0.03	0.88±0.04	0.85±0.07	0.95±0.05	0.91±0.05	0.95±0.08	0.90±0.06
Validation HC	0.88±0.04	0.77±0.07	0.94±0.03	0.82±0.11	0.96±0.03	0.90±0.04	0.87±0.06	0.93±0.05	0.92±0.04	0.91±0.09	0.92±0.05
Test HC	0.87±0.05	0.77±0.10	0.93±0.04	0.85±0.11	0.95±0.06	0.89±0.05	0.87±0.08	0.94±0.06	0.92±0.04	0.91±0.12	0.92±0.05
Test PDnonRBD	0.86±0.10	0.74±0.16	0.93±0.07	0.89±0.12	0.93±0.10	0.88±0.08	0.85±0.12	0.92±0.13	0.92±0.07	0.84±0.26	0.93±0.07
Test iRBD	0.76±0.10	0.60±0.15	0.84±0.10	0.91±0.10	0.81±0.15	0.82±0.08	0.76±0.15	0.90±0.13	0.86±0.08	0.56±0.32	0.93±0.09
Test PD+RBD	0.77±0.10	0.60±0.16	0.82±0.11	0.95±0.08	0.78±0.16	0.82±0.10	0.77±0.15	0.88±0.14	0.90±0.05	0.47±0.30	0.98±0.03

Table 6.2: Classification performances achieved by the proposed algorithm when all epochs (i.e. *certain* and *uncertain*) were included. The values are shown as mean and standard deviation across the participants in each group.

Group	Overall		W			NREM			REM		
	ACC [-]	κ [-]	**ACC** [-]	**SENS** [-]	**SPEC** [-]	**ACC** [-]	**SENS** [-]	**SPEC** [-]	**ACC** [-]	**SENS** [-]	**SPEC** [-]
Test HC	0.91±0.05	0.81±0.10	0.96±0.03	0.83±0.15	0.98±0.03	0.92±0.04	0.91±0.07	0.94±0.07	0.93±0.05	0.95±0.08	0.93±0.06
Test PDnonRBD	0.90±0.08	0.80±0.17	0.97±0.04	0.88±0.16	0.98±0.05	0.91±0.07	0.91±0.09	0.92±0.15	0.93±0.07	0.86±0.24	0.93±0.08
Test iRBD	0.85±0.09	0.73±0.15	0.93±0.07	0.91±0.13	0.93±0.11	0.88±0.07	0.87±0.11	0.89±0.16	0.89±0.08	0.66±0.35	0.93±0.10
Test PD+RBD	0.88±0.08	0.76±0.16	0.93±0.07	0.95±0.10	0.90±0.12	0.90±0.07	0.89±0.10	0.88±0.17	0.94±0.05	0.59±0.35	0.98±0.02

Table 6.3: Classification performances achieved by the proposed algorithm when only *certain* epochs were included. The values are shown as mean and standard deviation across the participants in each group.

6.4 Paper VI: A data-driven system to identify REM sleep behavior disorder and to predict its progression from the prodromal stage in Parkinson's disease

This Section presents a new method to identify RBD and prodromal RBD in PD patients. For this purpose, the following strategy was adopted. First, a fully automated data-driven system for identifying RBD in PD patients was developed. The developed system used features extracted from sleep micro-structure, EEG and EOG signals. In this way, it was possible to evaluate whether PD patients with RBD presented sleep electrophysiological abnormalities compared to PD patients without RBD, besides RSWA. Then, the developed system was used to evaluate whether PD patients with prodromal RBD had similar electrophysiological abnormalities to PD patients with RBD and whether they could be automatically identified.

The methodology of the proposed new system is shown in Fig. 6.8. The system used C4-A1, ROC-A1 and LOC-A2 signals to perform micro and macro-scale sleep staging (Section 6.3). After that, the epochs labeled as *certain* were considered and features describing micro-sleep structure, EEG spectral content, EEG coherence between left and right central derivations, EEG complexity, and EOG energy were extracted. These features were used to train a machine learning system based on random forest, aiming to identify RBD. The system returned as outputs the complementary probabilities that the given input PD patient had RBD ($P(RBD)$) or not ($P(nonRBD)$). The system was validated on a cohort of PD patients with and without RBD, and finally applied to evaluate whether it could identify PD patients with prodromal RBD.

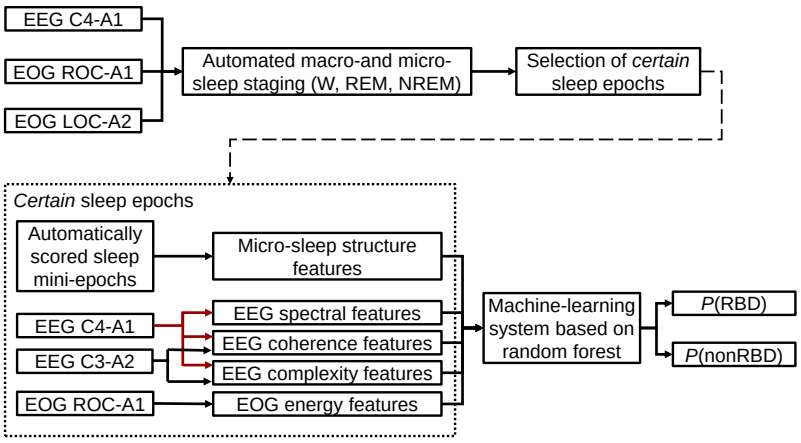


Figure 6.8: Schematic view of the methodology of Paper VI [155]. Macro- and micro-sleep staging was performed according to Section 6.3. In the *certain* epochs, features describing micro-sleep structure, EEG spectral content, EEG coherence, EEG complexity and EOG energy in the time-frequency domain were extracted. The features were used as input of a machine learning system returning the probability that a given PD patient had RBD or not ($P(RBD)$ and $P(nonRBD)$ respectively).

Parameter	PDnonRBD	PD+RBE	PD+RBD	p-value
Count	54	27	26 -	
M/F	31/23	20/7	17/9	0.33
Age [years]	62.13±10.05	63.19±10.39	64.85±7.90	0.48
Sleep efficiency [% of TIB]	76.71±11.17	75.41±9.88	76.03±8.83	0.69
N1 [% of TST]	21.71±9.81	24.76±9.07	23.58±7.74	0.07
N2 [% of TST]	50.56±9.44	46.77±9.58	48.75±9.68	0.33
N3 [% of TST]	9.53±7.87	6.35±8.86	7.33±8.60	0.07
REM [% of TST]	18.20±6.82	22.11±5.18	20.33±6.72	0.10
AHI [h^{-1}]	3.86±6.14	2.23±3.28	1.51±2.40	0.35
PLMS index [h^{-1}]	25.2±32.18	22.93±35.19	44.46±43.68	**0.02**
RSWA [% REM]	6.70±8.58	6.61±4.05	37.03±21.97	**<0.001**
Awakening index [h^{-1}]	4.74±2.62	5.29±3.97	3.74±1.88	0.13

Table 6.4: Demographic information of the cohort for identification of RBD and prodromal RBD in PD patients. Values are reported as mean and one standard deviation. Statistical analyses were performed with Kruskal-Wallis tests and chi-square test for gender. Significant p-values are highlighted in bold font. TIB: time in bed; TST: total sleep time; AHI: apnea/hypopnea index.

6.4.1 Methods: Development and validation of an EEG and EOG-based data driven system for RBD and prodromal RBD identification in PD patients

Subjects and recordings

The cohort here investigated included the baseline v-PSG of 107 PD patients, which were part of the "DeNoPa" cohort [14, 56, 150]. With respect to the PD patients shown in Table 5.6, 8 PD patients were excluded due to technical problems in applying the proposed system to them. More specifically, the cohort included 54 PD patients without RBD (PDnonRBD), 26 PD patient with RBD (PD+RBD) and 27 PD patients with RBEs (PD+RBE). This last group was defined as PD patients who showed at least two RBEs during the considered night, but who did not reach the RSWA cut-off criteria for definite RBD diagnosis [56]. Demographics and sleep information of these groups are shown in Table 6.4.

As part of the "DeNoPa" study, the patients were re-evaluated with v-PSG in a 2-year follow-up (FU). During the FU study, it was again evaluated whether the patients had normal REM sleep, had RBD or RBEs by applying the same criteria as in the baseline evaluation (Section 6.3.1). The only difference was that EMG signals from the flexor digitorum superficialis muscle (Fig. 2.3e) were also recorded at this time and used for making RBD diagnosis, according to the SINBAR method [100]. The diagnoses given at FU evaluation are shown in Table 6.5. The 9 PD+RBE patients at baseline who developed RBD at FU were considered the PD patients with prodromal RBD to identify. Only the group classification shown in Table 6.5 was available as information from FU and no v-PSG data from the FU evaluation was analyzed in this study.

Machine learning system development and evaluation

A compact overview of the machine learning system developed in this study is shown in Fig. 6.8. The full process is described below in details.

Baseline	**Follow-up**			
	PDnonRBD	**PD+RBE**	**PD+RBD**	**No REM sleep**
PDnonRBD (54)	31	11	9	3
PD+RBE (27)	8	8	9	2
PD+RBD (26)	0	0	23	3

Table 6.5: Group classification at FU evaluation for the cohort studied.

Automated macro- and micro-sleep staging algorithm The macro and micro-sleep staging algorithm described in Section 6.3 was applied to the participants included in this study. The algorithm classified each 5-s mini-epoch and each 30-s epoch as either W, REM or NREM sleep and additionally labeled each 30-s sleep epoch as either *certain* or *uncertain*. To further confirm the validity of the automated sleep staging classifier in this cohort, the macro-sleep staging predictions were compared to the manual sleep stages on the *certain* epochs and the overall accuracy (Eq. 5.14), overall Cohen's kappa (Eq. 6.9), and stage-specific accuracies, sensitivities, and specificities (Eq. 3.9) were calculated.

Feature extraction Feature extraction was performed by considering *certain* epochs only for all parameters. The extracted features could be divided into 5 categories, which are described below.

Features describing micro-sleep structure These features were obtained from the stage probabilities and predicted stages at a micro-scale level and were derived under the hypothesis that PD patients with RBD have altered sleep micro-structure compared to the ones without RBD. These features included:

- Wake-sleep transition index:

$$W - S_{transIndex} = \frac{(\#W \rightarrow S) + (\#S \rightarrow W)}{TST} \tag{6.11}$$

  calculated as the number of transitions between wakefulness (W) and sleep (S) and vice-versa per hour of total sleep time (TST). The TST was calculated by summing the amount of mini-epochs classified as either REM or NREM sleep.

- REM-NREM transition index:

$$REM - NREM_{transIndex} = \frac{(\#REM \rightarrow NREM) + (\#NREM \rightarrow REM)}{TST} \tag{6.12}$$

  calculated as the number of transitions between REM and NREM sleep and vice-versa per TST.

- Percentages of W, REM and NREM sleep:

$$\%_W = \frac{hW}{TIB} \quad \%_{NREM} = \frac{hNREM}{TST} \quad \%_{REM} = \frac{hREM}{TST} \tag{6.13}$$

  where hW, $hNREM$ and $hREM$ are the hours of wakefulness, NREM and REM sleep. TIB is the time in bed.

- W, REM and NREM stability indices:

$$W_{stabilityIndex} = \frac{\#W \rightarrow W}{hW} \tag{6.14}$$

  calculated as the number of transitions from W to W per hour of W. Similar equations hold for $REM_{stabilityIndex}$ and $NREM_{stabilityIndex}$.

- NREM and REM fragmentation indices:

$$NREM_{fragIndex} = \frac{\#NREM \rightarrow (W \text{ or } REM)}{hNREM} \tag{6.15}$$

calculated as the number of transitions from NREM sleep to another stage per hour of NREM sleep. A similar equation holds for $REM_{fragIndex}$.

- Global certainty [87]:

$$GlobCert = \frac{1}{N}\sum_{i=1}^{N} 1_{p^i_{ME}(k)\geq th} \tag{6.16}$$

This feature calculates the number of mini-epochs having at least one of the three probabilities $p^i_{ME}(k)$ with $k \in \{W, REM, NREM\}$ higher than a threshold th, with i representing the i^{th} of the N mini-epochs. Ten values of th were used ($th \in \{0.50, 0.55, ..., 0.95\}$), thus obtaining 10 sets of global certainty features in total.

- W, REM and NREM certainty:

$$WCert = \frac{1}{N_W}\sum_{i=1}^{N_W} 1_{p^i_{ME}(W)\geq th} \tag{6.17}$$

This feature represents the percentage of mini-epochs classified as W with probability of W higher than the threshold th. N_W is the amount of mini-epochs classified as W. The values of th used were $th \in \{0.50, 0.55, ..., 0.95\}$, thus obtaining 10 sets of features. Similar equations hold for NREM and REM sleep.

- Global stability [87]:

$$GlobStab = \frac{1}{\bar{S}}\sum_{i=1}^{\bar{S}} \bar{L}_i \quad \bar{L}_i = \frac{\bar{l}_i - \min\{\bar{l}_1, \bar{l}_2, ..., \bar{l}_{\bar{S}}\}}{\max\{\bar{l}_1, \bar{l}_2, ..., \bar{l}_{\bar{S}}\} - \min\{\bar{l}_1, \bar{l}_2, ..., \bar{l}_{\bar{S}}\}} \tag{6.18}$$

In this feature, $\bar{l}_i$ is the length of the i^{th} mini-epoch segment where the sleep stage predicted was the same. $\bar{S}$ is the total number of mini-epoch segments with the same sleep stage prediction. Such lengths were normalized ($\bar{L}_i$) for each participant with respect to the minimum and maximum lengths of segments.

- W, REM and NREM stabilities:

$$WStab = \frac{1}{\bar{S}_W}\sum_{i=1}^{\bar{S}_W} \bar{L}_{W,i} \quad \bar{L}_{W,i} = \frac{\bar{l}_{W,i} - \min\{\bar{l}_{W,1}, \bar{l}_{W,2}, ..., \bar{l}_{W,\bar{S}_W}\}}{\max\{\bar{l}_{W,1}, \bar{l}_{W,2}, ..., \bar{l}_{W,\bar{S}_W}\} - \min\{\bar{l}_{W,1}, \bar{l}_{W,2}, ..., \bar{l}_{W,\bar{S}_W}\}} \tag{6.19}$$

This feature is similar to the global stability, but only the wake segments were included. Analogous equations hold for NREM and REM stabilities.

- Amount of W, REM and NREM [87]:

$$W_{amount} = \frac{1}{N}\sum_{i=1}^{N} 1_{p^i_{ME}(W)\geq th} \tag{6.20}$$

This feature represents the amount of mini-epochs having probabilities of W higher then the threshold th. A total of 10 values for th were used ($th \in \{0.50, 0.55, ..., 0.95\}$). N is the total number of mini-epochs. Similar equations hold for the amount of NREM and REM sleep.

- Features based on Markov matrix probability [180]. A Markov transition matrix was built based on the number of stage transitions at micro-scale level:

$$\mathbf{TM} = \begin{bmatrix} \#W \rightarrow W & \#W \rightarrow NREM & \#W \rightarrow REM \\ \#NREM \rightarrow W & \#NREM \rightarrow NREM & \#NREM \rightarrow REM \\ \#REM \rightarrow W & \#REM \rightarrow NREM & \#REM \rightarrow REM \end{bmatrix} \tag{6.21}$$

  The matrix was normalized for each row by calculating probability of transitions between stages, after setting to 0 the values in the diagonal:

$$\mathbf{T\bar{M}} = \begin{bmatrix} 0 & p(W \rightarrow NREM) & p(W \rightarrow REM) \\ p(NREM \rightarrow W) & 0 & p(NREM \rightarrow REM) \\ p(REM \rightarrow W) & p(REM \rightarrow NREM) & 0 \end{bmatrix} \tag{6.22}$$

  Due to complementary values in each row, the features used were $p(W \rightarrow REM)$, $p(NREM \rightarrow W)$, and $p(REM \rightarrow W)$. Moreover, the zero-memory Markov model entropy rate was calculated for each stage and used as feature:

$$\begin{gathered} H0_W = -[p(W-NREM) \cdot log(p(W-NREM)) + \\ p(W-REM) \cdot log(p(W-REM))] \\ H0_{NREM} = -[p(NREM-W) \cdot log(p(NREM-W)) + \\ p(NREM-REM) \cdot log(p(NREM-REM))] \\ H0_{REM} = -[p(REM-W) \cdot log(p(REM-W)) + \\ p(REM-NREM) \cdot log(p(REM-NREM))] \end{gathered} \tag{6.23}$$

EEG spectral features These features describe the EEG frequency content and were derived based on the hypothesis of more pronounced EEG slowing in PD patients with RBD compared to the ones without RBD [159].

The C4-A1 signal was considered for computing such features. The signal was pre-processed by applying the adaptive filtering described in Section 6.3.1 to remove EOG and ECG interference. Moreover, the algorithm proposed by Brunner et al. [181] was used to identify areas of the C4-A1 signal with EMG artifacts. Briefly, the EEG signal was divided into non-overlapping 4-s windows and for each window the spectral power density in the frequency range 26.25-32.0 Hz was calculated with fast Fourier transform [182]. For each 4-s window, the median of the 45 values in the surrounding symmetric 3-min window was chosen to estimate the *local* level of high frequency activity. A 4-s window was considered containing muscular activity artifact if its high frequency content was higher than 4 times the local level of high frequency.

After the described pre-processing, the 30-s *certain* sleep epochs classified as REM or NREM sleep were considered, and the ones located at a sleep stage transitions were removed for having more stable spectral estimation [183]. The remaining epochs were divided into 50% overlapping 2-s windows and the ones containing EMG artifacts removed from analysis. For each 2-s window, the spectral density was estimated with the auto/cross-covariance matrix technique [184]. This consists in the following process. An EEG matrix $\mathbf{X}$ with N samples and M channels can be represented as

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1M} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{NM} \end{bmatrix} \tag{6.24}$$

From this, the auto-covariance ($\hat{\gamma}_{11}(s)$) of the EEG signal $\mathbf{x}_1 = \{x_{11}, x_{21}, ..., x_{N1}\}$ can be estimated as

$$\hat{\gamma}_{1,1}(s) = \begin{cases} \frac{1}{N-s} \sum_{n=0}^{N-s-1} \left(x_{n+s,1} - \frac{1}{N} \sum_{i=0}^{N-1} x_{i,1} \right) \cdot \left(x_{n,1} - \frac{1}{N} \sum_{i=0}^{N-1} x_{i,1} \right) & s \geq 0 \\ \hat{\gamma}_{11}(-s) & s < 0 \end{cases} \tag{6.25}$$

The cross-covariance can be calculated similarly, by considering two different EEG channels. The auto/cross-coveriance matrix is then defined as:

$$\hat{\Gamma}(s) = \begin{bmatrix} \hat{\gamma}_{1,1}(s) & \cdots & \hat{\gamma}_{1,M}(s) \\ \vdots & \ddots & \vdots \\ \hat{\gamma}_{M,1}(s) & \cdots & \hat{\gamma}_{M,M}(s) \end{bmatrix} \tag{6.26}$$

In other words, the auto/cross-covariance describes the similarity between EEG samples at different time points as a function of the time lag s between them. From this, it is possible to estimate the spectral density $\mathbf{S}(f)$ at the (normalized) frequency f as:

$$\mathbf{S}(f) = \frac{1}{2\pi} \sum_u \hat{w}(u) \hat{\Gamma}(u) e^{-ifu} = \begin{bmatrix} S_{1,1}(f) & \cdots & S_{1,M}(f) \\ \vdots & \ddots & \vdots \\ S_{M,1}(f) & \cdots & S_{M,M}(f) \end{bmatrix} \tag{6.27}$$

where $\hat{w}(u)$ is the Parzen window (of length 255 in this case) [185]. This means that the diagonal (i, i) elements represent the auto-spectral density of the i^{th} channel and the off-diagonal (i, j) elements represent the cross-spectral density between the i^{th} and j^{th} channel.

As only the C4-A1 channel was used, its auto-spectral density was available for each 2-s window. Each auto-spectral density was cut in the frequency band [0-35] Hz and normalized to have unitary area. The average spectral density in REM and NREM sleep ($\hat{S}^{REM}_{C4A1,C4A1}(f)$ and $\hat{S}^{NREM}_{C4A1,C4A1}(f)$) were finally obtained by averaging the spectral densities of the 2-s windows included in the analysis. The following features were then extracted from $\hat{S}^{i}_{C4A1,C4A1}(f)$ with $i \in \{REM, NREM\}$:

- Relative spectral power in specific frequency bands, calculated as the ratio between the area under the spectral density in a specific frequency band and the area of the spectral density in the range 0.5-35 Hz. The following frequency bands were considered: δ ([0.5-4) Hz), δ_1 ([0.5-2) Hz), δ_2 ([2-4) Hz), θ ([4-8) Hz), θ_1 ([4-6) Hz), θ_2 ([6-8) Hz), α ([8-12) Hz), α_1 ([8-10) Hz), α_2 ([10-12) Hz), σ ([12-16) Hz), σ_1 ([12-14) Hz), σ_2 ([14-16) Hz), β ([16-35) Hz), β_1 ([16-25) Hz), β_2 ([25-35) Hz).

- Central frequency, defined as:

$$CF = \frac{\sum_{t=1}^{N} \hat{S}^{i}_{C4A1,C4A1}(f_t) \cdot f_t}{\sum_{t=1}^{N} \hat{S}^{i}_{C4A1,C4A1}(f_t)} \tag{6.28}$$

  where f_t is the t^{th} frequency sample and N is the total number of frequency samples.

- Normalized spectral entropy, defined as:

$$SpecEnt = \frac{-\sum_{t=1}^{N} p(t) \cdot \log_2 p(t)}{\log_2 N} \quad \text{with } p(t) = \frac{\hat{S}^{i}_{C4A1,C4A1}(f_t)}{\sum_{r=1}^{N} \hat{S}^{i}_{C4A1,C4A1}(f_r)} \tag{6.29}$$

  where f_t is the t^{th} frequency sample and N is the total number of frequency samples.

- Maximum peak of the average spectral density in the δ band.
- Slow-to fast ratio defined as the ratio between the *log*-transformed relative power in δ_1 and β_1 bands.

An additional analysis was performed for REM sleep [107]. Each spectral density in a 2-s window in REM sleep was standardized with respect to $\hat{S}^{NREM}_{C4A1,C4A1}(f)$. The so obtained standardized spectral densities were then averaged, thus obtaining a final average REM sleep standardized spectral density ($\tilde{S}^{REM}_{C4A1,C4A1}(f)$). From this, the following features were calculated:

- Average values of $\tilde{S}^{REM}_{C4A1,C4A1}(f)$ in δ, δ_1, δ_2, θ, θ_1, θ_2, α, α_1, α_2, σ, σ_1, σ_2, β, β_1, and β_2 bands.

EEG coherence features These features were extracted to investigate functional connectivity, with the hypothesis that PD patients with RBD have lower connectivity than the ones without RBD. To calculate that, the auto and cross-spectral density for C4-A1 and C3-A2 EEG signals were calculated according to Eqs. 6.25, 6.26 and 6.27 for 2-s windows. Only the 2-s windows that were included for calculating EEG spectral features were considered. For each of these windows, the coherence was calculated as:

$$Coh(f) = \frac{|S_{C3A2,C4A1}(f)|^2}{S_{C4A1,C4A1}(f) \cdot S_{C3A2,C3A2}(f)} \tag{6.30}$$

The average coherence curves in NREM and REM sleep were calculated and from these the following features computed for both curves:

- Median coherence value in δ, δ_1, δ_2, θ, θ_1, θ_2, α, α_1, α_2, σ, σ_1, σ_2, β, β_1, and β_2 bands.
- Area under the coherence curve.

EEG complexity features A number of features describing EEG complexity were also derived, under the hypothesis that PD patients with RBD have lower EEG complexity than the ones without RBD. Several measures of complexity were calculated, which are described below.

EEG entropy: Entropy was calculated for C4-A1 EEG signal and the same 2-s windows used for computing EEG spectral features were used. For each window, the EEG values were divided into a histogram with $\lceil\sqrt{N}\rceil$ bins, where N is the number of EEG samples in a 2-s window. From the histogram, the probabilities for each bin were computed, thus obtaining a probability vector $\mathbf{p} = \{p(bin_1), p(bin_2), ..., p(bin_{\lceil\sqrt{N}\rceil})\}$. The vector was used to compute the entropy as:

$$Ent = \frac{-\sum_{i=1}^{\lceil\sqrt{N}\rceil} p(bin_i) \cdot \log_2 p(bin_i)}{\log_2 \lceil\sqrt{N}\rceil} \tag{6.31}$$

From this, the following features were calculated:

- Average entropies in REM and NREM sleep, obtained by averaging the entropy values in 2-s windows during REM and NREM sleep, respectively.
- Standard deviation of REM and NREM entropies, calculated by considering the entropy values in 2-s windows during REM and NREM sleep, respectively.

Multiscale sample entropy: Multiscale sample entropy was calculated for each of the 2-s windows of C4-A1 EEG used also for computing EEG spectral features. From each EEG window $\mathbf{x} =$

$[x_1, x_2, ..., x_N]$, template vectors with scale $\bar{\delta}$ were obtained as $\mathbf{x}_{m,\bar{\delta}}(i) = [x_i, x_{i+\bar{\delta}}, x_{i+2\cdot\bar{\delta}}, ..., x_{i+(m-1)\cdot\bar{\delta}}]$. The sample entropy at scale $\bar{\delta}$ was calculated as:

$$SampEn = -\log \frac{A_{\bar{\delta}}}{B_{\bar{\delta}}} \tag{6.32}$$

where $A_{\bar{\delta}}$ is the number of template vectors with $\bar{d}[\mathbf{x}_{m+1,\bar{\delta}}(i), \mathbf{x}_{m+1,\bar{\delta}}(j)] < \bar{r}$ and $B_{\bar{\delta}}$ is the number of template vectors with $\bar{d}[\mathbf{x}_{m,\bar{\delta}}(i), \mathbf{x}_{m,\bar{\delta}}(j)] < \bar{r}$ (where $\bar{d}[\cdot,\cdot]$ was defined as the Chebychev distance) [186, 187]. The sample entropy was computed for $m = 2$, $\bar{r} = 0.15$ and $\bar{\delta}$ varying in the range [1,2,...,20], thus obtaining the *multiscale* sample entropy for each 2-s EEG window. The following features were calculated:

- The average multiscale sample entropy values in REM and NREM sleep, as average over the multiscale sample entropy over the 2-s windows in REM and NREM sleep respectively.

Mutual information: Given two signals signals $\mathbf{x}$ and $\mathbf{y}$, their cross-mutual information is defined as:

$$MI(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{x},\mathbf{y}} p_{\mathbf{x},\mathbf{y}}(\mathbf{x}, \mathbf{y}) \log \frac{p_{\mathbf{x},\mathbf{y}}(\mathbf{x}, \mathbf{y})}{p_{\mathbf{x}}(\mathbf{x}) \cdot p_{\mathbf{y}}(\mathbf{y})} \tag{6.33}$$

where $p_{\mathbf{x},\mathbf{y}}(\mathbf{x}, \mathbf{y})$ is the joint probability distribution between the signals $\mathbf{x}$ and $\mathbf{y}$, and $p_{\mathbf{x}}(\mathbf{x})$ and $p_{\mathbf{y}}(\mathbf{y})$ are the marginal probability distributions. These probability distributions were estimated in the same way as for entropy. The auto-mutual information for the signal $\mathbf{x}$ is obtained by considering the signal $\mathbf{y}$ equal to $\mathbf{x}$. For each 2-s window selected also for EEG spectral frequency calculation, the auto-mutual information for C4-A1 and the cross-mutual information between C4-A1 and C3-A2 channels were calculated. The cross-mutual information between C4-A1 and C3-A2 was computed by using a shifted version of C3-A2 up to 34 samples. The same was done for the auto-mutual information for the signal C4-A1. The following features were then extracted:

- The average auto-mutual and cross-mutual information values in REM and NREM sleep for lags spanning from 0 to 34 samples.

EOG energy features Based on previous studies showing alterations of EOG related to neurodegeneration [79, 81], EOG features were extracted. The sleep epochs classified as *certain* by the sleep staging algorithm were considered and, to make the analysis more stable, the ones located at sleep stage transitions removed. For the remaining epochs, the ROC-A1 EOG signal was divided into 50% overlapping windows of 10 s. For each window, the discrete wavelet transform (DWT) with Daubechies 4 mother wavelet (Section 4.3.1) was calculated up to the 8^{th} level, thus obtaining detail coefficients $d_1, d_2, ..., d_8$ and approximation coefficients a_8. From these coefficients, the total energy was calculated as:

$$E_{tot} = \sum_{i=1}^{8} \sum_{j=1}^{N_i} d_i(j)^2 + \sum_{j=1}^{N_8} a_8(j)^2 \tag{6.34}$$

where N_i is the length of the coefficient vectors $i = \{1, ..., 8\}$. Moreover, the energies for each detail band d_i (with $i = \{1, ..., 8\}$) and for the approximation band were calculated as:

$$E_{d_i} = \sum_{j=1}^{N_i} d_i(j)^2 \quad E_{a_8} = \sum_{j=1}^{N_8} a_8(j)^2 \tag{6.35}$$

The relative energies in each detail and approximation bands were obtained as the ration of the energies in the bands and the total energy. The following features were calculated:

- Average relative energy in each detail and approximation band during NREM and REM sleep.
- Standard deviation of the relative energy in each detail and approximation band during NREM and REM sleep.

In total, 396 features were extracted from each participant.

Construction of the machine learning system The features were used to build a machine learning system giving as outputs the probabilities that a given participant had RBD ($P(RBD)$) or not ($P(nonRBD)$) (Fig. 6.8). Random forest (RF) was the classifier used in this system because of its good performances, robustness to overfitting, and the useful property of ranking features (Section 5.3.1). To build a robust model capturing the electrophysiological characteristics of RBD in PD patients, PDnonRBD and PD+RBD patients that were classified in the same groups at both baseline and FU evaluations were considered for training and validation (Table 6.5). Therefore, the training and validation cohort included a total of 31 PDnonRBD and 23 PD+RBD patients. Due to the relative small number of patients, R=10 runs of K=5 fold cross-validation were implemented to make the system more generalizable. By doing so, a total of 50 training and 50 validation datasets were available.

Preliminary feature reduction A preliminary feature reduction was performed to make the machine learning training faster. For simplicity of explanation, the global certainty feature (Eq. 6.16) is now considered. This feature was calculated for 10 values of threshold th_t: $th_1 = 0.50$, $th_2 = 0.55$,...,$th_{10} = 0.95$. The objective of the feature reduction was to keep only the global certainty in correspondence of the threshold value that could best distinguish PD+RBD from PDnonRBD patients.

To do so, for each value of th_t and for each of the r, k^{th} training sets (with $r = \{1, 2, ..., R\}$ and $k = \{1, 2, ..., K\}$), a receiver operating characteristic analysis discriminating between PDnonRBD and PD+RBD was performed, and the corresponding area under the curve ($AUC_{r,k,t}$) calculated. The average AUC_t was calculated across the 50 training sets and TH was identified as the threshold th_t corresponding to the highest AUC_t value. Only the global certainty values calculated for TH were kept for further analysis. This process is illustrated in Fig. 6.9.

The same procedure was repeated for W/NREM/REM certainty (Eq. 6.17), W/NREM/REM amount (Eq. 6.20), multiscale entropy in NREM/REM (Eq. 6.32, where the best scale value was selected among the 20 computed), auto-mutual and cross-mutual information in NREM/REM (Eq. 6.33, where the best lag was chosen among the 35 computed). By doing so, the features were reduced from 396 to a subset of 159 features, which still contained features from all categories described above.

Structure, training and validation of the system The strategy adopted to train and validate the machine learning system is depicted in Fig. 6.10. In each r, k^{th} run of the cross-validation, a total of 10 RF models were trained on the training set by using all 159 features. The 159 features were then ranked according to their average feature importance across the 10 trained RF models. From the ranked features, the 50 most important ones were considered and used to train, on the same training set, J=50 RF models: the first one (j=1) included the most important feature, the second one (j=2) the two most important features, etc.

As this procedure was repeated for each r, k^{th} run of the cross-validation, a total of 2500 RF models were available ($\bar{M}_{r,k,j}$ with $r = \{1, 2, ..., R\}$, $k = \{1, 2, ..., K\}$ and $j = \{1, 2, ..., J\}$), of which 50 were trained with one feature ($\bar{M}_{r,k,1}$), 50 with two features ($\bar{M}_{r,k,2}$), etc. Each of these models

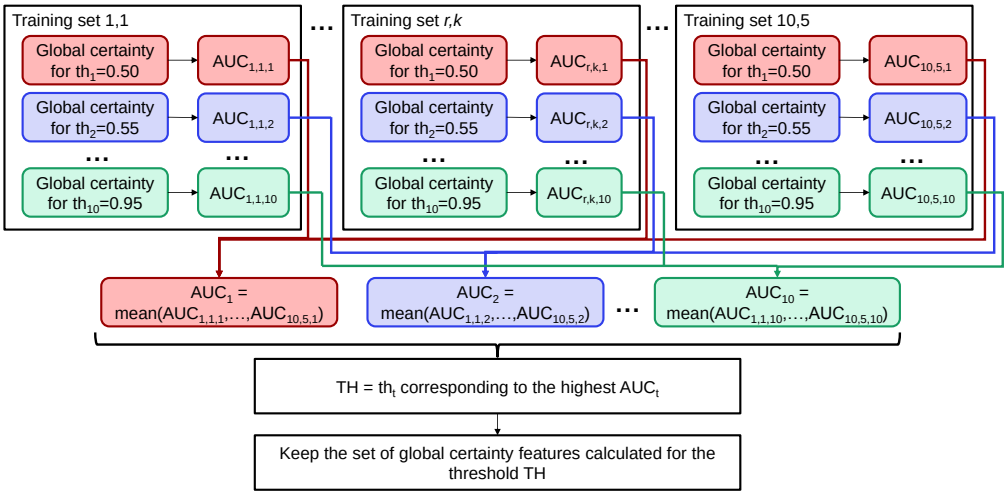


Figure 6.9: Schematic overview of the preliminary feature reduction applied to the global certainty feature to select the optimal threshold. For each of the r, k^{th} training sets and for each set of global certainty features corresponding to the ten threshold values th_t, ROC to distinguish between PDnonRBD and PD+RBD were calculated. The respective area under the curve ($AUC_{r,k,t}$) values were computed and the average value across the r, k training sets computed. The threshold TH corresponding to the highest AUC_t value was selected.

returned the probabilities (summing to 1) of having RBD ($p(RBD)$) or not ($p(nonRBD)$) for the respective training and validation sets.

During the training of each RF model, the following hyperparameters were optimized with a Bayesian optimization approach (Section 4.3.1) in an inner 5-fold cross validation (Fig. 4.8): the number of trees (positive integers, log-scaled in the range [10, 500]), their depth (which consisted in the optimization of the maximum number of splits and the minimum leaf size, which were searched among the integers log-scaled in the range $[1, \max(2, N-1)]$ and $[1, \max(2, \lfloor N/2 \rfloor))]$, with N being the number of training samples) and the number of features selected randomly (integers in the range $[1, \max(2, N_{feat})]$, where N_{feat} is the number of features).

Selection of the optimal set of RF models The optimal set of RF models was selected as the set of 50 models trained with N_{opt} features (i.e. $\bar{M}_{r,k,N_{opt}}$). The value N_{opt} was identified by following the procedure illustrated in Fig. 6.11. The 50 RF models trained with one feature ($\bar{M}_{r,k,1}$) were first considered. Because of the 10 runs of 5-fold cross-validation, each participant was included 40 times in a training set and 10 times in a validation set. For each participant, it was then possible to calculate the final training and validation probabilities of having RBD or not ($P(RBD)$ and $P(nonRBD)$ respectively) as the average of the 40 training and 10 validation probabilities respectively. A participant was automatically classified as PD+RBD if $P(RBD) \geq P(nonRBD)$ (i.e. $P(RBD) \geq 0.5$), and as PDnonRBD otherwise. By comparing the predictions with the true groups for all participants, the training and validation accuracy, sensitivity and specificity for RBD identification were computed (Eq. 3.8). This procedure was repeated for the set of 50 RF models trained with 2 features ($\bar{M}_{r,k,2}$), for the set of 50 RF models trained with 3 features ($\bar{M}_{i,k,3}$), etc. N_{opt} was set equal to the value of j corresponding to the set of 50 RF models achieving the highest validation accuracy.

Feature relevance To evaluate the relevance of each feature, it was counted in how many of

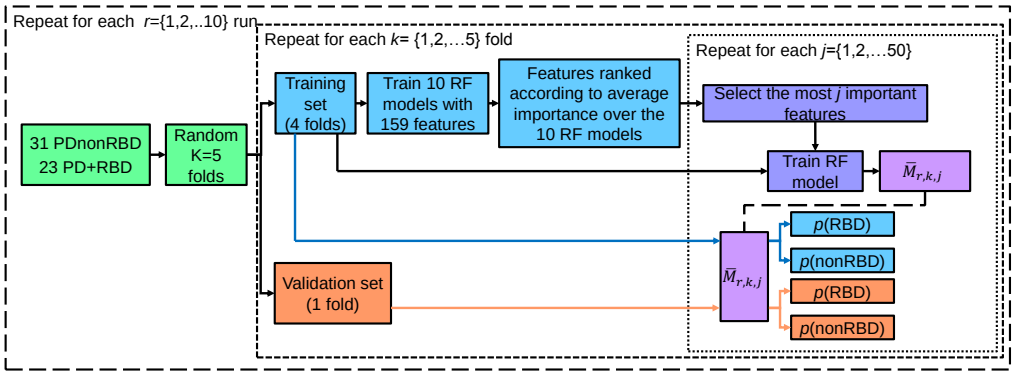


Figure 6.10: Structure of training and validation of the proposed system. For each training set, 10 RF models were trained based on all 159 features and the feature importance was calculated for each model. The features were then ranked according to the average feature importance and J=50 RF models were trained by considering the j most important features (with $j = \{1, 2, ..., 50\}$). Overall, 2500 models $\bar{M}_{r,k,j}$ were obtained (with $r = \{1, 2, ..., R\}$, $k = \{1, 2, ..., K\}$ and $j = \{1, 2, ..., J\}$). Each trained model was applied to the respective training (blue line) and validation (orange line) sets, obtaining training and validation probabilities of having RBD ($p(RBD)$) or not ($p(nonRBD)$) for each participant.

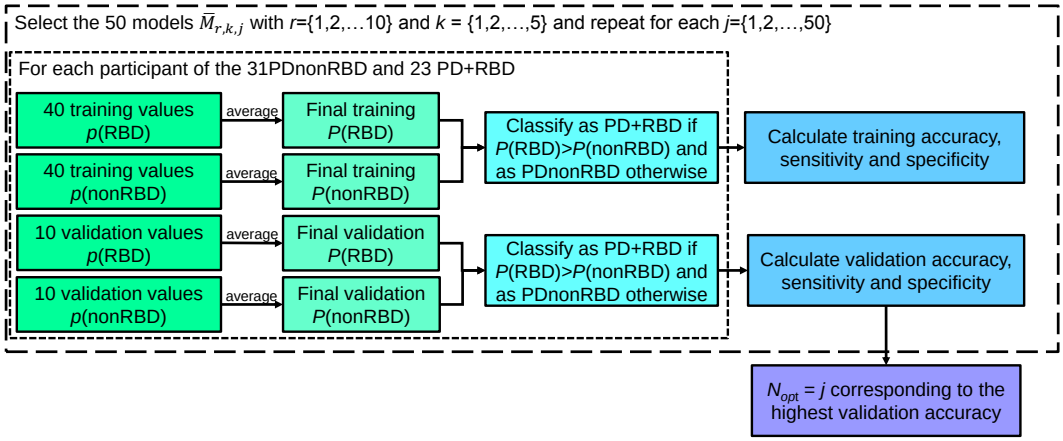


Figure 6.11: Schematic overview of the selection of the optimal set of 50 RF models, trained with N_{opt} features. For each set of 50 $\bar{M}_{r,k,j}$ RF models trained with j features, each participant was included 40 times in a training set and 10 times in a validation set, therefore 40 values of training $p(RBD)$ and $p(nonRBD)$ and 10 values of validation $p(RBD)$ and $p(nonRBD)$ were available. The average values were calculated ($P(RBD)$ and $P(nonRBD)$). These were used to classify the participant as PD+RBE if $P(RBD) \geq P(nonRBD)$ and as PDnonRBD otherwise, and to compute training and validation accuracy, sensitivity, and specificity. The parameter N_{opt} was set to the value of j corresponding to the highest validation accuracy.

the 50 optimal RF models ($\bar{M}_{r,k,N_{opt}}$) each feature φ_i (with $i = \{1, 2, ..., 159\}$) was included as:

$$rel_{\varphi_i} = \sum_r \sum_k 1_{r,k} \tag{6.36}$$

where $1_{r,k}$ was set equal to 1 if the i^{th} feature was included in the model $\bar{M}_{r,k,N_{opt}}$ and 0 otherwise. From this, it was possible to calculate the relative relevance of each feature φ_i as:

$$REL_{\varphi_i} = \frac{rel_{\varphi_i}}{50 \cdot N_{opt}} \tag{6.37}$$

Test of the optimal machine learning system and identification of prodromal RBD The 50 optimal RF models trained with N_{opt} features were applied to the remaining test dataset of 23 PDnonRBD, 27 PD+RBE and 3 PD+RBD patients. For each patient, 50 values of $p(RBD)$ and $p(nonRBD)$ were therefore obtained, which were averaged to obtain the final probabilities $P(RBD)$ and $P(nonRBD)$.

For the 23 PDnonRBD and the 3 PD+RBD patients, each participant was automatically classified as PD+RBD if $P(RBD) \geq P(nonRBD)$ (i.e. $P(RBD) \geq 0.5$) and as PDnonRBD otherwise. From this, the percentage of PDnonRBD correctly classified as not having RBD and the percentage of PD+RBD correctly classified as having RBD were calculated.

Furthermore, it was evaluated whether the developed system was able to identify patients with prodromal RBD. To do so, the 27 PD+RBE patients were divided into two subgroups: one including the patients who showed normal REM sleep or RBEs at FU (16 participants), and the other including the ones that developed definite RBD at FU (9 participants), as shown in Table 6.5. The second sugroup includes the PD participants with prodromal RBD, as these participants developed RBD after 2 years. A participant was classified as having prodromal RBD if $P(RBD) \geq P(nonRBD)$ (i.e. $P(RBD) \geq 0.5$). The accuracy, sensitivity and specificity for prodromal RBD detection was calculated.

6.4.2 Results

In total, 88.78±8.01%, 86.94±6.88% and 81.80±7.60% of the sleep epochs were classified as *certain* by the sleep staging algorithm in PDnonRBD, PD+RBE and PD+RBD groups, respectively. Table 6.6 shows the performances of the sleep staging algorithm in *certain* epochs, when compared to manual scoring. The overall agreement between manual and automatic sleep staging is good, with the main limitation being the low sensitivity for REM sleep in PD+RBD patients.

N_{opt} was found to be 14 and Table 6.7 shows the performance of the system during training and validation: PD+RBD patients could be successfully distinguished from PDnonRBD patients with validation accuracy, sensitivity and specificity over 80%.

In Fig. 6.12 the distributions of the relative feature relevance (REL) across the five different feature categories are shown. The features describing micro-sleep structure were the most important in distinguishing PD+RBD from PDnonRBD, followed by EEG spectral features, EEG complexity features, EOG energy features, and finally by EEG coherence features. Table 6.8 shows the distributions of the features that were selected in at least half of 50 optimal RF models. From the distributions, it can be seen that PD+RBD patients had significantly higher amount of wake-sleep transitions, increased REM sleep instability, more pronounced EEG slowing in REM sleep, decreased EEG complexity in REM sleep and altered EOG patterns in REM sleep when compared to PDnonRBD patients.

Parameter		PDnonRBD	PD+RBE	PD+RBD
Overall	**ACC** [%]	90.45±7.80	90.70±7.68	89.30±7.63
	κ [-]	0.81±0.15	0.83±0.13	0.79±0.15
W	**ACC** [%]	96.84±3.77	95.41±7.09	93.44±6.79
	SENS [%]	89.06±13.75	87.75±16.39	95.66±8.77
	SPEC [%]	98.09±5.02	96.28±9.10	92.15±9.69
NREM	**ACC** [%]	91.48±6.81	92.20±5.42	90.62±7.44
	SENS [%]	90.19±9.34	91.29±8.91	90.39±9.48
	SPEC [%]	94.31±6.53	93.34±6.12	89.05±13.98
REM	**ACC** [%]	92.57±7.54	93.80±5.44	94.54±4.53
	SENS [%]	91.09±13.74	84.22±22.67	64.46±31.36
	SPEC [%]	92.88±8.28	95.01±6.94	98.22±1.89

Table 6.6: Performances of the automated macro-sleep staging algorithm compared to manual sleep annotations when only *certain* epochs were considered

	Accuracy [%]	**Sensitivity** [%]	**Specificity** [%]
Training	83.33	77.42	91.30
Validation	81.48	80.65	82.61

Table 6.7: Performance of the proposed machine learning system in the training and validation datasets.

In Fig. 6.13 the distribution of $P(RBD)$ in the test dataset is shown. As expected, PDnonRBD patients had low values of $P(RBD)$ and 82.61% of them could be correctly classified as not having RBD. Accordingly, PD+RBD patients had high values of $P(RBD)$ and 2 out of 3 (66.67%) could be correctly classified as having RBD. PD+RBE patients showed an intermediate distribution of $P(RBD)$. The distribution of $P(RBD)$ in PD+RBE patients, divided in subgroups according to their classification at FU, is shown in Fig. 6.14. PD+RBE patients who developed RBD at FU could be differentiated by the ones that did not with accuracy of 80%, sensitivity of 66.67% and specificity of 87.5%. These results suggest that PD patients with prodromal RBD already show electrophysiological features similar to patients with definite RBD. Despite absence of RSWA, prodromal RBD could be identified with high sensitivity in PD patients.

6.4.3 Discussion

The previous sections presented the development and validation of a fully automated data-driven system that, based only on two EEG and two EOG signals recorded during sleep, is able to i) perform macro- and micro-sleep staging; ii) extract features describing micro-sleep structure, spectral EEG content, EEG coherence, EEG complexity, and EOG energy; iii) use these features for estimating the probability of RBD for a given PD patient. The proposed system could identify RBD in PD patients with over 80% accuracy, sensitivity, and specificity and prodromal RBD in PD patients with accuracy and specificity over 80%. The analysis of features relevance revealed that features describing micro-sleep structure were the most important for identifying patients with RBD and it was found that PD+RBD had increased micro-sleep instability compared to PDnonRBD. It was also found that PD+RBE patients who developed RBD after two years had already high values of probability of RBD at baseline. This suggests that PD patients with prodromal RBD already

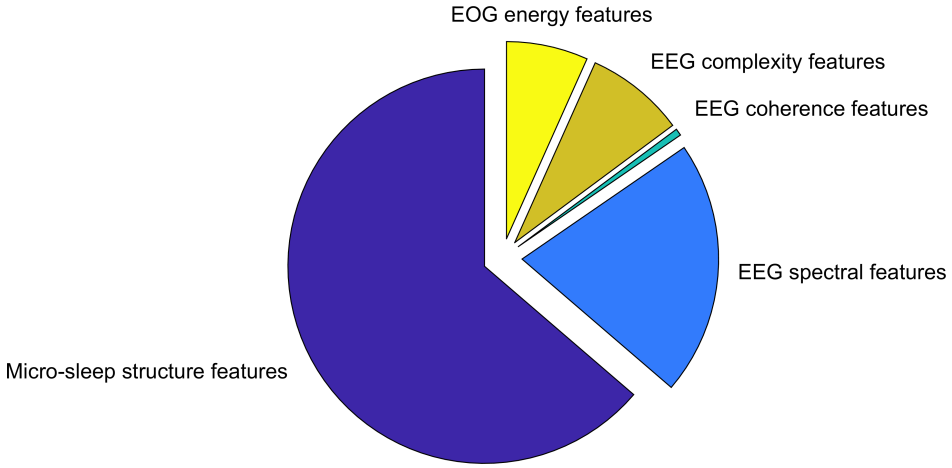


Figure 6.12: Distribution of the relative relevance of features in the optimal set of 50 random forest models. Features describing micro-sleep structure sum up to a relative importance of 63.7%, EEG spectral features of 20.9%, EEG complexity features of 8.1%, EOG energy features of 6.7% and EEG coherence of 0.6%.

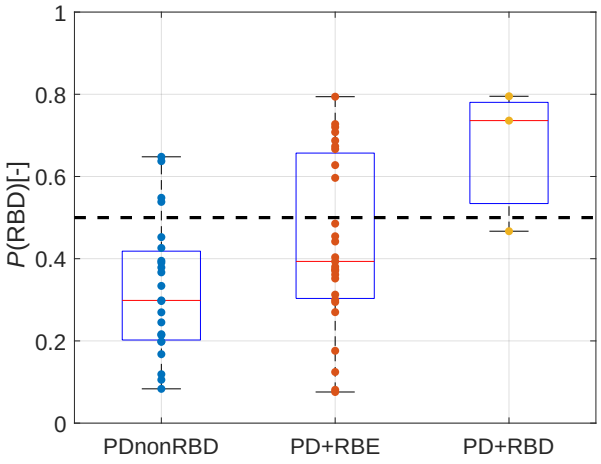


Figure 6.13: Distribution of $P(RBD)$ in the test dataset of 23 PDnonRBD, 27 PD+RBE and 3 PD+RBD. The dashed black line represents the probability value of 0.5, which was used as cut-off for classification.

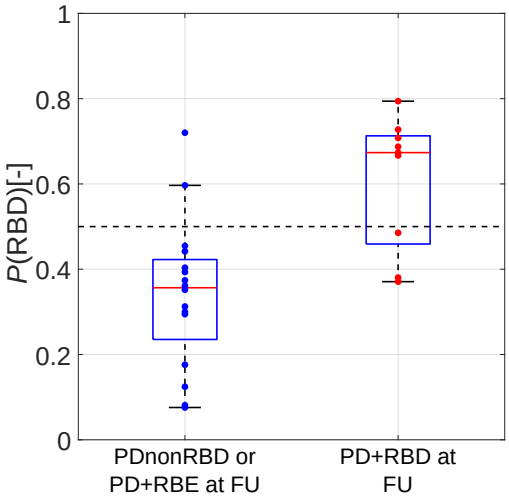


Figure 6.14: Distribution of $P(RBD)$ in PD+RBE patients at baseline subgroupped according to their classification FU. The dashed black line represents the probability value of 0.5, which was used as cut-off for classification.

Category	Feature	# models where selected	PDnonRBD (A)	PD+RBE (B)	PD+RBD (C)	p-value
Micro-sleep structure features	$W - S_{transIndex}$ $[h^{-1}]$	48	119.82±53.99	137.66±53.03	174.61±71.39	**<0.001**, A<C, B<C
	$REM_{stabilityIndex}$ $[h^{-1}]$	46	1154.31±107.01	1110.19±159.59	980.44±166.15	**<0.001**, A>C, B>C
	$REM_{fragIndex}$ $[h^{-1}]$	37	276.99±103.74	317.44±149.37	438.15±160.76	**<0.001**, A<C, B<C
	$REMCert$ $(th = 0.50)$ [-]	46	0.94±0.03	0.92±0.06	0.89±0.05	**<0.001**, A>C, B>C
	REM_{amount} $(th = 0.50)$ [-]	49	0.19±0.07	0.17±0.09	0.10±0.05	**<0.001**, A>C, B>C
	$p(NREM \rightarrow W)$ [-]	49	0.23±0.16	0.34±0.27	0.46±0.22	**<0.001**, A<C
	$p(REM \rightarrow W)$ [-]	43	0.24±0.13	0.33±0.16	0.37±0.16	**<0.001**, A<B, A<C
	$H0_{REM}$ [-]	41	0.51±0.12	0.57±0.12	0.6±0.11	**<0.001**, A<B, A<C
EEG spectral features	Relative power in δ_1 (REM sleep) [-]	36	0.27±0.03	0.27±0.03	0.30±0.04	**<0.01**, A<C, B<C
	Maximum peak in δ band (REM sleep) [-]	36	0.09±0.01	0.09±0.01	0.10±0.02	**<0.01**, A<C, B<C
EOG energy features	Standard deviation of the relative energy at d_4 in REM sleep [-]	32	0.04±0.01	0.04±0.01	0.03±0.01	**<0.001**, A>C, B>C
EEG complexity features	Multi-scale entropy in REM sleep (scale=8) [-]	27	2.13±0.05	2.14±0.05	2.10±0.08	**<0.01**, A>C, B>C

Table 6.8: Distribution of the features selected in at least half of the 50 optimal models. The distributions are shown as mean and standard deviation. Kruskal-Wallis tests were performed to make multi-group comparisons and tailed Mann-Whitney tests with post-hoc Bonferroni correction were used to analyze which pairs of groups were significantly different.

have EEG, EOG and micro-sleep abnormalities typical of definite RBD. From this, it is possible to speculate that micro-sleep instability might be a biomarker of progression from prodromal RBD to definite RBD in PD patients.

Identification of RBD in PD patients

Despite no features describing RSWA were included in this study, RBD could be successfully identified in PD patients by using only two EEG and two EOG signals. The obtained performances are comparable to the results presented when only EMG signals were used (Section 5.4.2). This suggests that EEG and EOG signals could be used as complementary electrophysiological signals to EMG for identifying RBD in PD patients.

The final proposed system consists of an ensemble of RF models. This strategy has been used to increase the generalization of the system and because ensemble methods generally have increased performances compared to single classifiers [143].

Features describing micro-sleep structure had the highest impact to identify RBD in PD patients and, among these, the wake-sleep transition index was found to have high importance for correct discrimination of PD patients with and without RBD (Table 6.8). The increased wake-sleep transitions might reflect the sleep fragmentation seen in neurodegenerative diseases [188], which may be more severe in presence of RBD. PD patients with RBD report slightly higher amounts of subjective sleep problems compared to those without RBD [189]; future studies could investigate whether increased micro-sleep transitions in PD+RBD patients correlates with the subjective sleep complaints. Interestingly, PD+RBD and PDnonRBD groups did not differ on manually quantified macro-sleep scale measures such as sleep efficiency and awakening index (Table 6.4). Concerning micro-sleep structure, PD+RBD patients also showed altered REM micro-sleep structure (i.e. increased REM sleep fragmentation and lower REM sleep stability, Table 6.8). These micro-sleep alterations might be caused by neurodegeneration in the brainstem and in the thalamus [9, 158, 190], which are structures having a critical role for wake and sleep regulation (Section 2.2.2).

The micro-sleep instability seen in PD+RBD participants might also be caused by the automated sleep staging method, which could classify some of the REM sleep segments containing abnormal muscular activity as wakefulness. However, this should not be considered as a flaw, because the algorithm simply classifies each 5-s mini-epoch by identifying which of the typical patterns of W, REM or NREM sleep is dominant (Section 6.3.1). As there are no definite rules for scoring REM sleep in RBD patients [191], the proposed data-driven approach can be also considered as a first attempt to define them objectively.

The results also revealed more pronounced EEG slowing in REM sleep in PD+RBD patients compared to PDnonRBD patients (Table 6.8). A previous study found more pronounced EEG slowing in wakefulness but not in REM sleep in PD+RBD compared to PDnonRBD patients [159]. The different finding here reported might be caused by the difference in sleep staging (automatic versus manual). However, the present results can be considered a confirmation of the EEG slowing seen in PD+RBD patients compared to PDnonRBD patients.

A previous study found lower values of multi-scale entropy in PDnonRBD patients compared to healthy controls during REM sleep [187]. The present study shows that PD+RBD patients have lower values of multi-scale entropy compared to PDnonRBD patients (Table 6.8). This indicates that PD+RBD patients exhibit a further decrease of complexity in the EEG structure during REM sleep.

Alterations in EOG have previously been suggested as a measure to distinguish healthy controls from iRBD and PD patients [79, 81]. The present results substantiate this notion by revealing a significantly lower standard deviation of the EOG energy during REM sleep in the 8-16 Hz frequency range (i.e. 4^{th} level of decomposition in the wavelet transform) in PD+RBD patients compared to PDnonRBD (Table 6.8). This feature suggests a reduction of the variability in the time-frequency structure of EOG signals in presence of RBD.

Overall, the micro-sleep, EEG, and EOG alterations which are relevant to distinguish PDnonRBD from PD+RBD can be considered as further evidence that PD+RBD patients might have more severe neurodegeneration compared to PDnonRBD patients, in agreement with studies focusing on brain imaging [40, 157, 158] and clinical features [156].

Identification of prodromal RBD in PD patients

The results in Fig. 6.14 show that the proposed system could separate PD patients with RBEs who developed definite RBD after two years from the ones that remained without RBD at 2-year FU with accuracy and specificity over 80%. This finding supports the emerging concept of prodromal RBD (Section 2.3.4, [14, 15]) and further confirms the hypothesis of RBD as a continuum in the neurodegenerative process of alpha-synucleinopathies.

These results indicate that PD+RBE patients, who developed definite RBD after two years, already show micro-sleep, EEG and EOG characteristics typical of definite RBD at baseline study. These patients might also have more advanced neurodegeneration, which affects the sleep and REM regulation system, but not yet the neurons controlling the REM atonia to such a degree that is seen as in PD+RBD patients. This confirms the distinctive patterns of REM-on/off neurons and REM-atonia neurons [49] (Section 2.2.2).

PDnonRBD patients who developed definite RBD at 2-year FU did not show increased values of $P(RBD)$ (see $P(RBD)$ distribution for test PDnonRBD group in Fig. 6.13). It could be hypothesized that conversion to definite RBD in PDnonRBD patients occur at a faster rate than in PD+RBE patients and therefore the proposed system could not identify abnormalities in micro-sleep structure, EEG and EOG signals. This means that the fully automated system can be used as supportive tool to identify with high specificity and accuracy PD patients with RBEs who will progress to definite RBD in a short time range.

Finally, these results support the progression from prodromal RBD to definite RBD in overt PD patients. When data will be available, the developed system could be applied to subjects with RBEs without PD or even to subjects with isolated RSWA [91, 92]. Due to the high specificity, this system has the potential to be used to better predict the time of conversion to any Parkinsonian syndrome and to better identify neurodegeneration in its very early stages.

Limitations

The main limitation of this study is the small cohort used and the small number of PD+RBE who converted to definite RBD at 2-year follow-up. However, this is an emerging field of research and more data will be available in the future. Moreover, only central EEG derivations were used as they were available for all participants; future studies should investigate whether other EEG derivations could be more powerful for the same purposes presented in this study. In addition, the combination of muscular activity features with micro-sleep, EEG and EOG features was not investigated. This combination may be useful for more objective diagnostic purposes.

6.5 Conclusive remarks

The results presented in Paper V [154] and Paper VI [155] answer the research questions and address the research objectives outlined at the beginning of this Chapter. Below, the research questions are reported again, together with the research outputs achieved in this Chapter:

- ***Chapter research question 6.1:*** Which electrophysiological changes, besides RSWA, characterize RBD in PD? Can a fully automated method use them to correctly identify RBD in PD patients?
  Chapter research output 6.1: A fully-automated data-driven method, based on two EEG and two EOG channels, has been developed and it could identify with over 80% accuracy, sensitivity and specificity RBD in PD patients. The results revealed that micro-sleep alterations had a major role in correctly identifying RBD in PD patients. In addition, it has been found that PD patients with RBD have more pronounced EEG slowing, reduced EEG complexity and EOG alterations in REM sleep compared to PD patients without RBD. Such results support the idea that PD patients with RBD have more severe neurodegeneration compared to PD patients without RBD.

- ***Chapter research question 6.2:*** Do PD patients with prodromal RBD have the same electrophysiological changes as PD patients with RBD? Can PD patients with prodromal RBD be identified with the same automated methods developed for identification of PD patients with RBD?
  Chapter research output 6.2: When the developed system was applied to PD patients with RBEs, it could distinguish with accuracy and specificity over 80% the ones that developed definite RBD after two years from the ones that did not. These results indicate that PD+RBE patients, who developed definite RBD after two years, already show micro-sleep, EEG and EOG characteristics typical of definite RBD at baseline study. In this way, the developed system could be used as a supportive tool for identification of progression from RBEs, as prodromal RBD, to definite RBD in PD patients.

Chapter 7

Conclusions

The main focus of this thesis was to propose new automatic algorithms for faster and objective identification and characterization of subjects suffering from RBD or prodromal RBD, as they represent early stages of neurodegeneration. The application of these algorithms has the potential to identify a homogeneous patient group, which can be the future target for neuroprotective trials aiming at slowing down or even stopping the ongoing neurodegeneration. This thesis addresses three research questions, which are reported and answered below based on the findings presented in the previous Chapters.

Thesis research question 1: Which one of the available automated methods is the optimal one to identify patients with RBD?
Thesis research output 1: Chapter 3 presented a systematic comparison of five previously developed automated methods for RBD identification [18, 19, 20, 21, 23, 24]. The different methods were compared based on their capability of identifying RBD in a cohort including healthy controls, PLMD patients, iRBD patients, PD patients without RBD, and PD patients with RBD. The comparison showed that none of the automated methods can be elected as the optimal one to identify patients with RBD. Three of the analyzed methods (the REM atonia index - RAI [18, 19], Frandsen index - FRI [23] and Kempfner index - KEI [24]) were generally the most accurate in identifying patients with RBD, with RAI being the one with the highest sensitivity. However, these automated methods presented varying performances when different participant groups were considered. This suggests that a more robust and generalized algorithm should be implemented.

Thesis research question 2: Can a new automated method, based on machine learning techniques applied to EMG signals, overcome the currently available ones to identify RBD? Is such a method robust to inter-clinical variability?
Thesis research output 2: Chapters 4 and 5 presented the development and validation of a new data-driven method, that, based on machine learning techniques, can recognize muscular activity and identify RBD patients with accuracy and specificity over 80% (which are higher compared to RAI, FRI and KEI), and with sensitivity around 70% (which is comparable to RAI, FRI and KEI). The new proposed data-driven method is superior to the other methods based on traditional programming techniques. This might be because data patterns hidden to the human eye are automatically learned during the training of a machine learning model. The validation of the new data-driven method also revealed that muscular activity occurring in NREM sleep contributed for a better identification of patients with RBD and that the removal of muscular activity related to apneas

and arousals is not required for identifying patients with RBD. Finally, the new method has been applied to data from another clinic and it was found that it identified patients with RBD with similar performances as it did in the original cohort. This proves the robustness of the proposed method.

Thesis research question 3: Can an automated machine learning-based method, built on EEG ad EOG signals, identify and characterize RBD and prodromal RBD in PD patients?

Thesis research output 3: Chapter 6 presented the development and validation of a fully automated method that, based only on two EEG and two EOG channels, can identify RBD in PD patients with accuracy, sensitivity and specificity over 80%. The results show that, in PD patients, RBD is not just related to abnormal muscular activity in REM sleep, but also to increased micro-sleep transitions, more pronounced EEG slowing in REM sleep, reduced EEG complexity in REM sleep, and EOG alterations in REM sleep. These findings suggest that PD patients with RBD might have more severe neurodegeneration than PD patients without RBD. Based on these alterations, the proposed fully automated method recognized PD patients with RBEs who developed definite RBD after two years (i.e. the PD patients with prodromal RBD) with accuracy and specificity over 80%. This suggests that PD patients with prodromal RBD already have micro-sleep, EEG and EOG abnormalities typical of definite RBD two years before diagnosis of RBD. Therefore, the proposed algorithm can be used as a supportive tool to identify prodromal RBD in PD patients. These results are promising with the perspective of identifying biomarkers of prodromal RBD in subjects without overt PD.

Overall, the research presented in this thesis has revealed that data-driven methods based on machine learning techniques can provide good, consistent, fast and objective identification of RBD and prodromal RBD. In particular, the application of these algorithms requires an amount of time significantly lower than the one needed for manual analysis. Moreover, the intrinsic objectiveness of automatic algorithms allows identification of homogeneous patient groups, which can be the target for future neuroprotective trials. At this stage, the developed methods could be used as clinical supportive tools, with the aim of helping clinicians in making faster decisions and diagnoses. After validation of the described algorithms in bigger cohorts, it is proposed that it will be possible to use them as full diagnostic tools, thus significantly reducing the burden of manual analysis of sleep data. Furthermore, the research here presented also revealed important insights concerning characterization of early neurodegeneration, including the abnormal muscular activity in RBD patients in NREM sleep and the increased micro-sleep instability as a biomarker of RBD and prodromal RBD in *de novo* PD patients. In conclusion, the research presented here proves the usefulness of applying data-driven methods in sleep research and contributes in the path of overcoming traditional subjective and time-consuming analysis methods.

Chapter 8

Future work

Objective and fast identification of RBD and prodromal RBD has a great potential for early identification of neurodegeneration. This thesis has proposed new methods which have the potential to help researchers and clinicians with this purpose. However, this research field needs further investigations which include the following aims:

- To integrate the developed method for RBD detection based on EMG signals with a sleep staging algorithm, to obtain a fully automated system.
- To repeat a similar analysis as the one proposed in Chapter 6 in subjects without PD, to evaluate whether it is possible to identify prodromal RBD prior to development of iRBD and α-synucleinopathies.
- To combine EEG, EOG and EMG for better and more precise identification and characterization of patients with RBD.
- To investigate whether EMG, EEG and EOG patterns identified with an automated data-driven approach can predict time of conversion from prodromal RBD or iRBD to an overt α-synucleinopathy.
- To further validate the proposed methods in data recorded in several clinics, or to re-train the data-driven methods in a bigger multi-center cohort.
- To implement the proposed algorithms in clinical software.
- To test the proposed algorithms on a cohort representing a general as well as a clinical population, to better evaluate the sensitivity and specificity of the proposed methods.
- To investigate whether recent deep learning techniques, not requiring any explicit feature extraction, could lead to even better performances in the identification of patients with RBD and prodromal RBD.
- To investigate new modalities such as video or actigraphy, which are less expensive and invasive than PSG, for identification of RBD and prodromal RBD.

Bibliography

[1] M. G. Erkkinen, M. O. Kim, and M. D. Geschwind, "Clinical neurology and epidemiology of the major neurodegenerative diseases," *Cold Spring Harbor Perspectives in Biology*, vol. 10, no. 4, p. a033118, 2018.

[2] P. Rizek, N. Kumar, and M. S. Jog, "An update on the diagnosis and treatment of Parkinson's disease," *Cmaj*, vol. 188, no. 16, pp. 1157–1165, 2016.

[3] 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," *Lancet Neurology*, vol. 18, no. 5, pp. 459–480, 2019.

[4] W. R. Gibb and A. J. Lees, "The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease," *Journal of Neurology, Neurosurgery, and Psychiatry*, vol. 51, no. 6, pp. 745–752, 1988.

[5] R. B. Postuma, D. Berg, M. Stern, W. Poewe, C. W. Olanow, W. Oertel, J. Obeso, K. Marek, I. Litvan, A. E. Lang, G. Halliday, C. G. Goetz, T. Gasser, B. Dubois, P. Chan, B. R. Bloem, C. H. Adler, and G. Deuschl, "MDS clinical diagnostic criteria for Parkinson's disease," *Movement Disorders*, vol. 30, no. 12, pp. 1591–1601, 2015.

[6] A. Galvan and T. Wichmann, "Pathophysiology of Parkinsonism," *Clinical Neurophysiology*, vol. 119, no. 7, pp. 1459–1474, 2008.

[7] H. Braak, K. Del Tredici, U. Rüb, R. A. I. De Vos, E. N. H. Jansen Steur, and E. Braak, "Staging of brain pathology related to sporadic Parkinson's disease," *Neurobiology of Aging*, vol. 24, no. 2, pp. 197–211, 2003.

[8] H. Braak, E. Ghebremedhin, U. Rüb, H. Bratzke, and K. Del Tredici, "Stages in the development of Parkinson's disease-related pathology," *Cell and Tissue Research*, vol. 318, no. 1, pp. 121–134, 2004.

[9] B. F. Boeve, "Idiopathic REM sleep behaviour disorder in the development of Parkinson's disease," *Lancet Neurology*, vol. 12, no. 5, pp. 469–482, 2013.

[10] C. H. Schenck, S. R. Bundlie, M. G. Ettinger, and M. W. Mahowald, "Chronic behavioral disorders of human REM sleep: a new category of parasomnia," *Sleep*, vol. 9, no. 2, pp. 293–308, 1986.

[11] C. H. Schenck and M. W. Mahowald, "REM sleep behavior disorder: Clinical, developmental, and neuroscience perspectives 16 years after its formal identification in SLEEP," *Sleep*, vol. 25, no. 2, pp. 120–138, 2002.

[12] R. B. Postuma, A. Iranzo, M. Hu, B. Hoegl, B. F. Boeve, R. Manni, W. H. Oertel, I. Arnulf, L. Ferini-Strambi, M. Puligheddu, E. Antelmi, V. C. De Cock, D. Arnaldi, B. Mollenhauer, A. Videnovic, K. Sonka, K.-Y. Jung, D. Kunz, Y. Dauvilliers, F. Provini, S. J. Lewis, J. Buskova, M. Pavlova, A. Heidbreder, J. Y. Montplaisir, J. Santamaria, T. R. Barber, A. Stefani, E. K. St Louis, M. Terzaghi, A. Janzen, S. Leu-Semenescu, G. Plazzi, F. Nobili, F. Sixel-Doering, P. Dusek, F. Bes, P. Cortelli, K. E. Martens, J.-F. Gagnon, C. Gaig, M. Zucconi, C. Trenkwalder, Z. Gan-Or, C. Lo, M. Rolinski, P. Mahlknecht, E. Holzknecht, A. R. Boeve, L. N. Teigen, G. Toscano, G. Mayer, S. Morbelli, B. Dawson, and A. Pelletier, "Risk and predictors of dementia and parkinsonism in idiopathic REM sleep behaviour disorder: a multicentre study," *Brain*, vol. 142, no. 3, pp. 744–759, 2019.

[13] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, C. L. Marcus, B. V. Vaughn, *et al.*, *The AASM manual for the scoring of sleep and associated events, Rules, Terminology and Technical Specifications*, vol. 1. Darien, Illinois, American Academy of Sleep Medicine, 2012.

[14] F. Sixel-Döring, J. Zimmermann, A. Wegener, B. Mollenhauer, and C. Trenkwalder, "The evolution of REM sleep behavior disorder in early Parkinson's disease," *Sleep*, vol. 39, no. 9, pp. 1737–1742, 2016.

[15] B. Högl, A. Stefani, and A. Videnovic, "Idiopathic REM sleep behaviour disorder and neurodegeneration - An update," *Nature Reviews Neurology*, vol. 14, no. 1, pp. 40–56, 2018.

[16] R. B. Postuma, "Biomarkers of Neurodegenerative Disease in Idiopathic RBD," *Rapid-eye-movement Sleep Behavior Disorder*, pp. 527–540, 2018.

[17] R. V. Kogan, S. K. Meles, K. L. Leenders, K. Reetz, and W. H. Oertel, "Brain Imaging in RBD," *Rapid-eye-movement Sleep Behavior Disorder*, pp. 403–445, 2018.

[18] R. Ferri, M. Manconi, G. Plazzi, O. Bruni, S. Vandi, P. Montagna, L. Ferini-Strambi, and M. Zucconi, "A quantitative statistical analysis of the submentalis muscle EMG amplitude during sleep in normal controls and patients with REM sleep behavior disorder," *Journal of Sleep Research*, vol. 17, no. 1, pp. 89–100, 2008.

[19] R. Ferri, F. Rundo, M. Manconi, G. Plazzi, O. Bruni, A. Oldani, L. Ferini-Strambi, and M. Zucconi, "Improved computation of the atonia index in normal controls and patients with REM sleep behavior disorder," *Sleep Medicine*, vol. 11, no. 9, pp. 947–949, 2010.

[20] J. W. Burns, F. B. Consens, R. J. Little, K. J. Angell, S. Gilman, and R. D. Chervin, "EMG variance during polysomnography as an assessment for REM sleep behavior disorder," *Sleep*, vol. 30, no. 12, pp. 1771–1778, 2007.

[21] G. Mayer, K. Kesper, T. Ploch, S. Canisius, T. Penzel, W. Oertel, and K. Stiasny-Kolster, "Quantification of tonic and phasic muscle activity in REM sleep behavior disorder," *Journal of Clinical Neurophysiology*, vol. 25, no. 1, pp. 48–55, 2008.

[22] D. Guttowski, G. Mayer, W. H. Oertel, K. Kesper, and T. Rosenberg, "Validation of semiautomatic scoring of REM sleep without atonia in patients with RBD," *Sleep Medicine*, vol. 46, pp. 107–113, 2018.

[23] R. Frandsen, M. Nikolic, M. Zoetmulder, L. Kempfner, and P. Jennum, "Analysis of automated quantification of motor activity in REM sleep behaviour disorder," *Journal of Sleep Research*, vol. 24, no. 5, pp. 583–590, 2015.

[24] J. Kempfner, G. L. Sorensen, M. Nikolic, R. Frandsen, H. B. D. Sorensen, and P. Jennum, "Rapid eye movement sleep behavior disorder as an outlier detection problem," *Journal of Clinical Neurophysiology*, vol. 31, no. 1, pp. 86–93, 2014.

[25] J. Kempfner, H. B. Sorensen, M. Nikolic, and P. Jennum, "Early automatic detection of Parkinson's disease based on sleep recordings," *Journal of Clinical Neurophysiology*, vol. 31, no. 5, pp. 409–415, 2014.

[26] B. Frauscher, D. Gabelia, M. Biermayr, A. Stefani, H. Hackner, T. Mitterling, W. Poewe, and B. Högl, "Validation of an integrated software for the detection of rapid eye movement sleep behavior disorder," *Sleep*, vol. 37, no. 10, pp. 1663–1671, 2014.

[27] R. Boutaba, M. A. Salahuddin, N. Limam, S. Ayoubi, N. Shahriar, F. Estrada-Solano, and O. M. Caicedo, "A comprehensive survey on machine learning for networking: evolution, applications and research opportunities," *Journal of Internet Services and Applications*, vol. 9, no. 1, p. 16, 2018.

[28] D. A. Evans, H. H. Funkenstein, M. S. Albert, P. A. Scherr, N. R. Cook, M. J. Chown, L. E. Hebert, C. H. Hennekens, and J. O. Taylor, "Prevalence of Alzheimer's disease in a community population of older persons. Higher than previously reported," *Jama*, vol. 262, no. 18, pp. 2551–6, 1989.

[29] M. J. Martí, E. Tolosa, and J. Campdelacreu, "Clinical overview of the synucleinopathies," *Movement Disorders*, vol. 18, no. 6, pp. S21–S27, S21–S27, 2003.

[30] R. Savica, B. F. Boeve, and G. Logroscino, "Epidemiology of alpha-synucleinopathies: from Parkinson's disease to dementia with Lewy bodies," *Handbook of Clinical Neurology*, vol. 138, pp. 153–158, 2016.

[31] H. Douna, B. M. Bavelaar, H. Pellikaan, B. Olivier, and T. Pieters, "Neuroprotection in Parkinson's disease: A systematic review of the preclinical data," *Open Pharmacology Journal*, vol. 6, no. 1, pp. 12–26, 2012.

[32] R. G. Hart, L. A. Pearce, B. M. Ravina, T. C. Yaltho, and J. R. Marler, "Neuroprotection trials in Parkinson's disease: Systematic review," *Movement Disorders*, vol. 24, no. 5, pp. 647–654, 2009.

[33] A. C. Vernon, C. Ballard, and M. Modo, "Neuroimaging for Lewy body disease: Is the in vivo molecular imaging of alpha-synuclein neuropathology required and feasible?," *Brain Research Reviews*, vol. 65, no. 1, pp. 28–55, 2010.

[34] C. W. Olanow, M. B. Stern, and K. Sethi, "The scientific and clinical basis for the treatment of Parkinson's disease (2009)," *Neurology*, vol. 72, no. 21, Suppl. 4, pp. S1–S136, 2009.

[35] H. Braak, U. Rüb, W. P. Gai, and K. Del Tredici, "Idiopathic Parkinson's disease: Possible routes by which vulnerable neuronal types may be subject to neuroinvasion by an unknown pathogen," *Journal of Neural Transmission*, vol. 110, no. 5, pp. 517–536, 2003.

[36] K. A. Jellinger, "Is Braak staging valid for all types of Parkinson's disease?," *Journal of Neural Transmission*, vol. 126, no. 4, pp. 423–431, 2019.

[37] C. Marras and K. R. Chaudhuri, "Nonmotor features of Parkinson's disease subtypes," *Movement Disorders*, vol. 31, no. 8, pp. 1095–1102, 2016.

[38] J. Jankovic, M. McDermott, J. Carter, S. Gauthier, C. Goetz, L. Golbe, S. Huber, W. Koller, C. Olanow, and I. Shoulson, "Variable expression of Parkinson's disease: a base-line analysis of the DATATOP cohort. The Parkinson Study Group," *Neurology*, vol. 40, no. 10, pp. 1529–34, 1990.

[39] M. Selikhova, D. R. Williams, P. A. Kempster, J. L. Holton, T. Revesz, and A. J. Lees, "A clinico-pathological study of subtypes in Parkinson's disease," *Brain*, vol. 132, no. 11, pp. 2947–2957, 2009.

[40] S. Rahayel, M. Gaubert, R. B. Postuma, J. Montplaisir, J. Carrier, O. Monchi, D. Rémillard-Pelchat, P.-A. Bourgouin, M. Panisset, S. Chouinard, S. Joubert, and J.-F. Gagnon, "Brain atrophy in Parkinson's disease with polysomnography-confirmed REM sleep behavior disorder," *Sleep*, vol. 42, no. 6, p. zsz062, 2019.

[41] R. B. Postuma, A. Pelletier, D. Berg, J. F. Gagnon, F. Escudier, and J. Montplaisir, "Screening for prodromal Parkinson's disease in the general community: A sleep-based approach," *Sleep Medicine*, vol. 21, pp. 101–105, 2016.

[42] Y. Dauvilliers, C. H. Schenck, R. B. Postuma, A. Iranzo, P.-H. Luppi, G. Plazzi, J. Montplaisir, and B. Boeve, "REM sleep behaviour disorder," *Nature Reviews Disease Primers*, vol. 4, no. 1, p. 19, 2018.

[43] A. Iranzo, J. Santamaria, and E. Tolosa, "Idiopathic rapid eye movement sleep behaviour disorder: Diagnosis, management, and the need for neuroprotective interventions," *Lancet Neurology*, vol. 15, no. 4, pp. 405–419, 2016.

[44] A. Galbiati, L. Verga, E. Giora, M. Zucconi, and L. Ferini-Strambi, "The risk of neurodegeneration in REM sleep behavior disorder: A systematic review and meta-analysis of longitudinal studies," *Sleep Medicine Reviews*, vol. 43, pp. 37–46, 2019.

[45] A. Iranzo, E. Tolosa, E. Gelpi, J. L. Molinuevo, F. Valldeoriola, M. Serradell, R. Sanchez-Valle, I. Vilaseca, F. Lomeña, D. Vilas, A. Lladó, C. Gaig, and J. Santamaria, "Neurodegenerative disease status and post-mortem pathology in idiopathic rapid-eye-movement sleep behaviour disorder: An observational cohort study," *Lancet Neurology*, vol. 12, no. 5, pp. 443–453, 2013.

[46] L. Schneider, "Anatomy and physiology of normal sleep," *Sleep and Neurologic Disease*, pp. 1–28, 2017.

[47] H. R. Colten and B. M. Altevogt, *Sleep disorders and sleep deprivation: An unmet public health problem*. National Academies Press, 2006.

[48] C. B. Saper, T. E. Scammell, and J. Lu, "Hypothalamic regulation of sleep and circadian rhythms," *Nature*, vol. 437, no. 7063, pp. 1257–1263, 2005.

[49] C. B. Saper, P. M. Fuller, N. P. Pedersen, J. Lu, and T. E. Scammell, "Sleep state switching," *Neuron*, vol. 68, no. 6, pp. 1023–1042, 2010.

[50] J. R. Schwartz and T. Roth, "Neurophysiology of sleep and wakefulness: Basic science and clinical implications," *Current Neuropharmacology*, vol. 6, no. 4, pp. 367–378, 2008.

[51] M. M. Ohayon, M. A. Carskadon, C. Guilleminault, and M. V. Vitiello, "Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: Developing normative sleep values across the human lifespan," *Sleep*, vol. 27, no. 7, pp. 1255–1273, 2004.

[52] S. Redline, H. L. Kirchner, S. F. Quan, D. J. Gottlieb, V. Kapur, and A. Newman, "The effects of age, sex, ethnicity, and sleep-disordered breathing on sleep architecture," *Archives of Internal Medicine*, vol. 164, no. 4, pp. 406–418, 2004.

[53] B. A. Mander, J. R. Winer, and M. P. Walker, "Sleep and human aging," *Neuron*, vol. 94, no. 1, pp. 19–36, 2017.

[54] J. Carrier, I. Viens, G. Poirier, R. Robillard, M. Lafortune, G. Vandewalle, N. Martin, M. Barakat, J. Paquet, and D. Filipini, "Sleep slow wave changes during the middle years of life," *European Journal of Neuroscience*, vol. 33, no. 4, pp. 758–766, 2011.

[55] J. Haba-Rubio, B. Frauscher, P. Marques-Vidal, J. Toriel, N. Tobback, D. Andries, M. Preisig, P. Vollenweider, R. Postuma, and R. Heinzer, "Prevalence and determinants of rapid eye movement sleep behavior disorder in the general population," *Sleep*, vol. 41, no. 2, p. zsx197, 2018.

[56] F. Sixel-Döring, E. Trautmann, B. Mollenhauer, and C. Trenkwalder, "Rapid eye movement sleep behavioral events: A new marker for neurodegeneration in early Parkinson's disease?," *Sleep*, vol. 37, no. 3, pp. 431–438, 2014.

[57] J. Gagnon, M. Bedard, M. Fantini, D. Petit, M. Panisset, S. Rompre, J. Carrier, and J. Montplaisir, "REM sleep behavior disorder and REM sleep without atonia in Parkinson's disease," *Neurology*, vol. 59, no. 4, pp. 585–589, 2002.

[58] B. F. Boeve, M. H. Silber, T. J. Ferman, J. A. Lucas, and J. E. Parisi, "Association of REM sleep behavior disorder and neurodegenerative disease may reflect an underlying synucleinopathy," *Movement Disorders*, vol. 16, no. 4, pp. 622–630, 2001.

[59] R. Vetrugno, F. Provini, P. Cortelli, G. Plazzi, E. M. Lotti, G. Pierangeli, C. Canali, and P. Montagna, "Sleep disorders in multiple system atrophy: A correlative video-polysomnographic study," *Sleep Medicine*, vol. 5, no. 1, pp. 21–30, 2004.

[60] A. A. of Sleep Medicine, *International Classification of Sleep Disorders, Third Edition*, vol. 1. Darien, Illinois, American Academy of Sleep Medicine, 2014.

[61] B. Högl and A. Stefani, "REM sleep behavior disorder (RBD): Update on diagnosis and treatment," *Somnologie*, vol. 21, no. 1, pp. 1–8, 2017.

[62] B. F. Boeve, M. H. Silber, C. B. Saper, T. J. Ferman, D. W. Dickson, J. E. Parisi, E. E. Benarroch, J. E. Ahlskog, G. E. Smith, R. C. Caselli, M. Tippman-Peikert, E. J. Olson, S. C. Lin, T. Young, Z. Wszolek, C. H. Schenck, M. W. Mahowald, P. R. Castillo, K. Del Tredici, and H. Braak, "Pathophysiology of REM sleep behaviour disorder and relevance to neurodegenerative disease," *Brain*, vol. 130, no. 11, pp. 2770–2788, 2007.

[63] C. L. Rodriguez, N. Jaimchariyatam, and K. Budur, "Rapid eye movement sleep behavior disorder: A review of the literature and update on current concepts," *Chest*, vol. 152, no. 3, pp. 650–662, 2017.

[64] M. Cesari and P. Jennum, "Selective polysomnographic findings in REM sleep behavior disorder (RBD) and Parkinson's disease," *Rapid-eye-movement Sleep Behavior Disorder*, pp. 271–279, 2018.

[65] P. Jennum, J. A. Christensen, and M. Zoetmulder, "Neurophysiological basis of rapid eye movement sleep behavior disorder: Informing future drug development," *Nature and Science of Sleep*, vol. 8, pp. 107–120, 2016.

[66] M. L. Fantini, J. F. Gagnon, D. Petit, S. Rompré, A. Décary, J. Carrier, and J. Montplaisir, "Slowing of electroencephalogram in rapid eye movement sleep behavior disorder," *Annals of Neurology*, vol. 53, no. 6, pp. 774–780, 2003.

[67] A. Iranzo, V. Isetta, J. L. Molinuevo, M. Serradell, D. Navajas, R. Farre, and J. Santamaria, "Electroencephalographic slowing heralds mild cognitive impairment in idiopathic REM sleep behavior disorder," *Sleep Medicine*, vol. 11, no. 6, pp. 534–539, 2010.

[68] T. Sasai, M. Matsuura, and Y. Inoue, "Electroencephalographic findings related with mild cognitive impairment in idiopathic rapid eye movement sleep behavior disorder," *Sleep*, vol. 36, no. 12, pp. 1893–1899, 2013.

[69] J. Rodrigues Brazéte, J. F. Gagnon, R. B. Postuma, J. A. Bertrand, D. Petit, and J. Montplaisir, "Electroencephalogram slowing predicts neurodegeneration in rapid eye movement sleep behavior disorder," *Neurobiology of Aging*, vol. 37, pp. 74–81, 2016.

[70] G. Ruffini, D. Ibañez, M. Castellano, L. Dubreuil-Vall, A. Soria-Frisch, R. Postuma, J. F. Gagnon, and J. Montplaisir, "Deep learning with EEG spectrograms in rapid eye movement behavior disorder," *Frontiers in Neurology*, vol. 10, p. 806, 2019.

[71] R. Ferri, F. Rundo, A. Silvani, M. Zucconi, O. Bruni, L. Ferini-Strambi, G. Plazzi, and M. Manconi, "REM sleep EEG instability in REM sleep behavior disorder and clonazepam effects," *Sleep*, vol. 40, no. 8, 2017.

[72] I. H. Hansen, M. Marcussen, J. A. E. Christensen, P. Jennum, and H. B. D. Sorensen, "Detection of a sleep disorder predicting Parkinson's disease," *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, vol. 2013, pp. 5793–5796, 2013.

[73] V. Latreille, J. Carrier, J. Montplaisir, M. Lafortune, and J.-F. Gagnon, "Non-rapid eye movement sleep characteristics in idiopathic REM sleep behavior disorder," *Journal of the Neurological Sciences*, vol. 310, no. 1-2, pp. 159–162, 2011.

[74] J. A. E. Christensen, J. Kempfner, M. Zoetmulder, H. L. Leonthin, L. J. Arvastson, S. R. Christensen, H. B. D. Sorensen, and P. Jennum, "Decreased sleep spindle density in patients with idiopathic REM sleep behavior disorder and patients with Parkinson's disease," *Clinical Neurophysiology*, vol. 125, no. 3, pp. 512–519, 2014.

[75] C. O'Reilly, I. Godin, J. Montplaisir, and T. Nielsen, "REM sleep behaviour disorder is associated with lower fast and higher slow sleep spindle densities," *Journal of Sleep Research*, vol. 24, no. 6, pp. 593–601, 2015.

[76] A. Melpignano, L. Parrino, J. Santamaria, C. Gaig, I. Trippi, M. Serradell, C. Mutti, M. Riccó, and A. Iranzo, "Isolated rapid eye movement sleep behavior disorder and cyclic alternating pattern: is sleep microstructure a predictive parameter of neurodegeneration?," *Sleep*, 2019.

[77] V. Latreille, J. Carrier, B. Gaudet-Fex, J. Rodrigues-Brazéte, M. Panisset, S. Chouinard, R. B. Postuma, and J. F. Gagnon, "Electroencephalographic prodromal markers of dementia across conscious states in Parkinson's disease," *Brain*, vol. 139, no. 4, pp. 1189–1199, 2016.

[78] J. A. E. Christensen, M. Nikolic, S. C. Warby, H. Koch, M. Zoetmulder, R. Frandsen, K. K. Moghadam, H. B. D. Sorensen, E. Mignot, and P. J. Jennum, "Sleep spindle alterations in patients with Parkinson's disease," *Frontiers in Human Neuroscience*, vol. 9, p. 233, 2015.

[79] J. A. Christensen, R. Frandsen, J. Kempfner, L. Arvastson, S. R. Christensen, P. Jennum, and H. B. Sorensen, "Separation of Parkinson's patients in early and mature stages from control subjects using one EOG channel," *34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, vol. 2012, pp. 2941–2944, 2012.

[80] J. A. E. Christensen, M. Cesari, F. Pizza, G. Plazzi, and P. Jennum, "Nocturnal eye movements in patients with idiopathic rapid eye movement sleep behaviour disorder and patients with Parkinson's disease," *Journal of Sleep Research*, vol. 27, no. Suppl. 1, p. P209, 2018.

[81] J. A. E. Christensen, H. Koch, R. Frandsen, J. Kempfner, L. J. Arvastson, S. R. Christensen, H. B. D. Sorensen, and P. Jennum, "Classification of iRBD and Parkinson's disease patients based on eye movements during sleep," *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society Conference (EMBC)*, vol. 2013, pp. 441–444, 2013.

[82] R. H. S. Carpenter, "The neural control of looking," *Current Biology*, vol. 10, no. 8, pp. 291–293, 2000.

[83] G. L. Sorensen, J. Kempfner, M. Zoetmulder, H. B. D. Sorensen, and P. Jennum, "Attenuated heart rate response in REM sleep behavior disorder and Parkinson's disease," *Movement Disorders*, vol. 27, no. 7, pp. 888–894, 2012.

[84] G. L. Sorensen, J. Mehlsen, and P. Jennum, "Reduced sympathetic activity in idiopathic rapid-eye-movement sleep behavior disorder and Parkinson's disease," *Autonomic Neuroscience: Basic and Clinical*, vol. 179, no. 1-2, pp. 138–141, 2013.

[85] R. B. Postuma, P. A. Lanfranchi, H. Blais, J. F. Gagnon, and J. Y. Montplaisir, "Cardiac autonomic dysfunction in idiopathic REM sleep behavior disorder," *Movement Disorders*, vol. 25, no. 14, pp. 2304–2310, 2010.

[86] T. Miyamoto, M. Miyamoto, Y. Inoue, Y. Usui, K. Suzuki, and K. Hirata, "Reduced cardiac 123I-MIBG scintigraphy in idiopathic REM sleep behavior disorder," *Neurology*, vol. 67, no. 12, pp. 2236–2238, 2006.

[87] J. A. E. Christensen, M. Zoetmulder, H. Koch, R. Frandsen, L. J. Arvastson, S. R. Christensen, P. Jennum, and H. B. D. Sorensen, "Data-driven modeling of sleep EEG and EOG reveals characteristics indicative of pre-Parkinson's and Parkinson's disease," *Journal of Neuroscience Methods*, vol. 235, pp. 262–276, 2014.

[88] J. A. E. Christensen, P. Jennum, H. Koch, R. Frandsen, M. Zoetmulder, L. J. Arvastson, S. R. Christensen, and H. B. D. Sorensen, "Sleep stability and transitions in patients with idiopathic REM sleep behavior disorder and patients with Parkinson's disease," *Clinical Neurophysiology*, vol. 127, no. 1, pp. 537–543, 2016.

[89] H. Koch, J. A. E. Christensen, R. Frandsen, M. Zoetmulder, L. J. Arvastson, S. R. Christensen, P. Jennum, and H. B. D. Sorensen, "Automatic sleep classification using a data-driven topic model reveals latent sleep states," *Journal of Neuroscience Methods*, vol. 235, pp. 130–137, 2014.

[90] T. Sasai-Sakuma, B. Frauscher, T. Mitterling, L. Ehrmann, D. Gabelia, E. Brandauer, Y. Inoue, W. Poewe, and B. Högl, "Quantitative assessment of isolated rapid eye movement (REM) sleep without atonia without clinical REM sleep behavior disorder: Clinical and research implications," *Sleep Medicine*, vol. 15, no. 9, pp. 1009–1015, 2014.

[91] R. Ferri, D. Aricó, F. I. Cosentino, B. Lanuzza, G. Chiaro, and M. Manconi, "REM sleep without atonia with REM sleep-related motor events: Broadening the spectrum of REM sleep behavior disorder," *Sleep*, vol. 41, no. 12, 2018.

[92] A. Stefani, D. Gabelia, B. Högl, T. Mitterling, P. Mahlknecht, H. Stockner, W. Poewe, and B. Frauscher, "Long-term follow-up investigation of isolated rapid eye movement sleep without atonia without rapid eye movement sleep behavior disorder: A pilot study," *Journal of Clinical Sleep Medicine*, vol. 11, no. 11, pp. 1273–1279, 2015.

[93] D. L. Bliwise, J. Fairley, S. Hoff, R. S. Rosenberg, D. B. Rye, D. A. Schulman, and L. M. Trotti, "Inter-rater agreement for visual discrimination of phasic and tonic electromyographic activity in sleep," *Sleep*, vol. 41, no. 7, 2018.

[94] M. Cesari, J. A. E. Christensen, L. Kempfner, A. N. Olesen, G. Mayer, K. Kesper, W. H. Oertel, F. Sixel-Döring, C. Trenkwalder, H. B. D. Sorensen, and P. Jennum, "Comparison of computerized methods for rapid eye movement sleep without atonia detection," *Sleep*, vol. 41, no. 10, 2018.

[95] S. J. McCarter, E. K. St. Louis, D. J. Sandness, E. J. Duwell, P. C. Timm, B. F. Boeve, and M. H. Silber, "Diagnostic REM sleep muscle activity thresholds in patients with idiopathic REM sleep behavior disorder with and without obstructive sleep apnea," *Sleep Medicine*, vol. 33, pp. 23–29, 2017.

[96] M. Figorilli, R. Ferri, M. Zibetti, P. Beudin, M. Puligheddu, L. Lopiano, A. Cicolin, F. Durif, A. Marques, and M. L. Fantini, "Comparison between automatic and visual scorings of REM sleep without atonia for the diagnosis of REM sleep behavior disorder in Parkinson's disease," *Sleep*, vol. 40, no. 2, p. zsw060, 2017.

[97] R. Ferri, J. F. Gagnon, R. B. Postuma, F. Rundo, and J. Y. Montplaisir, "Comparison between an automatic and a visual scoring method of the chin muscle tone during rapid eye movement sleep," *Sleep Medicine*, vol. 15, no. 6, pp. 661–665, 2014.

[98] A. N. Olesen, M. Cesari, J. A. E. Christensen, H. B. D. Sorensen, E. Mignot, and P. Jennum, "A comparative study of methods for automatic detection of rapid eye movement abnormal muscular activity in narcolepsy," *Sleep Medicine*, vol. 44, pp. 97–105, 2018.

[99] J. Montplaisir, J. F. Gagnon, M. L. Fantini, R. B. Postuma, Y. Dauvilliers, A. Desautels, S. Rompré, and J. Paquet, "Polysomnographic diagnosis of idiopathic REM sleep behavior disorder," *Movement Disorders*, vol. 25, no. 13, pp. 2044–2051, 2010.

[100] B. Frauscher, A. Iranzo, C. Gaig, V. Gschliesser, M. Guaita, V. Raffelseder, L. Ehrmann, N. Sola, M. Salamero, E. Tolosa, W. Poewe, J. Santamaria, and B. Högl, "Normative EMG values during REM sleep for the diagnosis of REM sleep behavior disorder," *Sleep*, vol. 35, no. 6, pp. 835–847, 2012.

[101] S. J. McCarter, E. K. St. Louis, E. J. Duwell, P. C. Timm, D. J. Sandness, B. F. Boeve, and M. H. Silber, "Diagnostic thresholds for quantitative REM sleep phasic burst duration, phasic and tonic muscle activity, and REM atonia index in REM sleep behavior disorder with and without comorbid obstructive sleep apnea," *Sleep*, vol. 37, no. 10, pp. 1649–1662, 2014.

[102] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.

[103] T. MathWorks, "friedman." `https://se.mathworks.com/help/stats/friedman.html`, 2019. [Online; accessed 22-August-2019].

[104] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, 1979.

[105] R. Ferri, C. Franceschini, M. Zucconi, S. Vandi, F. Poli, O. Bruni, C. Cipolli, P. Montagna, and G. Plazzi, "Searching for a marker of REM sleep behavior disorder: Submentalis muscle EMG amplitude analysis during sleep in patients with narcolepsy/cataplexy," *Sleep*, vol. 31, no. 10, pp. 1409–1417, 2008.

[106] A. Iranzo and J. Santamaria, "Severe obstructive sleep apnea/hypopnea mimicking REM sleep behavior disorder," *Sleep*, vol. 28, no. 2, pp. 203–206, 2005.

[107] T. C. Wetter, G. Dirlich, J. Streit, C. Trenkwalder, A. Schuld, and T. Pollmächer, "An automatic method for scoring leg movements in polygraphic sleep recordings and its validity in comparison to visual scoring," *Sleep*, vol. 27, no. 2, pp. 324–328, 2004.

[108] R. Ferri, M. Zucconi, M. Manconi, O. Bruni, S. Miano, G. Plazzi, and L. Ferini-Strambi, "Computer-assisted detection of nocturnal leg motor activity in patients with restless legs syndrome and periodic leg movements during sleep," *Sleep*, vol. 28, no. 8, pp. 998–1004, 2005.

[109] H. Moore, E. Leary, S.-Y. Lee, O. Carrillo, R. Stubbs, P. Peppard, T. Young, B. Widrow, and E. Mignot, "Design and validation of a periodic leg movement detector," *Plos One*, vol. 9, no. 12, p. e114565, 2014.

[110] A. S. Huang, P. Skeba, M. S. Yang, F. P. Sgambati, C. J. Earley, and R. P. Allen, "MATPLM1, A MATLAB script for scoring of periodic limb movements: Preliminary validation with visual scoring," *Sleep Medicine*, vol. 16, no. 12, pp. 1541–1549, 2015.

[111] D. Alvarez-Estevez, "A new automatic method for the detection of limb movements and the analysis of their periodicity," *Biomedical Signal Processing and Control*, vol. 26, pp. 117–125, 2016.

[112] A. Stefani, A. Heidbreder, H. Hackner, and B. Högl, "Validation of a leg movements count and periodic leg movements analysis in a custom polysomnography system," *Bmc Neurology*, vol. 17, no. 1, p. 42, 2017.

[113] J. A. Fairley, G. Georgoulas, N. A. Mehta, A. G. Gray, and D. L. Bliwise, "Computer detection approaches for the identification of phasic electromyographic (EMG) activity during human sleep," *Biomedical Signal Processing and Control*, vol. 7, no. 6, pp. 606–615, 2012.

[114] J. A. Fairley, G. Georgoulas, O. L. Smart, G. Dimakopoulos, P. Karvelis, C. D. Stylios, D. B. Rye, and D. L. Bliwise, "Wavelet analysis for detection of phasic electromyographic activity in sleep: Influence of mother wavelet and dimensionality reduction," *Computers in Biology and Medicine*, vol. 48, no. 1, pp. 77–84, 2014.

[115] M. Cesari, J. A. E. Christensen, P. Jennum, and H. B. D. Sorensen, "Probabilistic data-driven method for limb movement detection during sleep," *40th International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2018, pp. 163–166, 2018.

[116] M. Akay, "Wavelets in biomedical-engineering," *Annals of Biomedical Engineering*, vol. 23, no. 5, pp. 531–542, 1995.

[117] A. Phinyomark, C. Limsakul, and P. Phukpattaranont, "Application of wavelet analysis in EMG feature extraction for pattern classification," *Measurement Science Review*, vol. 11, no. 2, pp. 45–52, 2011.

[118] S. Mallat, "Zero-crossings of a wavelet transform," *IEEE Transactions on Information Theory*, vol. 37, no. 4, pp. 1019–1033, 1991.

[119] P. Bentley and J. McDonnel, "Wavelet transforms - an introduction," *Electronics and Communication Engineering Journal*, vol. 6, no. 4, pp. 175–186, 1994.

[120] R. Merletti, "Standards for Reporting EMG Data," *Journal of Electromyography and Kinesiology*, vol. 9, pp. III–IV, 1999.

[121] A. N. Akansu and Y. P. Liu, "On-signal decomposition techniques," *Optical Engineering*, vol. 30, no. 7, pp. 912–920, 1991.

[122] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Systems With Applications*, vol. 39, no. 8, pp. 7420–7431, 2012.

[123] M. A. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal Processing*, vol. 99, pp. 215–249, 2014.

[124] M. Kristan, A. Leonardis, and D. Skocaj, "Multivariate online kernel density estimation with Gaussian kernels," *Pattern Recognition*, vol. 44, no. 10-11, pp. 2630–2642, 2011.

[125] F. Siddiqui and Q. M. Ali, "Performance of non-parametric classifiers on highly skewed data," *Global Journal of Pure and Applied Mathematics*, vol. 12, no. 2, pp. 1547–1565, 2016.

[126] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Elsevier Inc., 2009.

[127] T. MathWorks, "Predict and classification score svm." `https://se.mathworks.com/help/stats/classreg.learning.classif.compactclassificationsvm.predict.html`, 2019. [Online; accessed 26-August-2019].

[128] M. Feurer and F. Hutter, "Hyperparameter optimization," *Automated Machine Learning*, pp. 3–33, 2019.

[129] E. Brochu, V. M. Cora, and N. de Freitas, "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," *arXiv*, 2010.

[130] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," *Advances in Neural Information Processing Systems*, vol. 4, pp. 2951–2959, 2012.

[131] P. I. Frazier, "A tutorial on bayesian optimization," *arXiv*, 2018.

[132] M. Cesari, J. A. E. Christensen, F. Sixel-Döring, C. Trenkwalder, G. Mayer, W. H. Oertel, P. Jennum, and H. B. D. Sorensen, "Validation of a new data-driven automated algorithm for muscular activity detection in REM sleep behavior disorder," *Journal of Neuroscience Methods*, vol. 312, pp. 53–64, 2019.

[133] M. Cesari, J. A. E. Christensen, H. B. D. Sorensen, P. Jennum, B. Mollenhauer, M.-L. Muntean, C. Trenkwalder, and F. Sixel-Döring, "External validation of a data-driven algorithm for muscular activity identification during sleep," *Journal of Sleep Research*, p. In press, 2019.

[134] R. Miguel and I. Arnulf, "Phasic activity during non REM sleep," *Sleep Medicine*, vol. 29, pp. 37–40, 2017.

[135] U. Hanif, L. Trap, P. Jennum, M. Zoetmulder, and H. B. D. Sorensen, "Quantification of muscle activity during sleep for patients with neurodegenerative diseases," *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2015, pp. 6002–6005, 2015.

[136] C. Schenck, T. Hurtwitz, and M. Mahowald, "REM-sleep behavior disorder - an update on a series of 96 patients and a review of the world literature," *Journal of Sleep Research*, vol. 2, no. 4, pp. 224–231, 1993.

[137] R. Ferri, S. Fulda, F. I. I. Cosentino, F. Pizza, and G. Plazzi, "A preliminary quantitative analysis of REM sleep chin EMG in Parkinson's disease with or without REM sleep behavior disorder," *Sleep Medicine*, vol. 13, no. 6, pp. 707–713, 2012.

[138] R. Ferri, S. Marelli, F. I. Cosentino, F. Rundo, L. Ferini-Strambi, and M. Zucconi, "Night-to-night variability of automatic quantitative parameters of the chin EMG amplitude (atonia index) in REM sleep behavior disorder," *Journal of Clinical Sleep Medicine*, vol. 9, no. 3, pp. 253–258, 2013.

[139] T. MathWorks, "fitcecoc." `https://se.mathworks.com/help/stats/fitcecoc.html#bufm0zb-1`, 2019. [Online; accessed 22-July-2019].

[140] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[141] M. N. S. Tue Herlau and M. M, "Introduction to machine learning and data mining, course notes," *Technical University of Denmark*, 2018.

[142] H. A. Simon, E. B. Hunt, J. Marin, and P. Stone, "Experiments in induction," *American Journal of Psychology*, vol. 80, no. 4, p. 651, 1967.

[143] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits and Systems Magazine*, vol. 6, no. 3, pp. 21–44, 2006.

[144] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*. Springer, 2001.

[145] S. Nembrini, I. R. König, and M. N. Wright, "The revival of the Gini importance?," *Bioinformatics*, vol. 34, no. 21, pp. 3711–3718, 2018.

[146] T. MathWorks, "Prediction for ensemble method." `https://se.mathworks.com/help/stats/compactclassificationensemble.predict.html#bst1mxt-1-score`, 2019. [Online; accessed 27-August-2019].

[147] A. Stefani, A. Heidbreder, E. Brandauer, M. Guaita, L. M. Neier, T. Mitterling, J. Santamaria, A. Iranzo, A. Videnovic, C. Trenkwalder, F. Sixel-Döring, G. K. Wenning, A. Chade, W. Poewe, O. S. Gershanik, and B. Högl, "Screening for idiopathic REM sleep behavior disorder: Usefulness of actigraphy," *Sleep*, vol. 41, no. 6, p. zsy053, 2018.

[148] R. Boostani, F. Karimzadeh, and M. Nami, "A comparative review on sleep stage classification methods in patients and healthy individuals," *Computer Methods and Programs in Biomedicine*, vol. 140, pp. 77–91, 2017.

[149] N. Cooray, F. Andreotti, C. Lo, M. Symmonds, M. T. Hu, and M. De Vos, "Detection of REM sleep behaviour disorder by automated polysomnography analysis," *Clinical Neurophysiology*, vol. 130, no. 4, pp. 505–514, 2019.

[150] B. Mollenhauer, E. Trautmann, F. Sixel-Döring, T. Wicke, J. Ebentheuer, M. Schaumburg, E. Lang, N. K. Focke, K. R. Kumar, K. Lohmann, C. Klein, M. G. Schlossmacher, R. Kohnen, T. Friede, and C. Trenkwalder, "Nonmotor and diagnostic findings in subjects with de novo Parkinson's disease of the DeNoPa cohort," *Neurology*, vol. 81, no. 14, pp. 1226–1234, 2013.

[151] C. Iber, *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*, vol. 1. American Academy of Sleep Medicine Westchester, IL, 2007.

[152] A. A. of Sleep Medicine, *International Classification of Sleep Disorders, Second Edition*, vol. 1. Darien, Illinois, American Academy of Sleep Medicine, 2005.

[153] M. A. Sobreira-Neto, M. A. Pena-Pereira, E. S. T. Sobreira, M. H. N. Chagas, V. Tumas, R. M. F. Fernandes, and A. L. Eckeli, "Excessive fragmentary myoclonus in patients with Parkinson's disease: prevalence and clinico-polysomnographic profile," *Sleep and Breathing*, vol. 19, no. 3, pp. 997–1002, 2015.

[154] M. Cesari, J. A. E. Christensen, F. Sixel-Döring, M.-L. Muntean, B. Mollenhauer, T. Claudia, P. Jennum, and H. B. D. Sorensen, "A clinically applicable interactive micro and macro-sleep staging algorithm for elderly and patients with neurodegeneration," *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2019, pp. 3649–3652, 2019.

[155] M. Cesari, J. A. E. Christensen, M.-L. Muntean, B. Mollenhauer, F. Sixel-Döring, H. B. D. Sorensen, C. Trenkwalder, and P. Jennum, "A data-driven system to identify REM sleep

behavior disorder and to predict its progression from the prodromal stage in Parkinson's disease," p. Submitted, 2019.

[156] R. L. Zhu, C. J. Xie, P. P. Hu, and K. Wang, "Clinical variations in Parkinson's disease patients with or without REM sleep behaviour disorder: A meta-analysis," *Scientific Reports*, vol. 7, no. 1, p. 40779, 2017.

[157] D. Arnaldi, S. Morbelli, A. Brugnolo, N. Girtler, A. Picco, M. Ferrara, J. Accardo, A. Buschiazzo, F. de Carli, M. Pagani, and F. Nobili, "Functional neuroimaging and clinical features of drug naive patients with de novo Parkinson's disease and probable RBD," *Parkinsonism and Related Disorders*, vol. 29, pp. 47–53, 2016.

[158] M. Salsone, A. Cerasa, G. Arabia, M. Morelli, A. Gambardella, L. Mumoli, R. Nisticó, B. Vescio, and A. Quattrone, "Reduced thalamic volume in Parkinson's disease with REM sleep behavior disorder: Volumetric study," *Parkinsonism and Related Disorders*, vol. 20, no. 9, pp. 1004–1008, 2014.

[159] J.-F. Gagnon, M. L. Fantini, M.-A. Bédard, D. Petit, J. Carrier, S. Rompré, A. Décary, M. Panisset, and J. Montplaisir, "Association between waking EEG slowing and REM sleep behavior disorder in PD without dementia," *Neurology*, vol. 62, no. 3, pp. 401–6, 401–406, 2004.

[160] L. Fiorillo, A. Puiatti, M. Papandrea, P. L. Ratti, P. Favaro, C. Roth, P. Bargiotas, C. Bassetti, and F. D. Faracia, "Automated sleep scoring: A review of the latest approaches," *Sleep Medicine Reviews*, p. In press, 2019.

[161] A. B. Klok, J. Edin, M. Cesari, A. N. Olesen, P. Jennum, and H. B. D. Sorensen, "A new fully automated random-forest algorithm for sleep staging," *40th Annual International Conference of the Ieee Engineering in Medicine and Biology Society*, vol. 2018, pp. 4920–4923, 2018.

[162] A. R. Hassan and M. I. H. Bhuiyan, "Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating," *Biomedical Signal Processing and Control*, vol. 24, pp. 1–10, 2016.

[163] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 1998–2008, 2017.

[164] H. Phan, F. Andreotti, N. Cooray, O. Y. Chen, and M. De Vos, "Joint classification and prediction CNN framework for automatic sleep stage classification," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 5, pp. 1285–1296, 2019.

[165] J. Stephansen, A. Olesen, M. Olsen, A. Ambati, E. Leary, H. Moore, O. Carrillo, L. Lin, F. Han, H. Yan, Y. Sun, Y. Dauvilliers, S. Scholz, L. Barateau, B. Hogl, A. Stefani, S. Hong, T. Kim, F. Pizza, G. Plazzi, S. Vandi, E. Antelmi, D. Perrin, S. Kuna, P. Schweitzer, C. Kushida, P. Peppard, H. Sorensen, P. Jennum, and E. Mignot, "Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy," *Nature Communications*, vol. 9, no. 1, p. 5229, 2018.

[166] H. Danker-Hopfe, P. Anderer, J. Zeitlhofer, M. Boeck, H. Dorn, G. Gruber, E. Heller, E. Loretz, D. Moser, S. Parapatics, B. Saletu, A. Schmidt, and G. Dorffner, "Interrater

reliability for sleep scoring according to the Rechtschaffen and Kales and the new AASM standard," *Journal of Sleep Research*, vol. 18, no. 1, pp. 74–84, 2009.

[167] H. Danker-Hopfe, D. Kunz, G. Gruber, G. Klösch, J. L. Lorenzo, S. L. Himanen, B. Kemp, T. Penzel, J. Röschke, H. Dorn, A. Schlögl, E. Trenker, and G. Dorffner, "Interrater reliability between scorers from eight European sleep laboratories in subjects with different sleep disorders," *Journal of Sleep Research*, vol. 13, no. 1, pp. 63–69, 2004.

[168] J. Kempfner, P. Jennum, H. B. D. Sorensen, J. A. E. Christensen, and M. Nikolic, "Automatic SLEEP staging: From young adults to elderly patients using multi-class support vector machine," *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2013, pp. 5777–5780, 2013.

[169] A. Patanaik, J. L. Ong, J. J. Gooley, S. Ancoli-Israel, and M. W. Chee, "An end-to-end framework for real-time automatic sleep stage classification," *Sleep*, vol. 41, no. 5, 2018.

[170] H. Koch, P. Jennum, and J. A. E. Christensen, "Automatic sleep classification using adaptive segmentation reveals an increased number of rapid eye movement sleep transitions," *Journal of Sleep Research*, vol. 27, no. 2, p. In press, 2018.

[171] M. Younes, S. T. Kuna, A. I. Pack, J. K. Walsh, C. A. Kushida, B. Staley, and G. W. Pien, "Reliability of the American Academy of Sleep Medicine rules for assessing sleep depth in clinical practice," *Journal of Clinical Sleep Medicine*, vol. 14, no. 2, pp. 205–213, 2018.

[172] P. He, G. Wilson, and C. Russell, "Removal of ocular artifacts from electro-encephalogram by adaptive filtering," *Medical and Biological Engineering and Computing*, vol. 42, no. 3, pp. 407–412, 2004.

[173] A. Muñoz, R. Ertlé, and M. Unser, "Continuous wavelet transform with arbitrary scales and O(N) complexity," *Signal Processing*, vol. 82, no. 5, pp. 749–757, 2002.

[174] M. Unser, A. Aldroubi, and S. J. Schiff, "Fast implementation of the continuous wavelet transform with integer scales," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3519–3523, 1994.

[175] T. MathWorks, "cwt." `https://se.mathworks.com/help/wavelet/ref/cwt.html`, 2019. [Online; accessed 06-August-2019].

[176] Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang, "The expressive power of neural networks: A view from the width," *Advances in Neural Information Processing Systems 30 (nips 2017)*, vol. 30, 2017.

[177] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press„ 2016.

[178] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd International Conference on Learning Representations*, 2015.

[179] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochemia Medica*, vol. 22, no. 3, pp. 276–282, 2012.

[180] R. Ferri, F. Pizza, S. Vandi, M. Iloti, and G. Plazzi, "Decreased sleep stage transition pattern complexity in narcolepsy type 1," *Clinical Neurophysiology*, vol. 127, no. 8, pp. 2812–2819, 2016.

[181] D. P. Brunner, R. C. Vasko, C. S. Detka, J. P. Monahan, C. F. Reynolds, and D. J. Kupfer, "Muscle artifacts in the sleep EEG: Automated detection and effect on all-night EEC power spectra," *Journal of Sleep Research*, vol. 5, no. 3, pp. 155–164, 1996.

[182] O. L. Frost, "Power-spectrum estimation," *Proceedings of the Nato Advanced Study Institute of Signal Processing*, pp. 125–62, 1977.

[183] J. A. E. Christensen, E. G. S. Munk, P. E. Peppard, T. Young, E. Mignot, H. B. D. Sorensen, and P. Jennum, "The diagnostic value of power spectra analysis of the sleep electroencephalography in narcoleptic patients," *Sleep Medicine*, vol. 16, no. 12, pp. 1516–1527, 2015.

[184] M. Waser, T. Benke, P. Dal-Bianco, H. Garn, J. A. Mosbacher, G. Ransmayr, R. Schmidt, S. Seiler, H. B. D. Sorensen, and P. J. Jennum, "Neuroimaging markers of global cognition in early Alzheimer's disease: A magnetic resonance imaging-electroencephalography study," *Brain and Behavior*, vol. 9, no. 1, p. e01197, 2018.

[185] E. Parzen, "On estimation of a probability density function and mode," *Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.

[186] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology*, vol. 278, no. 6 Part 2, pp. H2039–H2049, 2000.

[187] C. C. Chung, J. H. Kang, R. Y. Yuan, D. Wu, C. C. Chen, N. F. Chi, P. C. Chen, and C. J. Hu, "Multiscale entropy analysis of electroencephalography during sleep in patients with Parkinson's disease," *Clinical EEG and Neuroscience*, vol. 44, no. 3, pp. 221–226, 2013.

[188] S. Sohail, L. Yu, J. A. Schneider, D. A. Bennett, A. S. Buchman, and A. S. Lim, "Sleep fragmentation and Parkinson's disease pathology in older adults without Parkinson's disease," *Movement Disorders*, vol. 32, no. 12, pp. 1729–1737, 2017.

[189] C. S. Kim, Y. H. Sung, M. J. Kang, and K. H. Park, "Rapid Eye Movement Sleep Behavior Disorder in Parkinson's Disease: A Preliminary Study," *Journal of Movement Disorders*, vol. 9, no. 2, pp. 114–9, 2016.

[190] R. B. Postuma, C. H. Adler, B. N. Dugger, J. G. Hentz, H. A. Shill, E. Driver-Dunckley, M. N. Sabbagh, S. A. Jacobson, C. M. Belden, L. I. Sue, G. Serrano, and T. G. Beach, "REM sleep behavior disorder and neuropathology in Parkinson's disease," *Movement Disorders*, vol. 30, no. 10, pp. 1413–1417, 2015.

[191] J. Santamaria, B. Högl, C. Trenkwalder, and D. Bliwise, "Scoring sleep in neurological patients: The need for specific considerations," *Sleep*, vol. 34, no. 10, pp. 1283–1284, 2011.

Appendix **A**

Paper I

Title: Comparison of computerized methods for rapid eye movement sleep without atonia detection

Authors: Matteo Cesari, Julie A. E. Christensen, Lykke Kempfner, Alexander N. Olesen, Geert Mayer, Karl Kesper, Wolfgang H. Oertel, Friederike Sixel-Döring, Claudia Trenkwalder, Helge B. D. Sorensen and Poul Jennum

Journal: Sleep, vol. 41, n. 10, pp. zsy133, 2018

Status: Published

DOI: 10.1093/sleep/zsy133

URL: `https://academic.oup.com/sleep/article/41/10/zsy133/5053112`

Appendix **B**

Paper II

Title: Probabilistic Data-Driven Method for Limb Movement Detection during Sleep

Authors: Matteo Cesari, Julie A. E. Christensen, Poul Jennum and Helge B. D. Sorensen

Journal: Proceedings of 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), vol. 2018, pp. 163-166, 2018

Status: Published

DOI: 10.1109/EMBC.2018.8512254

URL: `https://ieeexplore.ieee.org/document/8512254`

Appendix C

Paper III

Title: Validation of a new data-driven automated algorithm for muscular activity detection in REM sleep behavior disorder

Authors: Matteo Cesari, Julie A. E. Christensen, Friederike Sixel-Döring, Claudia Trenkwalder, Geert Mayer, Wolfgang H. Oertel, Poul Jennum and Helge B. D. Sorensen

Journal: Journal of Neuroscience Methods, vol. 312, pp. 53-64, 2019

Status: Published

DOI: 10.1016/j.jneumeth.2018.11.016

URL: `https://www.sciencedirect.com/science/article/pii/S0165027018303844?via%3Dihub`

Appendix **D**

Paper IV

Title: External validation of a data-driven algorithm for muscular activity identification during sleep

Authors: Matteo Cesari, Julie A. E. Christensen, Helge B. D. Sorensen, Poul Jennum, Brit Mollenhauer, Maria-Lucia Muntean, Claudia Trenkwalder and Friederike Sixel-Döring

Journal: Journal of Sleep Research, vol. 28, issue 6, pp. e12868, 2019

Status: Published

DOI: 10.1111/jsr.12868

URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/jsr.12868`

Appendix **E**

Paper V

Title: A Clinically Applicable Interactive Micro and Macro-Sleep Staging Algorithm for Elderly and Patients with Neurodegeneration

Authors: Matteo Cesari, Julie A. E. Christensen, Friederike Sixel-Döring, Maria-Lucia Muntean, Brit Mollenhauer, Claudia Trenkwalder, Poul Jennum and Helge B. D. Sorensen

Journal: Proceedings of 41th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), vol. 2019, pp.3649-3652, 2019

Status: Published

DOI: 10.1109/EMBC.2019.8856705

URL: `https://ieeexplore.ieee.org/document/8856705`

Appendix F

Paper VI

Title: A data-driven system to identify REM sleep behavior disorder and to predict its progression from the prodromal stage in Parkinson's disease

Authors: Matteo Cesari, Julie A. E. Christensen, Maria-Lucia Muntean, Brit Mollenhauer, Friederike Sixel-Döring, Helge B.D. Sorensen, Claudia Trenkwalder and Poul Jennum

Status: Submitted

DOI: N/A