Doctoral Dissertation
Doctoral Program in Computer and Control Engineering (36^{th} cycle)

Sleep in Neurodegenerative Diseases: an Integrated Approach to Diagnosis and Monitoring

Irene Rechichi

* * * * *

Supervisor
Prof. Gabriella Olmo

Politecnico di Torino
19th June, 2024

I hereby declare that the contents and structure of this dissertation constitute my own original work and do not compromise in any way the rights of third parties, including those relating to the security of personal data.

. . . . . .

Irene Rechichi

Turin, 19th June, 2024

Summary

Sleep disorders encompass a diverse spectrum of motor and non-motor manifestations, representing an exceedingly common condition with a global prevalence of up to 70% in older adults. In recent years, they have emerged as a growing challenge worldwide due to their association with an increased risk of neuronal degeneration. Indeed, they may serve as predisposing factors of co-morbidities, negatively impacting the quality of life. In particular, REM sleep parasomnias have been acknowledged among the earliest markers of neurodegenerative disorders, such as α-synucleinopathies.

In this perspective, sleep embodies a reservoir of significant clinical information, and its role becomes twofold.

First, in light of the development of neuroprotective pharmacological treatment, identifying prodromal conditions, such as REM Sleep Behaviour Disorder (RBD), may offer a potential window for disease-modifying interventions with beneficial effects on the quality of life. Second, an objective characterisation of sleep becomes necessary to deliver effective monitoring strategies for neurodegenerative diseases, resulting in personalised healthcare outcomes and an improved quality of care.

However, state-of-the-art diagnostic procedures entail complex tests and visual, rule-based assessments, ofttimes resulting in protracted manual labour.

Hence, this Thesis addressed two aspects: *Diagnostic Support Systems* and *Monitoring Systems* by exploiting polysomnographic biosignals and Machine Learning techniques.

The first part addressed the development of minimally intrusive diagnosis support strategies. First, a single-channel EEG framework for automatic sleep staging was proposed, to assess the feasibility of lightweight sleep studies. The pipeline yielded encouraging results when tested in healthy subjects and patients with RBD. Then, a method to support the automatic removal of artefacts in EMG recordings when assessing REM Sleep Without Atonia was proposed. The approach exploited the morphology of EMG activity during sleep and provided high agreement with manual procedures. Subsequently, biosignals collected during sleep (EMG and EEG, respectively) were exploited to characterise subjects with RBD and highlight the differences with healthy recordings. The extracted parameters exhibited good predictive power and tackled the automatic detection of RBD with reasonably high

validation accuracies. Finally, a continuous, rater-independent metric was devised, to assess the extent of disease progression in RBD.

The second part explored lightweight strategies to monitor sleep disorders in neurodegenerative diseases. First, a set of inertial metrics was proposed to objectively characterise sleep-related motor disturbances and sleep quality in Parkinson's Disease through a wearable set-up. Second, non-intrusive EMG metrics collected during REM sleep were employed, to build a predictive model of mortality risk in Amyotrophic Lateral Sclerosis in a longitudinal retrospective fashion.

To conclude, the presented research activities explored the feasibility of data-driven approaches for the identification and monitoring of neurodegenerative conditions through the analysis of sleep. The encouraging findings represent possible approaches for minimally intrusive, accessible, and lightweight sleep studies, for an improved quality of care.

Preface

This Dissertation, *Sleep in Neurodegenerative Diseases: an Integrated Approach to Diagnosis and Monitoring*, illustrates the research activities conducted within the PhD Programme in Computer and Control Engineering, at Politecnico di Torino (Italy), and included in 6 scientific contributions. The Programme had a total duration of 3 years (1st November 2020–31st October 2023).

All experimental activities envisaged a multidisciplinary approach, and were conducted in cooperation with domain specialists in hospitals both in Italy and abroad.

The Regional Centre for Sleep Medicine (Turin, Italy), and the Departments of Neurology and Neuroscience, respectively, of the University of Turin (Italy) were the main partners in the clinical collaborations. In addition, a part of the research work was conducted at the Medical University of Innsbruck, during a secondment in the period July 2022–October 2022, and continued remotely until the end of the PhD.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

> That we are not much sicker and much madder than we are, is due exclusively to that most blessed and blessing of all natural graces, sleep.
>
> *Aldous Huxley*

Sleep is an essential asset in human health, due to its crucial and intricate involvement in various physiological processes. Beyond providing physical and constitutional recovery and restoration, robust and healthy sleep enhances cognitive functioning, supports emotional regulation, and bolsters immune resilience [174].

Although no comprehensive definition has yet been provided, sleep is typically defined as a recurrent and reversible state of altered responsiveness, associated with a diminished reactivity to external stimuli [39]. As introduced, this apparent relational idleness serves a pivotal role in supporting the fundamental functioning of the organism and brain health. Indeed, besides the valuable contributions mentioned above, recent findings identified an association between healthy sleep and the glymphatic pathways in the central nervous system. These latter promote the clearance of potential neurotoxic products during the most quiescent phase of sleep [78].

Disruptions in ultradian rhythms, due to sleep deprivation or insomnia, breathing-related sleeping disorders, or alterations in sleep architecture, negatively impact sleep quality. These disturbances lead to the impairment of the physiological glymphatic function, ultimately leading to the accumulation of toxic metabolites in the brain, including the tau (τ) protein, the amyloid-β, and lactate [175].

With mounting evidence of their association with increased risks of neuronal degeneration, the spectrum of sleep disorders (SD) emerged as a significant challenge to global public health. Particularly, due to their involvement in the pathogenesis, development, and progression of various neurodegenerative processes, they have

been listed among the earliest prodromes of tauopathies and synucleinopathies, thus sharing important features with Parkinson's Disease (PD) and Alzheimer's Disease (AD). Indeed, it has been shown that an altered sleep architecture, entailing sleep fragmentation or reduced slow-wave sleep (SWS), or the presence of REM sleep parasomnias, may precede by several years the onset of typical symptoms [82].

Given their valuable potential as a reservoir of relevant clinical information, research on SD has increasingly gained attention, both for early diagnosis purposes and risk stratification. In this perspective, the identification of either biochemical markers or digital biomarkers as prodromal indicators of neurodegeneration holds pivotal importance for an effective, timely diagnosis, precise and personalised prevention and intervention strategies, and the development of specific disease-modifying therapies. The adoption of such elements in the clinical practice might potentially improve disease management, and promote personalised healthcare strategies, positively impacting the quality of life.

However, currently, the diagnosis, monitoring, and follow-up of sleep disorders presents with many challenges. To begin with, state-of-the-art diagnostic protocols for sleep disorders, such as polysomnography (PSG) involve complex infrastructures. Indeed, PSG entails the overnight collection of numerous electrophysiological recordings, including electroencephalography (EEG), electromyography (EMG), electrooculography (EOG), electrocardiography (ECG) and respiratory airflow, often in combination with an infrared camera for monitoring body movements. This diagnostic test, though highly accurate, is rather intrusive, as it entails wire electrodes and cumbersome instrumentation. A proper implementation generally demands specialised facilities, with trained medical personnel. Furthermore, in-hospital PSG is costly, resource-intensive, and requires visual analysis for screening, regulated by a set of international rules [11], often leading to protracted manual labour and subjectivity.

Similarly, follow-ups procedures for the vast majority of sleep disorders are infrequent, and rely on the visual analysis of PSG recordings to retrieve threshold-based clinical metrics, thus making clinical assessments prone to inter- and intra-rater variability.

Finally, conditions such as REM sleep parasomnias or neurodegenerative disorders are associated with sleep-related motor manifestations. These latter encompass a broad spectrum, ranging from violent and abrupt motor behaviour – such as in the case of REM Sleep Behaviour Disorder (RBD) – to the inability of turning in bed (nocturnal hypokinesia and akinesia). Monitoring these symptoms remains a challenging task, as follow-up relies primarily on subjective reports by the patient or caregiver, and lacks adequate assessment scales.

In recent years, the development of computer-assisted diagnosis systems, as well as advancements in sleep monitoring technologies, including portable PSG devices, actigraphy, and wearable sensors, provided an encouraging scenario for the development of minimally-intrusive sleep studies and continuous monitoring

strategies. Lightweight data collection, in combination with Machine Learning (ML) algorithms, offers the possibility to identify specific digital biomarkers for an objective disease characterisation, in order to facilitate the diagnostic procedures and allowing for timely diagnosis and effective disease management.

In this context, this Thesis seeks to explore the advantages of lightweight, data-driven approaches for monitoring sleep and propose accessible monitoring strategies for individuals at risk of, or diagnosed with, neurodegenerative diseases (ND). Specifically, with the purposes of: **(1)** supporting minimally invasive sleep studies for early diagnosis, and **(2)** providing lightweight TeleHealth frameworks for continuous monitoring.

Therefore, it covers two main themes: *Diagnostic Support Systems*, and *Monitoring Systems*, reflecting the two parts of the Thesis; Figure 1.1 provides a graphical outline.

The first part of this dissertation (*Diagnostic Support Systems*, Chapters 4–6) will focus on the tools developed for the diagnosis, and the feasibility of minimally-intrusive sleep studies, while the second part, (*Monitoring Systems*, Chapters 7–8) will illustrate the proposed methodology for accessible sleep monitoring (home-based, or through in-patient settings), and continuous monitoring of sleep disorders in neurodegenerative diseases.

In more detail, the document is structured as follows.

Chapter 2 offers an insight into the neurobiology and electrophysiology of human sleep, and the clinical background on sleep disorders and neurodegenerative diseases, with a focus on REM parasomnias and Parkinson's Disease.

Chapter 3 provides an overview of the employed Machine Learning methods.

Chapter 4 presents a method for automatic sleep staging based on single-channel EEG data, to test the feasibility of minimally intrusive PSG.

Chapter 5 is dedicated to the assessment of muscle activity during REM sleep, and provides a preliminary approach to the automatic detection of artefacts in the quantification of REM Sleep Without Atonia.

Chapter 6 presents a data-driven, diagnosis support strategy to automatically identify REM Sleep Behaviour Disorder from polysomnography data, to investigate the feasibility of lightweight, sensor-based detection of RBD. Specifically, it focuses on the characterisation of RBD through EEG and EMG signals. Finally, it prototypes a distance-based method for the evaluation of RBD progression over time.

Chapter 7 describes a wearable set-up for the assessment of sleep quality and motor disturbances during sleep in subjects with Parkinson's Disease, in unsupervised, real-world, scenarios.

Chapter 8 illustrates a preliminary study for the prediction of Amyotrophic Lateral Sclerosis through simple, EMG based metrics.

Chapter 9 reports the principal findings of the research work included in the Thesis, and concludes the dissertation by outlining possible future trajectories.

The PhD project was carried out at the Department of Control and Computer Engineering of Politecnico di Torino, Italy. The research activities presented in this dissertation were conducted in cooperation with the Regional Centre for Sleep Medicine at the Molinette University Hospital (Turin, Italy) and the Sleep Disorders Unit at the Medical University of Innsbruck (Austria). The experimental work involved the analysis of electrophysiological signals, inertial measurements and medical data, to provide a data-driven, integrated approach to the diagnosis and monitoring of sleep and related disorders. All data processing was carried out on MATLAB® and Python, through both custom-written code and open-source libraries.

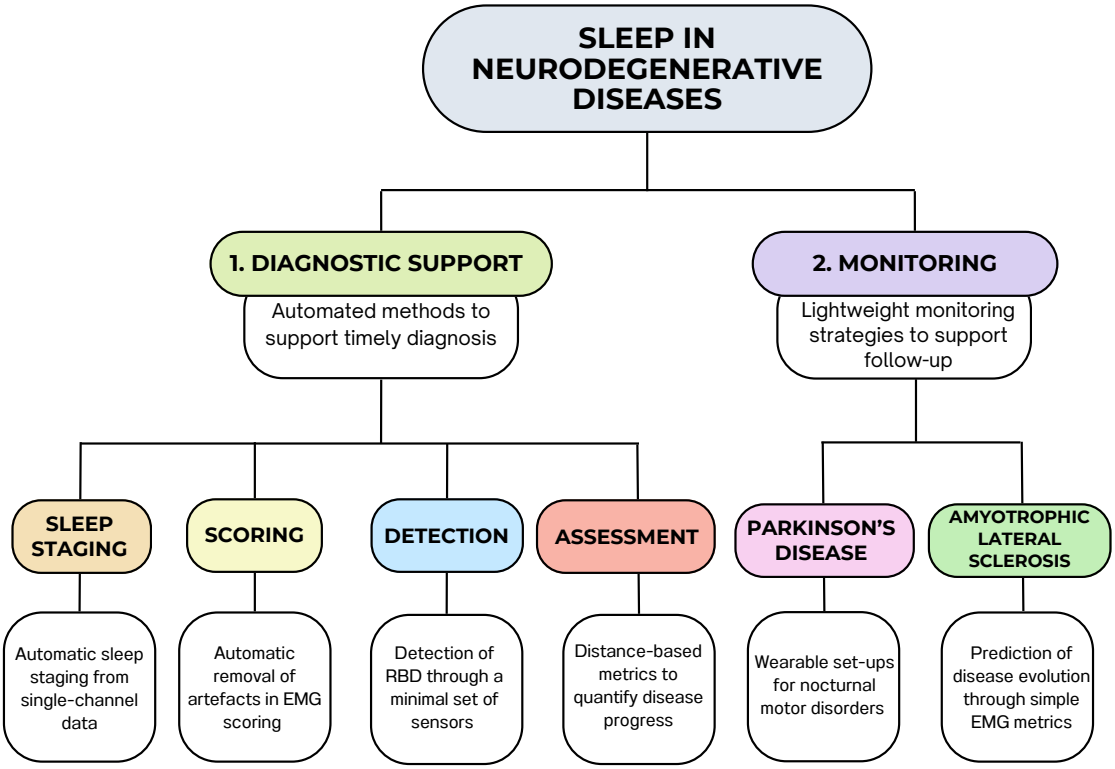


Figure 1.1: Graphical overview of the thesis structure. The topics are developed around the two macro-themes (Diagnosis support and Monitoring).

1.1 Scientific Contributions

The research activities presented in this Thesis were published in the following scientific contributions, including journal articles, conference papers, and conference abstracts.

Journal Articles

- Rechichi, I., Zibetti, M., Borzì, L., Olmo, G., and Lopiano, L. (2021). **Single-channel EEG classification of sleep stages based on REM microstructure**. Healthcare technology letters, 8(3), 58-65. [141]

- Rechichi, I., Iadarola, A., Zibetti, M., Cicolin, A., and Olmo, G. (2021). **Assessing rem sleep behaviour disorder: From machine learning classification to the definition of a continuous dissociation index.** International Journal of Environmental Research and Public Health, 19(1), 248. [139]

Conference Papers

- Rechichi, I., Amato, F., Cicolin, A., and Olmo, G. (2022, June). **Single-Channel EEG Detection of REM Sleep Behaviour Disorder: The Influence of REM and Slow Wave Sleep.** In International Work-Conference on Bioinformatics and Biomedical Engineering (pp. 381-394). Cham: Springer International Publishing. [142]

- Rechichi, I., Di Gangi, L., Zibetti, M., and Olmo, G. (2024, March). **Home Monitoring of Sleep Disturbances in Parkinson's Disease: A Wearable Solution.** In 2024 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops). IEEE.

- Rechichi, I., Amprimo, G., Cicolin, A., and Olmo, G. (2024). **Predicting Amyotrophic Lateral Sclerosis Progression: an EMG-based Survival Analysis** *(Accepted at IEEE EMBC2024.)*

Conference Abstracts

- Rechichi, I., Olmo, G., Stefani, A., Heidbreder, A., Holzknecht, E., Bergmann, M., Ibrahim, A., Brandauer, E., Högl, B., and Cesari, M. (2024). **Towards fully automatic quantification of REM sleep without atonia according to the Sleep Innsbruck Barcelona (SINBAR) scoring method**. Sleep Medicine, 115(1), 307 [143]

Over the course of the PhD, other scientific contributions were co-authored, focusing on Artificial Intelligence approaches for the detection, monitoring, and rehabilitation of Parkinson's Disease, and diagnostic support systems for sleep disorders in neurodegenerative disorders.

They were not included in the scientific content of this Thesis, and are listed below.

- Amato, F., Rechichi, I., Borzì, L., and Olmo, G. (2022, March). **Sleep Quality through Vocal Analysis: a Telemedicine Application**. In 2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops) (pp. 706-711). IEEE.

- Amprimo, G.†, Rechichi, I.†,[1], Ferraris, C., and Olmo, G. (2023). **Measuring Brain Activation Patterns from Raw Single-Channel EEG during Exergaming: A Pilot Study**. Electronics, 12(3), 623

- Masi, G., Amprimo, G., Rechichi, I., Ferraris, C., and Priano, L. (2023, March). **Electrodermal Activity in the Evaluation of Engagement for Telemedicine Applications**. In 2023 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops) (pp. 130-135). IEEE.

- Sigcha, L., Borzì, L., Amato, F., Rechichi, I., Ramos-Romero, C., Cárdenas, A., and Olmo, G. (2023). **Deep learning and wearable sensors for the diagnosis and monitoring of Parkinson's disease: a systematic review**. Expert Systems with Applications, 120541.

- Amprimo, G., Rechichi, I., Ferraris, C., and Olmo, G. (2023, June). **Objective Assessment of the Finger Tapping Task in Parkinson's Disease and Control Subjects using Azure Kinect and Machine Learning**. In 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS) (pp. 640-645). IEEE.

- Cesari, M., and Rechichi, I. (2023). **Automatic and machine learning methods for detection and characterization of REM sleep behavior disorder**. Handbook of AI and Data Sciences for Sleep Disorders, *(In Press)*

- Masi, G., Amprimo, G., Rechichi, I., Ferraris, C., and Olmo, G., (2024). **Does Baseline Stress Affect Electrodermal Activity? A Serious-Game-Based Pilot Study**. *Accepted for oral presentation at IEEE EMBC2024*

[1] †: shared first authorship.

Chapter 2

Clinical Background

The scientific content of this Thesis discusses technological strategies to leverage sleep for the extraction of significant clinical information in neurodegenerative diseases, with a special focus on Parkinson's Disease.

The following Sections will provide an overview on the clinical background, by introducing the electrophysiology of sleep, and describing sleep disorders affecting α-synucleinopathies. Finally, the two neurodegenerative disorders dealt with in this Thesis are presented.

2.1 Sleep

Sleep is a fundamental biological process, that is vital for human well-being. There is evidence that healthy sleep bolsters proper cognitive function in adults, supports physical health, and overall well-being [137]. The concept of *healthy* sleep entails not only adequate quantity, but also proper sleep hygiene and sleep structure. Indeed, disruption to physiological sleep-wake cycles, or to the macrostructure of sleep, have a significantly negative impact on overall health, with observable detriments on cognitive function and cardiovascular health, ultimately leading to an increased risk of mortality.

From a metabolical and electrophysiological perspective, sleep is a complex process that cyclically alternates between two major neurophysiological states: rapid eye movement (REM) sleep and non-rapid eye movement (NREM) sleep. The NREM state consists of various sub-phases, which can be interpreted as gradual transitions in waveform, detected by EEG.

2.1.1 Architecture of Sleep Stages

Conventionally, the electrophysiological macrostructure of sleep is arranged over distinct phases, commonly referred to as *sleep stages*. The first standardisation of sleep stages was provided by Rechtschaffen and Kales [89], and envisaged four

stages of REM sleep, from light to deep. An update to this schema was later proposed by the American Academy of Sleep Medicine (AASM), and encompassed the union of NREM stages 3 and 4. It is a rule-based method, and details about the characteristics of each stage are provided in the following paragraphs; the sleep stages according to these guidelines are displayed in Figure 2.1.

Wake (W) Although not representing a quiescent state, wakefulness is still included in the categorisation of sleep stages, as it indicates both awakening and arousals. Open-eye wakefulness is characterised by the prevalence of beta waves (13–21 Hz), which represent high-frequency activity. During drowsiness and the transition from quiet wakefulness to light sleep, the posterior dominant rhythm (alpha waves, 8–13 Hz) appears, and becomes the predominant brain rhythm. The amplitude of the EEG signal is low, and eye-movements are present.

Stage 1 (N1) It is the first sleep stage, generally known as light sleep. It serves as a transitional state between relaxed wakefulness to deeper stages, and accounts for 5% of the total duration of sleep. From an electrophysiological point of view, alpha waves are gradually replaced by slower, theta waves (4–7 Hz).

Stage 2 (N2) This stage is the first phase of deeper sleep, and features a slower heart rate, as well as lower body temperature. It comprises up to 45% of total sleep time, starting with shorter transitions and eventually consolidating over successive sleep cycles. Regarding electrical activity, it features peculiar waveforms in the EEG. Specifically, K-complexes and sleep spindles. The former are high-amplitude delta waves, in which a negative polarisation is followed by a positive peak, reaching up to 100 µV. These occur around every minute, and are frequently followed by sleep spindles. These latter, on the other hand, are rapid bursts of electrical activity, exhibiting frequency components in the range 12–15 Hz, and duration of 0.5 s. These oscillatory patterns are believed to support information processing, memory consolidation, and maintain sleep [7].

Stage 3 (N3) This stage comprises the deepest phases of sleep, and is also referred to as slow-wave sleep (SWS). It is characterised by the lowest observable rhythms, with high-amplitude delta waves (0.5–4 Hz), and constitutes up to 25% of sleep duration. It is the most quiescent of stages, and this reflects in idle muscle activity and the absence of eye movements. This stage is closely correlated with correct functioning of glymphatic pathways, a concept that will be explored in following Sections.

REM Stage REM sleep, or *paradoxical sleep*, accounts for up to 25% of total sleep time. Compared to the other stages, it exhibits a multifaceted fingerprint. The EEG activity during this stage reveals mixed frequency components, and low-amplitude waves, recalling an awake state. These characteristics reflect an extensively active underlying brain activity. Indeed, during REM sleep vivid dreaming occurs, though accompanied by total atonia in the skeletal muscles, except for the eyes and diaphragm. Although commonly treated as an homogeneous state, evidence showed the presence of two distinct micro-states [159]; namely, a *phasic* and a *tonic* state, characterised by different electrophysiological patterns, reflecting diverse cortical activations. The tonic phase is the most quiescent; eye movements are slow and rare (amplitude lower than 25 µV in a 4-second range), and skeletal muscles are atonic. Conversely, the phasic state is characterised by the peculiar, rapid bursts of electrical activity in the EOG – i.e., the rapid-eye movements – fast, low-amplitude EEG waves (*sawtooth* waves), irregular heart rhythm and erratic breathing, which results in increased metabolic rate of up to 20% [128]. Similarly to N2, the REM stage starts with shorter periods, which become longer toward the end of sleep time.

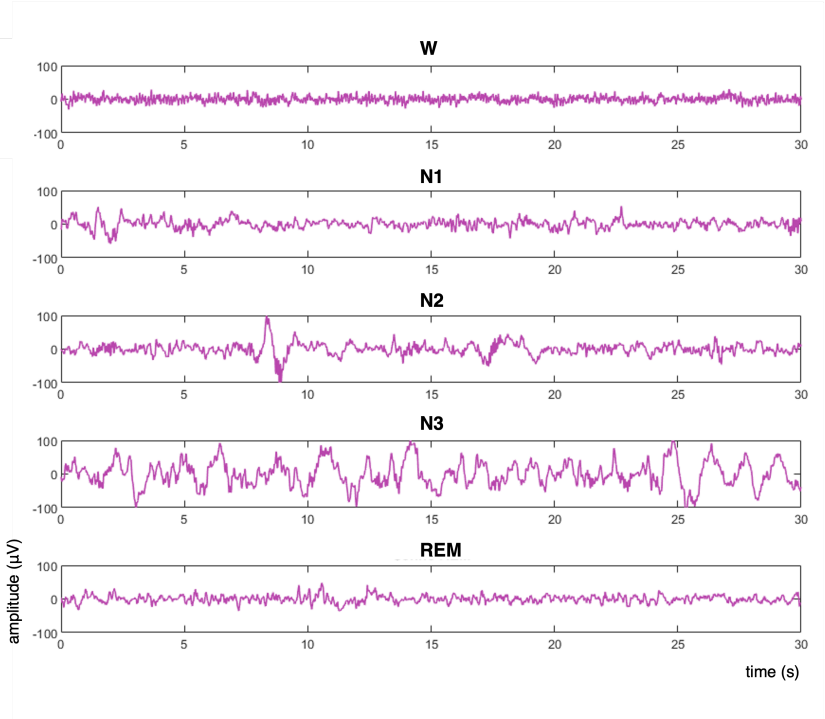


Figure 2.1: Representation of a 30-second EEG epoch for each sleep stage according to the AASM classification.

This overall architecture constitutes the ultradian sleep cycle, which envisages the cyclic alternation between NREM and REM sleep. The average cycle length in healthy adults is of about 90 minutes; over a night's sleep, up to 5 cycles may occur. The architecture is clinically described by the hypnogram (Figure 2.2), which exhibits the timing of each sleep cycle, as well as the inter-cycle duration transformation for each stage.

Variations in the sleep macrostructure and in the regularity of cycles commonly occur with aging. Indeed, the amount of slow-wave sleep decreases with age, with elderly adults presenting with a considerably shorter N3 duration, and a relatively fragmented sleep architecture, leading to an increased risk of developing neurological disorders, as will be discussed in Section 2.1.3.

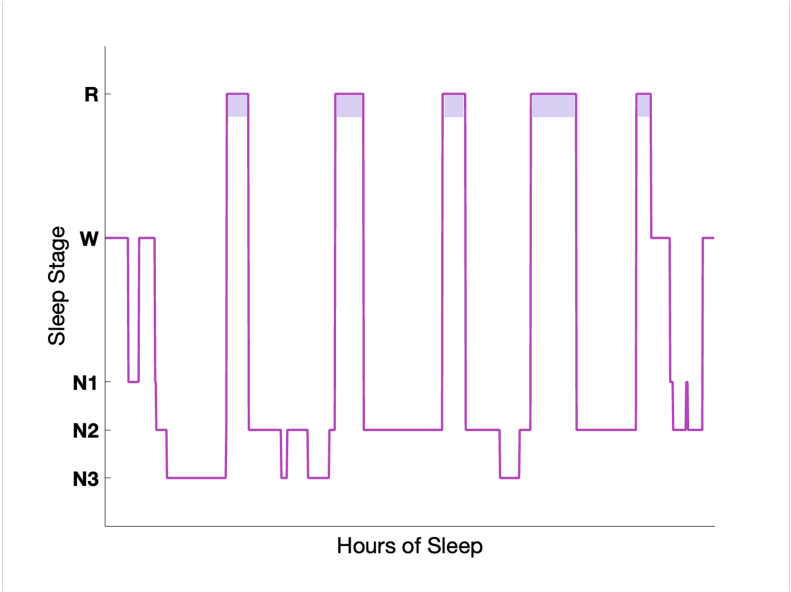


Figure 2.2: Hypnogram of a healthy subject. After a gradual transition to slow-wave sleep, this latter becomes less frequent in the second half of the night. The REM stage (R) becomes slightly lower and then decreases in duration, towards the awakening. The sleep data employed to generate this figure were downloaded with permission from PhysioNet [72].

2.1.2 Polysomnography

Polysomnography is a multi-parametric assessment that stands as the gold standard diagnostic tool in sleep studies. It allows the identification of sleep stages, and the subsequent evaluation of sleep architecture, salient features, sleep-disordered breathing, and sleep disorders.

It consists in the simultaneous recording of various electrophysiological signals during sleep, through electrodes and sensors, and is generally conducted in equipped

sleep labs or hospitals. A high number of wired electrodes is employed to record continuous EEG, EMG, EOG, ECG, and oxygen saturation. In addition, piezoelectric sensors, and, in some cases, inertial sensors, are employed to measure respiratory activity and body position overnight. In traditional PSG, all recordings are streamed on a display for concurrent visual inspection and technical artefact monitoring. Additionally, video recordings through infrared cameras have been introduced (vPSG), allowing for a more accurate diagnosis of complex sleep behaviours, including REM/NREM parasomnias, motor disturbances, and vocalisations.

A concise description of the employed electrophysiological recordings follows, to better comprehend their usage in the proposed research activities, along with the montage and technical specifications recommended by the AASM scoring manual. For the sake of clarity, in this Thesis, electrophysiological recordings will also be referred to as *biosignals*.

EEG It measures and records electrical activity in the brain, through wired electrodes placed on the scalp. It allows the classification of sleep stages by visual inspection of the detected activity. The AASM Scoring Manual requires the recording of derivations F4-M1, C4-M1, O2-M1 (Figure 2.3), in which M1 is the reference electrode, positioned at the left mastoid. This location is also commonly referred to as A1. Three additional derivations are recommended as backup, and are C3-M2, F3-M2, O1-M2, where M2 is located at the right mastoid. The minimum sampling frequency required for these derivations is 200 Hz, and a visual bandpass filter (range 0.3–35 Hz) is required for signal inspection.

EMG Surface EMG is employed to record the electrical activity of skeletal muscles, and retrieve information about muscle tone, which reflects voluntary movements during sleep or abrupt and periodic muscle contractions. It entails a non-invasive recording through electrodes placed above the muscle. In PSG, the first required derivation is chin EMG, revealing activity of the mylohyoid muscle, through three electrodes. Additionally, recordings of the tibialis anterior (TA) are required from both legs. Optional configurations envisage the bilateral recording of the flexor digitorum superficialis (FDS) muscle; all derivations are displayed in Figure 2.5. Technical requirements for sleep EMG entail a minimum sampling frequency of 200 Hz and a visual bandpass filter in the range 10-100 Hz.

EOG The electrooculogram measures the relative electrical potential between the cornea and retina. Essentially, it tracks horizontal and vertical eye movements, and is employed collectively with the EEG to accurately detect the REM stage (through the identification of rapid-eye movements) or slower saccades, commonly observed in NREM sleep. Technical requirements encompass a sampling frequency of at least 200 Hz, and visual bandpass filter of 0.3–35 Hz. It is bilaterally recorded through

wired electrodes in the derivations E1-M2 and E2-M2 for the left and right eyes, respectively (Figure 2.4); the electrodes are placed on the outer canthi, 1 above and below the eye axis.

ECG and Pulse Oximetry The electrocardiogram evaluates the cardiac rhythm and is recorded through a single-channel, modified lead II derivation. Oxygen saturation during sleep is measured through a wired pulse oximeter and used jointly with respiratory parameters for assessing breathing-related events.

Respiration Respiratory parameters are also monitored through diverse instrumentation. First, airflow is measured through an oronasal thermal airflow sensor, and a nasal pressure transducer, for the evaluation of sleep-disordered breathing, including hypopneas or apneas. For the differentiation between obstructive, central, and mixed apneic events, additional piezoelectric sensors, revealing the respiratory effort, are required, and placed on the chest through an elastic respiratory belt.

Body Position It is an optional parameter, assessed manually or through the use of position sensors. It is particularly relevant for the evaluation of position-dependent breathing events.

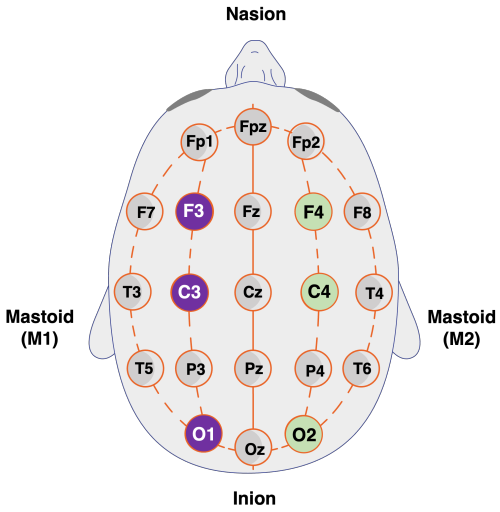
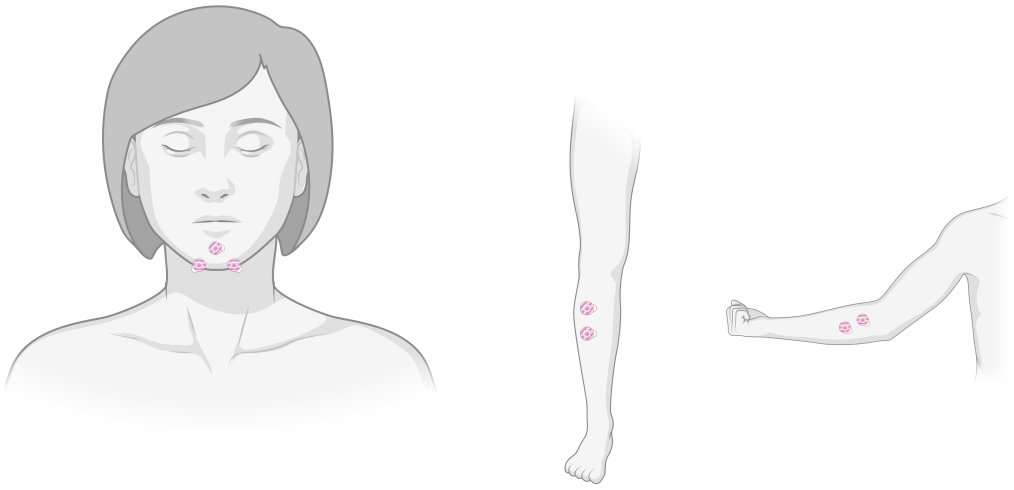


Figure 2.3: Recommended montage of the EEG channels for laboratory polysomnography. Green channels: required, Violet channels: backup.

Sleep staging is performed manually by sleep technologists or physicians. The overnight recordings are divided into windows of observation of length 30 seconds (i.e., *epochs*), and the above-mentioned biosignals are inspected simultaneously. Finally, each epoch is scored either as W, REM sleep, or a NREM sleep stage (N1,

Figure 2.4: Recommended montage of the EOG channels for laboratory polysomnography. Created with BioRender.com



(a) (b)

Figure 2.5: Recommended montage of the EMG channels for laboratory polysomnography. In detail, (a): chin channels. Lower channels are placed symmetrically under the inferior edge of the mandible. (b): Tibialis anterior channels and flexor digitorum superficialis channels. Created with BioRender.com

N2, N3), according to the specifications of each stage, provided above. Furthermore, visual inspection allows the scoring of sleep-related events, such as periodic limb movements, respiratory events, or arousals (i.e., a shift from deep to light sleep, or from sleep to wake).

Laboratory PSG is commonly scored through the use of medical-grade review softwares, which envisage custom channel visualisation for scoring sleep stages and sleep-related events, and allow raw data export for research applications.

2.1.3 The Glymphatic System

Although currently the function of sleep has not been completely understood [120], emerging evidence highlighted its role in facilitating the clearance of metabolic by-products. This process occurs in the central nervous system, through the so-called *glymphatic* system [78]. This latter consists in the perivascular pathways through which the cerebrospinal fluid (CSF) flows while filtrating toxins from the brain (Figure 2.6). This process, called the *glymphatic clearance*, is known to occur primarily during the deepest stage of sleep, while it is suppressed during quiet and active wakefulness.

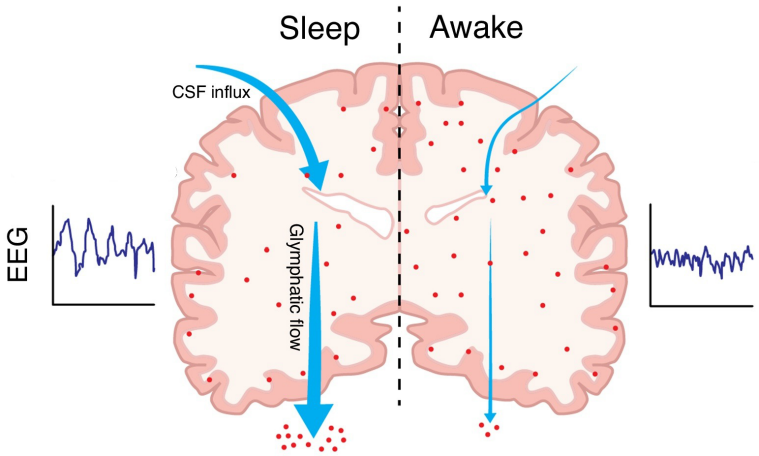


Figure 2.6: The process of glymphatic clearance, according to the type of EEG activity. Adapted from [78] with permission from Elsevier.

Recent findings suggested that glymphatic pathways are involved in the clearance of the protein Amyloid-β (Aβ) during slow-wave sleep. Aβ is a metabolic by-product that, if not properly cleared, aggregates into senile plaques and may lead to the development of neurodegenerative disorders. Indeed, gradual accumulation of Aβ is regarded as a hallmark of Alzheimer's Disease. In particular, very slow oscillations (<1Hz in the EEG) have been associated with higher rates of Aβ clearance, while increasing frequencies reflected positive correlation with the deposition of the protein [125].

Prolonged wakefulness and sleep deprivation have a negative impact on sleep quality, revealing a disruption of deep sleep, and significantly impairing glymphatic clearance. For this reason, these aspects may be identified as risk factors for neurodegeneration [102]. This concept will be further explored in Chapter 6.

2.2 Sleep Disorders

Disturbances in the physiological architecture, due to sleep fragmentation, insomnia, or environmental factors have an adverse effect on the timing, quantity, and quality of sleep. These diverse events converge on the spectrum of sleep disorders. These latter exhibit a global prevalence of over 60%, and are often accompanied by other medical conditions, mental, or mood disorders. They entail a huge variety of manifestations, and have been categorised by the International Classification of Sleep Disorders (ICSD-3) into the following [149]:

- **Insomnia**: Difficulty in falling and maintaining asleep,
- **Sleep-related breathing disorders**: Abnormal respiratory patterns during sleep, including apnea, hypoventilation, and hypoxemia [23],
- **Central Disorders of Hypersomnolence**: Excessive daytime sleepiness, tiredness and drowsiness,
- **Circadian Rhythm Sleep-Wake Disorders**: Disruption of the physiological sleep-wake cycle, including shift-work related and jet lag sleep disorders,
- **Parasomnias**: abnormal and excessive behaviours during sleep, including motor manifestations and vocalisations,
- **Sleep-related Movement Disorders**: Involuntary movements during sleep which lead to fragmented sleep architecture.

This Dissertation will focus on *Parasomnias* and *Sleep-Related Motor Disorders*. The two categories are strongly linked to neurodegenerative disorders, as concerns both the aetiology and secondary symptomatology.

2.2.1 REM Sleep Behaviour Disorder

REM Sleep Behaviour Disorder (RBD) is a REM Sleep parasomnia with a 1.15% prevalence in the aging population worldwide [38]. In resemblance to a *status dissociatus*, RBD manifests chiefly with loss of atonia of the skeletal muscles during REM sleep, accompanied by vocalisations and abrupt, and violent, behaviours. These manifestations reflect the so-called *dream enactment*. However, the majority of abnormal motor activities in RBD are identified in minor muscle jerks and increased myoclonic twitches during REM sleep, while complex motor and verbal activities due to dream enactment are more evident, but less frequent – though required for the clinical diagnosis. Indeed, this behaviour generally represents 13–31% of the recorded motor activity, with complex dream enactment entailing only

up to 1.8%. [15]. On the other hand, muscle twitches constitute the 66–83% of the observed loss of atonia [15].

The motor phenomenon is known as REM Sleep Without Atonia (RSWA), and is clinically considered the polysomnographic hallmark of RBD. Manual scoring of polysomnographic recordings overnight muscle activity allows the quantification of RSWA, which is ruled through amplitude thresholds. Chapter 5 will discuss in detail the current assessment methods.

However, besides evidence of increased muscle activity, the clinical diagnosis of RBD includes numerous criteria, and is performed through clinical interviews. Although presently no univocal method for scoring muscle activity has been identified, the symptoms criteria for the clinical diagnosis have been provided in the International Classification of Sleep Disorders (ICSD-3), and are the following:

- Repeated episodes of sleep-related vocalization and/or complex motor behaviours,
- The behaviours occur during REM sleep according to polysomnographic evidence or, based on the clinical history of dream enactment, are presumed to occur during REM sleep,
- Polysomnographic evidence of REM Sleep Without Atonia,
- Absence of epileptiform activity during REM sleep, unless RBD can be clearly distinguished from any concurrent REM sleep-related seizure disorder,
- The sleep disturbance is not better explained by another sleep disorder, medical or neurologic disorder, mental disorder, medication use, or substance use disorder.

From a physiological point of view, REM sleep atonia is regulated by the inhibition of the spinal motor neuron through the pre-coeruleus and sublaterodorsal (SLD) nuclei (Figure 2.7). Hence, the occurrence of REM Sleep Without atonia may be due to a disregulation in efferent inhibitory pathways, either as idiopathic, or due to an increased aggregation of proteins (such as α-Syn) in the nervous system [87].

When manifesting without prior medical conditions, RBD is known as *idiopathic* or *isolated* (iRBD). However, this parasomnia can appear as a co-morbidity in other pathologies, and is referred to as *symptomatic* or *secondary* RBD. The presence of secondary RBD has been prevalently associated to α-synucleinopathies, including multiple system atrophy, Lewy body dementia, and PD [119]. However, the involvement of RBD in neurodegenerative diseases is multifaceted. Indeed, longitudinal retrospective studies associated isolated RBD with a higher risk of development of α-synucleinopathies, exhibiting a phenoconversion rate to PD of up to 73.5% in a 12-year follow-up [181].

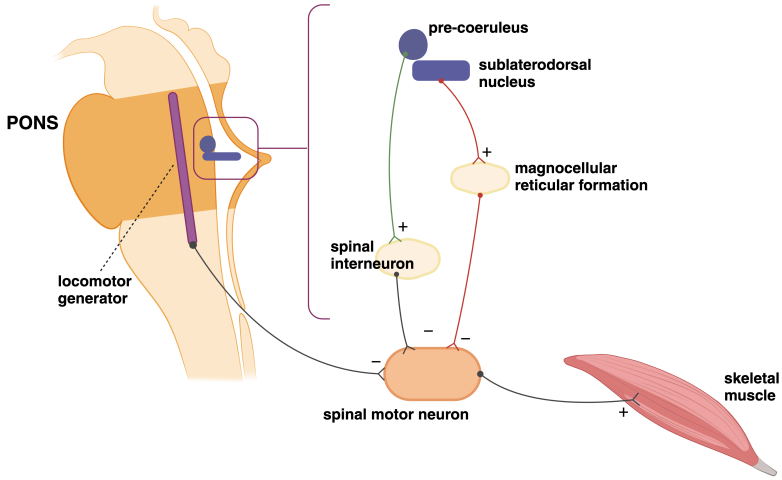


Figure 2.7: Physiology of REM atonia, originating in the brain stem. The pre-coeruleus and SLD nuclei activate the inhibitory pathways (direct and indirect routes, green and red lines, respectively). The pre-coeruleus nucleus stimulates an inhibitory spinal interneuron, and the SLD activates the MCRF to produce skeletal muscle atonia. In RBD, the SLD nucleus dysfunction, and consequent disregulation of the afferent or efferent pathways, cause the muscles to be activated during REM sleep. SLD: sublaterodorsal, MCRF: magnocellular reticular formation. Created with BioRender.com

On this premise, RBD has been accounted as among the earliest prodromes of neurodegeneration, with emerging evidence suggesting the presence of a preliminary stage – i.e., *prodromal* RBD – where REM sleep abnormalities are present, but do not meet the criteria for the actual diagnosis [82]. In this perspective, early diagnosis and the detection of prodromal markers become pivotal. Indeed, although neuro-protective treatment is still under study [45], recent findings suggested the ability of lifestyle modifications in slowing down the disease progression [57, 74].

2.3 Neurodegenerative Diseases

Neurodegenerative diseases (ND) include a spectrum of chronic disorders that affect the central nervous system and involve progressive neuronal degeneration, ultimately leading to the death of nerve cells. This condition translates to major non-traumatic disabilities, such as motor dysfunctions and cognitive decline. ND represent a dramatic and social burden, as they currently affect approximately 15% of the global population [170]. The three major conditions are Alzheimer's

Disease, Parkinson's Disease, along with increasing incidence on Amyotrophic Lateral Sclerosis. Although the aetiology of each is multifactorial, including genetic and environmental contributions, one common causal factor is age. Indeed, the presence of these diseases increases dramatically with aging, and their prevalence is expected to rise, with global demographic trends. This issue poses a significant burden on healthcare, demanding prompt intervention, as concerns disease diagnosis and prognosis, and prevention strategies. Pharmacological treatment is crucial for addressing overt symptomatology. However, a major role is held by non-pharmacological intervention, such as telehealth strageties for monitoring, physical and cognitive rehabilitation, potentially highlighting multimodal biomarkers to increase accessibility and enable an effective understanding of disease evolution for personalised care.

2.3.1 Parkinson's Disease

Parkinson's Disease (PD) is the second most common neurodegenerative disease. It affects a chronic up to 9 million people worldwide, though this number is expected to double by 2040, with the aging population worldwide. Indeed, generally disease onset appears after the age of 60, although an increasing number of subjects presents with a young onset – i.e., before the age of 50. PD exhibits an heterogeneous clinical fingerprint, entailing both motor and non-motor symptoms (MS and NMS, respectively).

The primary pathophysiological hallmark of PD is the neuronal degeneration in the substantia nigra pars compacta. Precisely, this process involves dopaminergic neurons in the basal ganglia, disregating the circuits that control movement [46]. Motor symptoms appear when about 70% of dopamine-secreting cells underwent degeneration. Secondly, abnormal aggregations of the α-Syn protein, called Lewy bodies, observed both inside and outside the substantia nigra, are a clear indication of PD [171]. These aggregates led to the definition of a six-stage system for the classification of the disease, based on the extent of the Lewy pathology [19]; this method is known as Braak staging (Figure 2.8). Early stages are primarily characterised by NMS, while most advanced states exhibit motor and cognitive manifestations. In more detail, stages I and II entail pre-motor symptoms, such as autonomic and olfactory disturbances. The first motor disorders appear in stages III and IV, which envisage also sleep disturbances. Finally, stages V and VI include emotional and cognitive impairment.

Motor symptoms are the most recognisable manifestations and include tremor at rest, increased rigidity, bradykinesia (reduced velocity of movement), and gait impairment, including freezing of gait. Stronger phenotypes may exhibit postural alterations such as camptocormia, a severe flexion of the torso, commonly observed in the sagittal and coronal planes.

NMS are less measurable, but highly common, and may precede the onset of

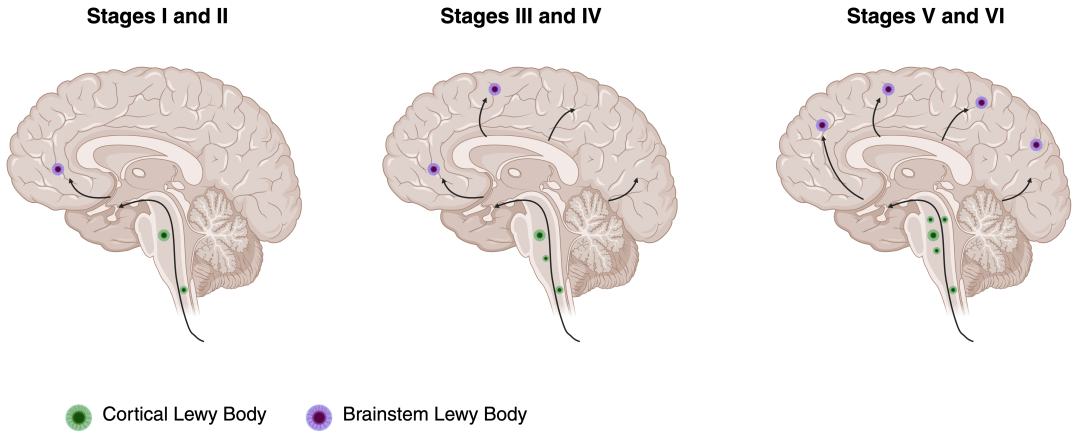


Figure 2.8: Braak Staging Method for Parkinson's Disease. Created with BioRender.com

MS. These include autonomic dysfunction, orthostatic hypotension, speech and vocal alterations, and neuropsychiatric conditions.

Although a definitive diagnosis relies on postmortem assessment of Lewy bodies in the midbrain region, from a clinical perspective, the medical assessment of PD is based on neurological examination and clinical interviews, with evidence of motor symptoms.

The Unified Parkinson's Disease Rating Scale (MDS-UPDRS) was proposed to assess the disease prognosis, by clinical observation of motor and non-motor manifestations, and approved by the Movement Disorder Society [138]. It is a useful tool to evaluate the effect of treatment and monitor the subject-specific progression of the disease.

Presently, there is no cure for PD, and pharmacological treatment aims at providing symptomatic relief. Generally, Levodopa (L-dopa) is administered to increase the level of dopamine in the brain. It is often combined with carbidopa, that helps reducing the side-effects of L-dopa therapy. In non-responders, surgical approaches may be adopted, including lesional surgery and deep brain stimulation (DBS) implantation. This latter involves brain neurostimulation in the areas responsible for motor control, such as the globus pallidus internus, the thalamus, subthalamic and pedunculopontine nuclei.

Sleep Disorders in Parkinson's Disease

Sleep disorders occur at the intersection of the MS and NMS macrocategories, and manifest both at an early stage or due to medication, affecting up to 96% of

patients with PD. They encompass insomnia, excessive daytime sleepiness, RBD, or motor disturbances during sleep. These latter include nocturnal hypokinesia and akinesia, two conditions that envisage the inability to move during sleep, causing difficulty in turning in bed, or getting up.

Nocturnal hypokinesia and akinesia were identified in up to 70% of people with PD [161, 104]. They are frequently observed in mid-stage patients, but, as mentioned previously, are present at a subclinical level even in early stages.

Impaired overnight mobility may lead to more severe co-morbidities, such as ulcers and asphyxia, and significantly affects independence in daily living activities [176], as well as contributing to a lower quality of sleep [161].

Clinical information on nocturnal hypokinesia and akinesia is collected through medical history, physical examination or subjective reports, as the MDS-UPDRS contains only one item for assessing sleep disorders. Generally, sleep diaries are employed to investigate sleep. Nevertheless, they provide fragmented information, as the patients subjectively report data related to sleep duration, and daytime naps. An alternative scale, the Parkinson's Disease Sleep Scale (PDSS) was proposed [35], to provide a more comprehensive description of the spectrum of sleep in PD. The PDSS is a visual assessment tool that includes 15 commonly reported symptoms, addressing:

- overall quality of sleep,
- sleep onset and maintenance insomnia,
- nocturnal restlessness,
- nocturnal psychosis,
- nocturia,
- nocturnal motor symptoms,
- sleep refreshment,
- daytime dozing.

The survey is completed by patients or caregivers, based on data from the previous week. Although proving as a reliable tool for the assessment of sleep, it stems from subjective reports, and is administered during outpatient visits, which are commonly scheduled on a 6- to 12-month basis. However, sleep disturbances in PD are manyfold and subjected to ultradian and circadian variations, thus demanding a precise and holistic approach in their assessment.

Hence, the attention on these conditions is often overlooked, resulting in inadequate follow-up and the lack of appropriate symptomatic treatment.

Chapter 7 will investigate sleep-related motor disturbances in PD, and discuss this topic.

2.3.2 Amyotrophic Lateral Sclerosis

Amyotrophic Lateral Sclerosis (ALS), also known as motor neuron disease, is a rapidly progressive and fatal neurodegenerative disease that entails the chronic, progressive loss of upper and lower motor neurons, responsible of voluntary muscle contractions [18].

In 95% of cases, the cause of ALS is unknown (*sporadic* ALS); the remainder are classified as genetic cases (*familial* ALS). Other classifications include the age of onset (typically 47–60 [94]), and the first-affected body part.

A key feature of the disease is found in motor disturbances. Indeed, the degeneration of motor neurons involves the gradual development of rigidity, weakness, and sporadic muscle twitches, ultimately leading to speech alterations, and respiratory failure. Due to the subtle onset of symptoms, the actual median time to diagnosis is 14 months [94]. This is reflected in therapeutic delay, with devastating consequences on psychosocial health and quality of life.

At present, no cure is available for ALS. Pharmacological strategies aim at preserving to some extent self-sufficiency and an adequate quality of life. Non-invasive ventilation, generally adopted in the later stages, proved effective in lengthening survival and improving the quality of life [58].

Although the disease is primarily characterised by motor manifestations, it presents with a diversified non-motor symptomatology, as many others neurodegenerative diseases. Specifically, sleep disruptions are frequently observed in ALS, and have a dramatic negative impact on overall health [58]. These include increased sleep latency, reduced slow-wave sleep, fragmented sleep, and REM parasomnias. REM Sleep Without Atonia, with or without a diagnosis of RBD, has been observed, although its impact on disease-related disability has not been investigated [106]. As emerging evidence suggested the impact of muscular parameters on the rapidity of disease progression [148], these parasomnias shall be investigated, to provide a comprehensive approach in the clinical assessment of the disease. Chapter 8 will explore this concept, and study the correlation between the extent of RSWA and disease severity.

Chapter 3

Machine Learning for Healthcare

In recent years, Machine Learning (ML) methods have drawn remarkable attention in healthcare, due to technological advancements in the clinical practice, and the increasing spread of Telehealth or eHealth applications.

This Chapter provides an overview of the ML models employed for the research activities presented in this Thesis, by describing their rationale, behaviour, and interpretation. For the sake of clarity, in the context of this Thesis, only shallow learning algorithms were employed, due to the moderately limited number of samples available in the employed datasets.

3.1 Background

Artificial Intelligence (AI) methods have rapidly evolved for years. In a scientific perspective, AI represents all technological applications that offer a simulation of human intelligence, through computer-based systems. ML is a subfield of AI, encompassing all methods that enable machines to learn from and echo human behaviour [2]. This delineates a multifactorial and diversified scenario of possible applications. Indeed, ML allows *data interpretation*, i.e., describing the relations between items in a dataset, *data prediction*, i.e., making predictions on the basis of current data, and *decision support*, meaning that the computer-based system may provide suggestions about future actions [21].

Hence, ML-based approaches find suitable application in healthcare frameworks. Indeed, these Tele- and eHealth seek to enhance medical care by facilitating patient involvement and accessibility, and AI offers encouraging benefits with regard to diagnosis support, treatment optimisation, and personalised health outcomes.

ML models are classified into three categories, depending on their behaviour. Namely, they are *Supervised Learning*, *Unsupervised Learning*, and *Reinforcement Learning*. Generally, the type of approach adopted depends on the study design and desired final outcome. In more detail, supervised models are trained from

previously labeled datasets, enabling the models to learn from known sets of data and later identify unlabelled data. Conversely, unsupervised approaches are trained on unknown data, and aim at identifying underlying data patterns and trends. Finally, reinforcement learning involves a heuristic approach of trial-and-error for training the models, and adopts a feedback reward system for optimising decision-making.

The research activities presented in this Dissertation adopted supervised machine learning approaches, as they were deemed more suitable for multidisciplinary data interpretation, a key feature for technological inclusion in the clinical practice. The following Sections provide an overview of the employed models, and present the adopted validation procedures.

3.2 Overview of the Employed Models

Supervised Learning models can be further divided according to their structure and learning strategies. Precisely, they can be linear, hierarchical, parametric or distance-based, and an ensemble of other structures. The research works included in this Thesis exploited various categories, to compare their performance and understand their suitability to the problem, and are described in the paragraphs below.

Support Vector Machine (SVM) This classifier is based on an algorithm that aims at finding the optimal hyperplane to separate classes in high-dimensional feature spaces, even with non-linearly separable data [44]. To do so, this classifier maximises the margin between data points (i.e., support vectors) by transforming data through a *kernel function*. This latter can be a linear or non-linear function. A linear SVM is recommended when the separation of data follows a linear trend.

K-Nearest Neighbour (KNN) This model is based on non-parametric classification algorithms. It categorises each data point based on its similarity to the K closest data points in the dataset. These latter are defined as *neighbours*. The choice of K is strongly dependent on the composition of the input data, and affects the performance and robustness of the models. In addition, the *distance* between the query point and its neighbours is a key parameter. It can rely on various distance metrics, such as the Euclidean, Manhattan and their generalisations [70]. A KNN is a simple algorithm that finds suitable applications with irregular data patterns; however, it is particularly affected by the curse of dimensionality, and generally features lower performance on high-dimensional dataset, resulting in overfitting.

Decision Trees (DT) A Decision Tree is a non-parametric hierarchical model, based on a tree-like structure. It stems from a *root* node, and follows with *internal* (or *decision*) node, and finally features *leaf* nodes, that represent the outcomes in

the dataset. It finds the optimal split point by testing observations at each internal node, through a set of decision rules based on a greedy search approach. The process is repeated in a top-down fashion for all observations in the dataset. The most popular splitting criteria are information gain (through entropy evaluations) or the Gini index [20].

Ensemble Methods Ensemble methods are variants of the initial DT model. They consist of a set of DT learners and provide a decision output based on the combination of predictions in the elemental learners. They are further subcategorised into:

- **Bagging**, an approach that aggregates predictions from multiple DTs trained parallelly on *bootstrap* portions of the initial dataset. A bootstrap sample is constructed with data sampled from the training set, with replacement. Each DT is trained on a subportion of the initial data. A typical example of this learning method is found in Random Forests (RF).
- **Boosting** Boosting, and its variant Gradient Boosting, work by combining a set of weak learners into a strong learner, to minimise the training error. With respect to Bagging methods, this category of classifiers trains the weak learners sequentially, gradually improving the training accuracy of the models.

Finally, other two approaches were explored, and are summarised in the following paragraphs.

Naïve-Bayes Classifiers (NB) They are a class of probabilistic classifiers that adopt the Bayes' theorem to yield predictions. For each query point, the posterior probability is computed through the Bayes' theorem; finally, the class output corresponds to the maximum posterior probability in a group of classes.

Linear Discriminant Analysis (LDA) This is an approach that aims at identifying the linear combination of features that best separates classes in the dataset. It finds its application both in binary dataset and multi-class problems. It classifies items in the dataset by modelling their distribution through a Bayesian approach. Specifically, the algorithm computes the probability of a query point of belonging to a particular class. High-dimensional data are projected into one dimension, for better classification performance.

3.3 Model Evaluation

Model optimisation, validation, and evaluation are three essential procedures for assessing the performance, generalisation capability, and overall robustness of

the classifiers. The following Sections will briefly describe these points.

3.3.1 Hyperparameters Optimisation

Hyperparameters tuning is a technique that allows to identify the optimal set of model's *parameters*, to enhance classification performance. The parameters are the features that help the model in conducting the learning process; hence, they are not learned during the training process, and need to be specified before. Generally, hyperparameters optimisation is conducted through automatic approaches, such as a Grid Search or Bayesian optimisation.

3.3.2 Validation Procedures

Validation procedures are commonly adopted after the training procedure of the ML models, to assess their performance and investigate their suitability to the study design. The following paragraph present three commonly adopted validation approaches, employed in the subsequent research activities.

Hold-Out Validation This method involves partitioning the initial dataset into two parts: the *training* and the *test* sets. The former is employed during model training, and the latter is used as unknown set of data to carry out predictions, and evaluate the model performance. Typically, the training set consists of a higher portion of data (70–80% of the initial dataset). Commonly employed splits set-ups are 70/30 % or 80/20 % (Train/Test configuration).

K-Fold Cross-Validation (CV) As the common hold-out method may lead to high variance, a more reliable solution is provided by the K-Fold CV. In this technique, the initial dataset is divided into K parts (*folds*), of similar size. The training process is carried out on $K-1$ subsets, and predictions are made on the remaining fold. The procedure is iterated K times, thus providing a more accurate assessment of model performance.

Leave-One-Out CV (LOO-CV) This method is a particular configuration of K-Fold CV, where the parameter K equals the number of observations in the dataset. Hence, the training/test procedure is repeated for all instances in the dataset. Similarly as traditional K-fold CV, this method is a reliable solution for small datasets. However, it is more computationally expensive. In healthcare applications, a variation of this method is proposed, through the Leave-One-Subject-Out CV (LOSO-CV). Within this framework, given a dataset with N subjects, at each iteration, all observations from one subject are held out for testing, and the remaining $(N-1)$ subjects are used in the training process. The procedure is repeated for a total of N iterations. It allows to maximise the classification accuracy by

limiting the effect of overfitting. This latter occurs when the trained model is too complex with respect to the data patterns, thus fitting too closely with the elements in dataset, resulting in inaccurate predictions.

3.3.3 Performance Metrics

Finally, to quantitatively assess the performance of a model, and evaluate its generalisation capability, performance metrics are computed from the obtained predictions. Commonly, this assessment relies on the definition of *confusion matrix*, where the rows represent the instances in the actual class, and the columns represent the class-wise predictions (Figure 3.1). Precisely, correct predictions are identified either as True Positives (TPs) or True Negatives (TNs), while misclassified instances are False Positives (FPs) and False Negatives (FNs).

The most common metrics in supervised ML are Accuracy, Recall, Specificity, Precision, and F1 score, defined as:

$$Accuracy = \frac{(TP + TN)}{(FP + FN)} \tag{3.1}$$

$$Recall = \frac{(TP)}{(TP + FN)} \tag{3.2}$$

$$Specificity = \frac{(TN)}{(FP + TN)} \tag{3.3}$$

$$Precision = \frac{(FP)}{(TP + FP)} \tag{3.4}$$

$$F1score = \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{3.5}$$

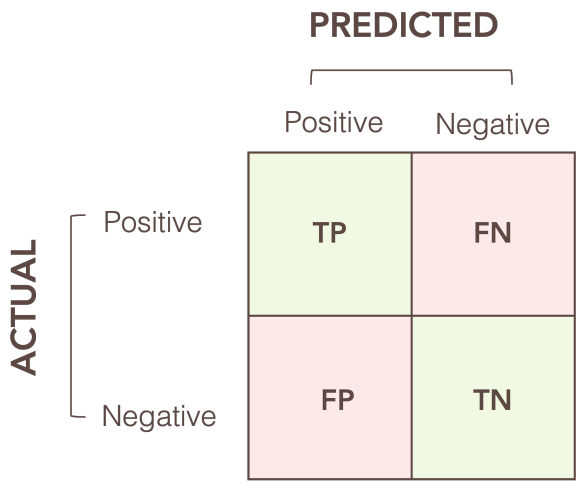


Figure 3.1: Confusion matrix for the evaluation of Machine Learning models.

In addition, metrics such as the Area Under the Curve (AUC) and Mean Average Error (MAE) can be employed. The former represents the overall performance of the model, computed as the integral of the Receiver Operating Characteristic (ROC) curve. This latter evaluates the performance of a model by plotting the Recall as function of the False Positive Rate (i.e., $1 - Specificity$). Values of AUC approaching 1 represent optimal performance.

The MAE, generally employed in continuous tasks, measures the variance between the actual (x) and predicted values (y), and is expressed as:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{3.6}$$

These metrics provide a reliable tool for the evaluation of models, and the interpretation of the predictions. As previously said, accurate model assessment allows for a robust implementation of such methods in healthcare scenarios.

Part I

Diagnostic Support Systems

Chapter 4

Automatic Sleep Staging

4.1 Context and Background

Polysomnography, a diagnostic test encompassing the study of electrophysiological activity during sleep, is the gold standard to investigate sleep structure and sleep disorders. Despite being accurate, it is often considered as invasive by the subjects, and its clinical validity depends on manual inspection, thus requiring protracted scoring times.

In view of developing minimally intrusive sleep studies, a pipeline for automatic sleep staging based on a single EEG channel has been proposed.

This Chapter illustrates the rationale behind this idea and the research activity included in Paper [141].

4.1.1 Machine Learning Approaches to Sleep Staging

The concept of *sleep staging* refers to the process of labelling sleep into different categories, relying on precise sets of rules in order to characterise polysomnographic activity over pre-defined period of times, typically referred to as *epochs* [79].

As depicted in Section 2.1, traditionally, these rules have been collected in the Rechtschaffen and Kales (R&K) criteria for sleep scoring [89], and later updated in the AASM Scoring Manual [11]. However, in the clinical practice, this translates to a manual and visual process, that is labour-intensive, time-consuming, and subjective, presenting high rates of inter- and intra-rater variability, significantly limiting the diagnosis' fluidity [179].

In recent years, there has been growing interest in developing automated methods for sleep staging, either by tackling sleep/wake detection [184], stage-specific [173, 177], or multi-stage classification [42, 84, 144]. These approaches leveraged PSG signals and Machine Learning environments to expedite the scoring process, achieving good agreement between the automated results and the manual annotations from sleep experts [42].

Though providing accurate and reliable results, a vast majority of the automated frameworks in the literature still carry a considerable amount of complexity. Indeed, some studies relied on several biosignals to achieve multi-stage classification [42], while other approaches encompassed the analysis of multiple EEG channels [84].

In this regard, automatic classification has been carried out both through feature-based approaches, and, more recently, through AI-derived probabilistic approaches [6, 163] or Deep Learning techniques [178]. These latter, in particular, do not require any signal processing or feature engineering, and appeared to provide reliable results, through spatial and temporal architectures [85, 59, 118, 182], or attention-based models [183, 53], though providing lower clinical interpretability.

Although providing an alternative to manual scoring, offering the potential to simplify the scoring process, while ensuring reliable results, these approaches still require multiple PSG signals, thus remaining impractical for use outside of clinical settings.

By contrast, efforts have been made to develop lightweight sleep monitoring systems, offer the potential for convenient and accessible sleep monitoring in home environments [97, 90], though most of them target respiratory events [145]. These frameworks involve wearable set-ups equipped with sensors, to record PSG biosignals (EEG, EOG, EMG), or photoplethysmography to retrieve heart rate during sleep, and tackle sleep staging [64].

Following these developments, sleep staging algorithms based on single-channel EEG have been proposed in the literature. These approaches allow for a reduction in the required instrumentation and hardware, having a significant impact on the diagnostic intrusiveness.

In this context, single-channel EEG-based sleep studies need to be explored, to investigate **(1)** the reliability of scoring sleep through a single EEG electrode, and **(2)** the feasibility of lightweight, home-based sleep monitoring devices.

4.2 Research Overview

In consideration of the presented context, the research activity included in this Chapter explored the feasibility of automatic scoring of sleep from single-channel raw EEG signals and its applicability to patients at risk for neurodegenerative diseases, exploiting supervised ML approaches.

More specifically, the automatic detection pipeline, detiled in the following Sections, consisted in:

1. Characterising REM sleep through specific predictors,
2. Performing 5-stage automatic classification of sleep based on a a single EEG channel,
3. Validating the model on subjects at risk for neurodegeneration.

4.3 Single-Channel EEG Sleep Staging

Manual annotation of sleep stages from full, in-hospital PSG relies on a fixed set of rules, based on the characteristics of the recorded biosignals, in terms of waveform and type of eye-movements.

The REM stage is commonly recognised by the presence of muscle atonia and the typical rapid-eye movements, observed with chin EMG, and the EOG channels, respectively. Though commonly treated as homogeneous, from an encephalographic point of view, the REM stage actually presents with mixed characteristics: generally low amplitude, and the alternation of two main frequency components (phasic and tonic) [159].

This activity focused primarily on the characterisation of the REM Sleep microstructure through specific metrics, employed in the automatic classification of sleep stages to enhance detection performance.

4.3.1 Materials

Subjects

The automated pipeline was designed on two publicly available datasets, employed in the design and training of the algorithms, and in the external validation of the models, respectively.

To train the models, the DREAMS Subjects Database [47] was employed. This dataset includes a collection of 20 PSG recordings of healthy subjects (4 males), with no sleep disorders nor underlying neurological conditions. The age distribution of the whole sample is 33.5 ± 14 years (20–65).

Annotations were provided according to the AASM criteria [11]; PSG recordings presented with a mean duration of 8 hours, 30 minutes, and were analysed in an epoch-wise fashion, with a standard epoch length of 30 seconds.

In consideration of the hypnographic variability of the recordings, a preliminary screening was conducted on the dataset, so as to include in the training models subjects presenting with an adequate duration of sleep stages, to limit class imbalance.

Following this concept, only recordings presenting with at least 50-minute REM were included in the analysis; this duration threshold was discussed with the clinical personnel. Finally, 10 subjects were discarded, either (*a*) due to the complete absence of REM sleep or SWS epochs, for which a possible explanation might be found in the first night effect, a common phenomenon in in-hospital PSG [50], or (*b*) the presence of irregular sleep cycles, likely not reflecting a healthy pattern. Table 4.1 reports an overview of the demographics of the included sample.

The final training configuration envisaged a total of 8382 epochs, and the stage distribution is displayed in the plot in Figure 4.1. As appreciable, a considerable

imbalance is present. This is a common issue in automatic sleep studies, as the N2 stage accounts for 45–55% of the total duration of sleep, and N1, due to its transitional nature, commonly features a very low number of epochs.

Figure 4.1: Distribution of sleep stages, in terms of number of 30-second epochs, in the training dataset, encompassing 10 healthy subjects.

In the final stage of this work, a second database was included, to validate the trained models on a set of subjects at possible risk for neurodegeneration, to explore the applicability of the presented framework.

This second batch included subjects from the CAP Sleep Database [166]. This dataset included 16 healthy subjects (HS), and 22 subjects with RBD, though considerably older in age with respect to the other group. Demographics for the two employed datasets are reported in Table 4.1. Being publicly available datasets, and due to the general scarcity of data, no preparatory age-matching was performed; however, the healthy cohort in the CAP Sleep Database features an average value in line with the subjects in the DREAMS Database. The higher variance observed in this latter is due to the fact that a number of elder subjects, compared to the healthy CAP group, are present; however, these represent only 20% of the whole group. No preparatory data screening was conducted on this cohort, as it only served the purpose of external validation, in an attempt to simulate a real-world scenario. The dataset consisted of a total of 14583 epochs, with over 5000 representing N2.

Data

As stated in the research objectives, the detection algorithm was based on data from a single EEG channel. In the DREAMS Database, three channels were available: a central (either CZ or C3), a pre-frontal (Fp1), and an occipital.

For the training stage, only the central channel (Cz, or C3 if the former was not available) was selected. Indeed, pre-frontal channels, besides being uncommon in

Table 4.1: Sample and demographics of the datasets employed in the study: the DREAMS Database (healthy subjects) and the CAP Sleep Database (healthy and RBD subjects).

Dataset	Sample	Age
DREAMS Subjects Database	10 HS (3 males)	36.4 ± 14.9 years
CAP Sleep Database	22 RBD (19 males)	70 ± 6 years
	16 HS (9 males)	32.5 ± 5 years

PSG, are also often affected by eye-blinking artefacts. Similarly, in the perspective of implementing a wearable and unsupervised set-ups in sleep studies, occipital channels likely suffer from poor sensor positioning.

4.3.2 Methods

Data Pre-Processing

As the EEG signals in the DREAMS Database were recorded at a sampling frequency of 200 Hz, they were first upsampled to 256 Hz. Despite knowing that upsampling could not compensate for missing information, it was carried out only to computationally ease the subsequent analysis, especially in the spectral domain.

To preserve the idea of raw single-channel analysis, minimal processing was carried out on the data. No artefact removal was carried out, and, given that a single EEG channel was employed, no spatial filtering on EEG data was performed.

In an attempt to mimic manual scoring, which employs review softwares with customisable signal display through *visual* filters, a similar filtering set-up was adopted. Specifically, high-frequency noise and slow drifts were removed, through a cascade of infinite-impulse-response (IIR) filters, detailed below. An anti-causal filtering approach was chosen, to avoid phase-distortion and preserve the waveform actual locations.

Slow drifts in the EEG signal were attenuated through a highpass, Chebyshev Type I filter, of order 1 and cut-off frequency (f_c) of 0.5 Hz. High-frequency noise, identified for this purpose in the spectral components above 40 Hz, was removed through a low-pass, Chebyshev Type I filter, of order 11 (f_c: 40 Hz).

Chebyshev Type I filters [132] present with a steep roll-off, thus offering a quite narrow transition band, though showing equiripple behaviour in the passband. The ripple factor (ϵ) is expressed as:

$$\epsilon = \sqrt{10^{(\delta/10)} - 1} \tag{4.1}$$

where δ is the ripple in the passband, expressed in dB. For this work, a passband ripple of 3 dB peak-to-peak was applied.

Identification of the REM Sleep Micro-Structure

The REM stage presents with a peculiar, underlying, structure. From the examination of REM segments, in terms of morphology and power spectral density, two main components emerge (cf. Section 2.1). These are identified as the *tonic* REM (TREM), with reduced EOG activity and no rapid-eye movements, and *phasic* REM (FREM), where rapid-eye movements are present, instead.

Several feature-based, automatic sleep staging algorithms target the REM stage as a whole, by leveraging linear and non-linear measures to characterise the overall signal morphology or its energy contributions [42]. However, given the mixed characteristics of REM sleep, exhibiting low-amplitude and multiple frequency components, this can lead to misdetection with wake or N1 segments.

Other approaches attempted to exploit the multiple frequency components of EEG during sleep, by identifying additional sub-bands [99, 84].

Following this strategy, and based on the intrinsic characteristics of TREM and FREM, this work aimed at identifying features for the precise characterisation of the two REM micro-states, in an attempt to enhance the detection performance of the overall stage.

Particularly, as previously highlighted in the literature [159], the tonic and phasic REM states present with distinctive traits. An increase in higher frequencies (α and low-β range) is observed during TREM, while FREM exhibits a higher density of slower waves, in particular in the δ and θ bands.

On this premise, in this study two frequency bands were experimentally defined for the subsequent feature extraction step on the REM stage micro-structure, and named accordingly, as TREM and FREM.

For this purpose, specific frequency boundaries need to be identified, in order to estimate the exact range for the two bands. Hence, the original REM stage was inspected in this way. Each manually scored REM epoch was selected individually, and the power spectral density (PSD) was estimated. From the obtained power spectrum, two distinct values were computed for each epoch: the median frequency, and the 95^{th} percentile.

For the sake of clarity, these two values will be referred to as SEF50 and SEF95, where SEF indicates the *spectral edge frequency*. In more detail, these represent the boundaries below which lie the 50% and 95% of the total EEG power, respectively.

Lastly, the bandwidths for each sub-state were estimated by averaging the 25^{th} and 75^{th} percentiles of the two SEFs, with the former yielding the lower and upper bounds for the FREM, and the latter the frequency bounds for TREM (Figure 4.2).

Thus, the two bands were identified as:

- **FREM**: frequency range 2–8 Hz,
- **TREM**: frequency range 7–16 Hz.

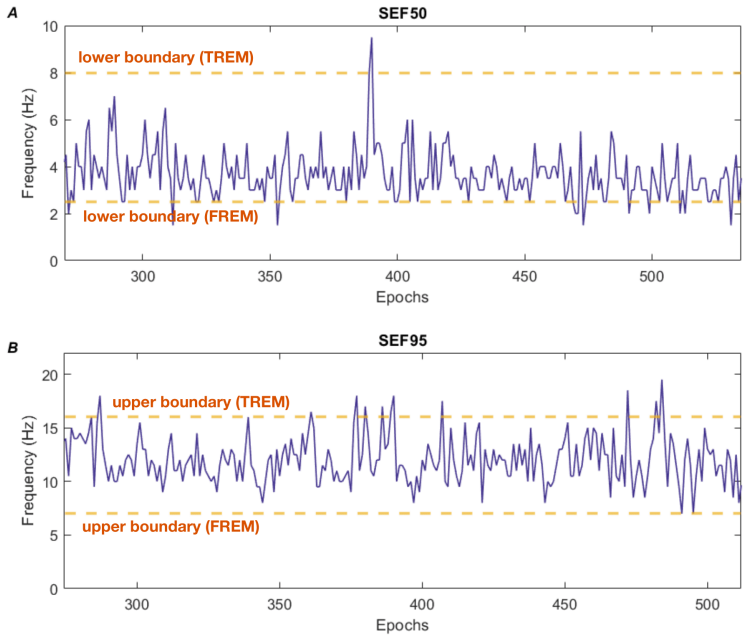


Figure 4.2: Estimation of the FREM and TREM bandwidths from the spectral edge frequency metrics. Adapted from Paper [141].

Similar boundary values for the two micro-states were identified previously in [84].

As observable, while partially overlapping at 7 Hz, the FREM sub-state covers the δ and θ ranges, while, as expected, TREM is skewed towards the higher α waves or sensorimotor rhythm (up to 15.5 Hz).

Feature Extraction

To carry out automatic sleep staging, a set of features were extracted, to characterise the EEG signal in each sleep stage.

Each sleep stage was inspected epoch-wise, and features were computed in four different domains: time, frequency, time-frequency and non-linear. This is a common approach in feature engineering from biosignals, and has also been adopted in various research activities presented in the following Chapters of this Thesis.

First, each sleep stage was assigned a class label. Then, for each label, a series of 30-second epochs was identified, and features in the time domain were computed from each epoch. For the remaining categories, the features were extracted on 1-second mini-epochs, and then averaged across each 30-second macro-epoch, following a uniform segmentation approach, widely adopted in sleep EEG pre-processing [122].

Time-domain features are computer over longer epochs, as within this length, it is possible to assume that the signal remains approximated as stationary [122]. Conversely, shorter epoch lengths are required to guarantee wide-sense stationarity when computing spectral and non-linear parameters from EEG signals. As mentioned, automatic sleep stage classification, epoch lengths of 2, 5, or 10 seconds are generally adopted [42, 84, 122]. In this study, a resolution of 1 Hz was chosen to provide for an adequate spectral representation of the analysed signal.

The hand-engineered features aimed at characterising the EEG signal in the different stages, to mimic visual inspection. The parameters were chosen so as to primarily describe the amplitude and morphology of the signal in each epoch, and, for a finer representation, the spectral distribution and regularity of the waveforms. To do so, a combination of clinical parameters, novel, and well-established features was employed; the complete list is displayed in Table 4.2. Various among the extracted parameters are traditionally employed in EEG signal processing, while others were adapted from other domains. In particular, the *Coastline* is an amplitude measure which relies on the derivative of the signal, and highlights the extent of fluctuations in a signal [177]. Let $x(t)$ be a continuous EEG epoch, and $\dot{x}(t)$ its derivative. The coastline factor is expressed as:

$$Coastline = \sum |\dot{x}(t) - \dot{x}(t-1)| \tag{4.2}$$

From a spectral point of view, and as previously discussed, this study proposed two bandwidths to accurately describe the dichotomy in the REM stage. To this end, the absolute and relative power, and the energy density in the FREM and TREM sub-states were computed and employed as features in the frequency domain. In addition to those, a differential metric for the power density (SEFd, difference between SEF95 and SEF50), was also computed, as seen previously in [84].

The feature extraction step also envisaged time-frequency and non-linear measures.

Regarding the former, the Short Time Fourier Transform (STFT) was exploited to retrieve finer details of the frequency contributions overtime; this becomes of particular importance in non-stationary signals such as the EEG.

This type of transform provides a two-dimensional representation of the signal, highlighting the frequency distribution at each time istant. In its continuous form, for an epoch $x(t)$, the STFT is defined as:

$$STFT_{x(t)}(\tau, \omega) = \int_{-\infty}^{\infty} x(t) w(t-\tau) e^{-i\omega t}\, dt \tag{4.3}$$

where $w(\tau)$ is the window function, analogous as the one employed in spectral density estimation. Fundamentally, the STFT can be interpreted as the Fourier transform of the compound function $x(t)w(t-\tau)$, which represent the phase and magnitude of $x(t)$ over time.

For this study, the STFT was estimated on 1-second epochs on the whole available spectrum (0–40 Hz), and its magnitude and maximum value were computed.

Among Time-Frequency parameters, the Discrete Wavelet Transform (DWT) coefficients were computed and employed as features for the characterisation of sleep stages. A comprehensive description of the DWT is provided in Chapter 5.

Finally, the Teager-Kaiser energy operator (TKEO) [88] was employed as non-linear source for the features. This metric provides information about the instantaneous energy of the signal, exploiting anew its amplitude and frequency. For an epoch $x(t)$, it is expressed as:

$$\psi[x(t)] = [\dot{x}(t)]^2 - x(t)\ddot{x}(t) \tag{4.4}$$

where $\dot{x}(t)$ and $\ddot{x}(t)$ represent the first and second order derivatives, respectively.

The TKEO was evaluated on the whole spectrum, and numerical and statistical metrics were extracted and employed as features for the subsequent classification step. Specifically, after computing the TKEO on each 1-second sub-epoch, the values were concatenated, and for each macro-epoch (i.e., 30-second length) the mean value, standard deviation, skewness, kurtosis, amplitude range, maximum and minimum values were retrieved.

In the Non-Linear category, the approximate entropy (ApEn) was also computed. It is a statistical measure that represent the extent of regularity of a given signal [130], and is employed in EEG processing to measure complexity [1]. It assesses the logarithmic probability that patterns observed over a given epoch length will remain stable as the epoch length increases. Given a time series of N data points (length of the EEG segment), ApEn is expressed as [130]:

$$ApEn(m, r, N) = \phi^m(r) - \phi^{m+1}(r) \tag{4.5}$$

where m is the embedding dimension, which can be identified as the length of a window, and r is the tolerance parameter, that is the maximum dissimilarity allowed between two sets of data, expressed through distance-based metrics. In more detail, it is a positive real number that represents the radius of similarity. Finally, $\phi^m(r)$ represents the average natural logarithm of the conditional probability that two runs of m data points will remain similar at $m + 1$ [130]. Low values of approximate entropy reflect higher regularity of a signal, likely underlying repetitive patterns.

Feature Scaling and Selection

As described, the extracted features belong to very diverse domains, and therefore present with different scales. This intrinsic variance may lead either to misclassification, or potentially erroneous performances.

To minimise this effect, all extracted features were normalised through the *min-max* scaling technique. This method consists in rescaling the original feature array to a defined target range, so as to provide uniformity in data distribution.

Table 4.2: Features adopted for the automatic sleep staging, along with their description and proper reference. ⋄: variables adapted from the cited study, †: variables first proposed in this work.

Category	**Features**	**Ref.**
Time	Numerical and statistical measures (mean, standard deviation, skewness, kurtosis, range, max, min)	various
	Hjorth parameters (signal and its derivative)	[123]
	Zero crossing rate	[165]
	25^{th}, 75th^{th}, 95th^{th} percentile and their differential	various
	Envelope: number of peaks, peak prominence, peak width	†
	Coastline (1^{st} and 2^{nd} derivative)	[177]
Frequency	Power percentage for each clinically relevant band	various
	SEF50, SEF95, SEFd, Absolute and Relative Power (TREM)	⋄[84]
	SEF50, SEF95, SEFd, Absolute and Relative Power (FREM)	†
	Entropy	various
	Fast Fourier Transform: numerical and statistical measures	various
	Relative power for each clinically relevant band (δ, θ, α, β)	various
	Energy density in tonic and phasic REM	†
Time-Frequency	Short Time Fourier Transform: magnitude and maximum value of its density (0 – 40 Hz)	†
	Discrete Wavelet Transform coefficients: Daubechies order 4 and Haar filter wavelet	[177]
Non-Linear	Teager-Kaiser Energy operator: numerical and statistical measures	⋄[99]
	Approximate entropy	⋄[1]

In particular, given a feature f, and a target range of $[a, b]$, the rescaling follows this definition:

$$f' = a + \frac{(f - min(f))(b - a)}{max(f) - min(f)} \tag{4.6}$$

For this analysis, the range was set to $[-1, 1]$. A total of 164 features across all categories resulted from the feature extraction step, for 8382 observations – i.e., epochs. Since the subsequent classification step envisaged the use of diverse classifiers, a feature selection step was implemented, to limit overfitting, and to highlight a subset of most relevant features for the task, to allow for a final clinical interpretation.

For this purpose, the variance threshold method was adopted [154]. Despite being a simple, straightforward technique, it requires low computational load, and might be scalable in tiny-ML frameworks.

The variance threshold was heuristically chosen; finally, all features with variance below 0.2 were removed from the initial set, leaving a total of 87 features to be implemented for automatic sleep stage classification.

Automatic Sleep Staging Classification

Automatic sleep staging was tackled through supervised models, in a multiclass fashion. Training data included the three stages of NREM sleep (N1, N2, N3), REM sleep, and wake epochs (W).

Two classifiers were explored with a 10-fold CV framework, and an ensemble method was employed in a hold-out validation fashion, with a 80/20 proportion.

Namely, a KNN and a RF were chosen as classifiers, along with a boosting algorithm based on random under sampling (RUSBoost), particularly suitable for imbalanced datasets, as in this case [156]. For this latter, at each iteration of the algorithm, the majority class is randomly undersampled so as to balance the overall class distribution (Algorithm 4.3.2).

The final configuration of the parameters of the classifiers is displayed in Table 4.3. All parameters were heuristically selected, with a trial-and-error approach, and the configuration yielding the lowest classification error was adopted.

Table 4.3: Parameters of the classifiers employed for the analysis.

Classifier	**Parameters**	**Validation**
KNN	*Number of Neighbours* (K): 10 *Distance Metric*: Euclidean	10-fold CV
RF	$N_{learners}$: 30 Maximum N_{nodes}: 0.2*$N_{features}$	10-fold CV
RUSBoost	$N_{learners}$: 30	80/20 hold-out validation

Algorithm 1 RUSBoost: Random Under-Sampling Boosting

1: **Input:** Given a training dataset S, number of iterations T
2: **Output:** Ensemble classifier hypothesis $H(x)$
3: Create a temporary dataset S' by randomly undersampling the majority class to obtain the desired distribution (D'_t)
4: Initialise weights for the new dataset S'
5: **for** $t = 1$ to T **do**
6: Call a weak learner on the weighted dataset S', and obtain the weak hypothesis h_t
7: Calculate the pseudo-loss parameter (ϵ_t)
8: Calculate the weight parameter α_t as ($\frac{1-\epsilon_t}{\epsilon_t}$)
9: Update and normalise weights
10: **end for**
11: Return the final hypothesis $H(x)$ as weighted vote of the weak hypotheses

4.3.3 Results

This Section illustrates the results of the proposed automatic sleep staging framework, in terms of validation of the employed models and external validation on an unseen dataset.

The results will be presented as *macro-averaged* and *micro-averaged* metric. In multiclass classification tasks, the former depicts the unweighted, class-wise average performance. On the other hand, micro-averaging targets individual classes by giving equal weight to each instance in the dataset, and finally the class-wise performance across all predictions. This method is particularly suitable for imbalanced datasets.

Classification Performance: 5-stage

First, a 5-stage configuration was explored, through the presented classifiers.

As regards the Random Forest and KNN classifier, a 10-fold cross-validation approach was adopted; Tables 4.4 and 4.5 report the achieved performance. The metrics were evaluated for each class according to the AASM guidelines – i.e., N1, N2, N3, REM and wake. The results are further detailed in Figures 4.3 and 4.4, that provide class-wise performance through the summative confusion matrices.

As appreciable from the illustrated metrics, detection performance of each stage is fairly good, with the RF achieving a macro-averaged accuracy across the five stages of 93.85%, and of 93.24% for the KNN. Satisfactory results were also observed in both classifiers in terms of F1 score, with values above 78%, except for the N1 stage.

This latter featured quite unsatisfactory values both in terms of Recall and F1 score. A possible explanation was found in its nature; indeed, the N1 stage is largely considered as a transition stage between a wake state and N2, a peculiarity that 30-second EEG epochs probably fail to describe. Additionally, the poor performance observed on this class may be due to the high class imbalance: the epochs belonging to stage N1 represent only 6% of the total dataset. A recent commentary on polysomnography assessment methods in adults stated that stage N1 suffers from the lowest detection reliability as compared to other stages, as it is generally characterised through exclusion rules, rather than by clearly recognizable electrophysiological hallmarks [80].

As regards the REM stage, quite satisfactory metrics were attained, across all the explored performance metrics. Recall values above 85% were observed, with F1 score above 79% in both classifiers.

In addition, multi-stage performance was assessed through Cohen's Kappa (κ) [41], a statistical metrics that provides a measure of the agreement between the manually annotated value and the one predicted by the algorithm. The RF classifier presented a κ of 0.846, and the KNN 0.831; both values suggest substantial agreement between the real and predicted scores.

Table 4.4: Performance (10-fold cross-validation) of the Random Forest classifier, in the 5-stage configuration

	W	**REM**	**N1**	**N2**	**N3**
Accuracy	97.5 %	93.9 %	92.9 %	89.5 %	95.3 %
Recall	90 %	91 %	18.8 %	90.1 %	86 %
Specificity	98.5 %	94.6 %	98.2 %	88.9 %	97.7 %
Precision	88.4 %	76.2 %	43.6 %	87 %	90.8 %
F1	89.2 %	82.9 %	26.3 %	88.6 %	88.3 %

Table 4.5: Performance (10-fold cross-validation) of the KNN classifier, in the 5-stage configuration

	W	**REM**	**N1**	**N2**	**N3**
Accuracy	97.6 %	92.7 %	92 %	88.5 %	95.2 %
Recall	88.9 %	85 %	19.3 %	88.5 %	87.9 %
Specificity	98.7 %	94.2 %	97.3 %	88.6 %	97.1 %
Precision	90.1 %	73.9 %	35.2 %	86.4 %	88.8 %
F1	89.6 %	79.1 %	24.9 %	87.4 %	88.4 %

The Random-Under-Sampling Boosted trees envisaged, on the other hand, a 80-20 hold-out validation. This alternative approach was explored due to the under-sampling approach adopted by the RUSBoost classifier in the training stage. The epochs in the training and test set were randomly selected. More specifically, the split was conducted both on subject and epoch level, in such a way to ensure inter-class balance, and limit the risk of data leakage. Classification performance is reported in Table 4.6, and class-wise performance is displayed through the summative confusion matrix in Figure 4.5.

A macro-averaged accuracy of 87.7% was attained through this boosting method; however, as a general trend, performance metrics were slightly lower than the two previously explored classifiers. Indeed, despite presenting with a modest rise in detection performance as regards sleep stage N1 – although unsatisfactory – a loss in performance was transversally appreciable. A possible interpretation of this outcome was found in the under-sampling method adopted by the classifier, which, on the one hand, ensures balanced class representation, but, on the other hand, provides the model with a scarce set of examples. However, detection performance of the REM stage remained reliable, with 87.2% Recall, and almost 69% F1 score. Furthermore, this classifier featured a κ of 0.7; although lower than the coefficient obtained by the two other models, this value is indicative of a moderate score agreement between the real and predicted stages.

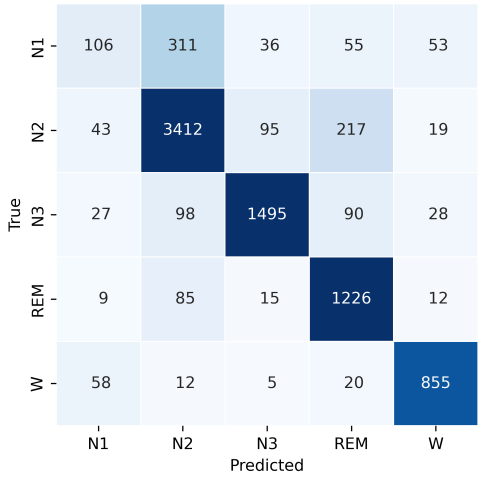


Figure 4.3: Confusion Matrix of the the Random Forest classifier.

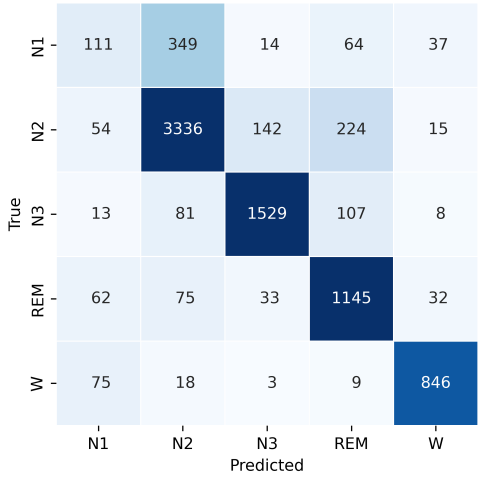


Figure 4.4: Confusion Matrix of the the KNN classifier.

Generally, the three tested classifiers proved good detection performance when considering only the sleep stages, with average accuracy above 86% in the least-performing configuration.

Validation on External Dataset: 5-stage

The previous paragraph illustrated the results of the 5-stage classification in the proposed configuration.

Table 4.6: Performance (80/20 hold-out) of the RUSBoost classifier, in the 5-stage configuration.

	W	**REM**	**N1**	**N2**	**N3**
Accuracy	92.1 %	89 %	85.4 %	82.2 %	89.7 %
Recall	49.8 %	87.2 %	43.6 %	57.9 %	98.4 %
Specificity	98.8 %	89.9 %	89.2 %	97.3 %	89.2 %
Precision	86.4 %	56.8 %	26.9 %	93 %	75.8 %
F1	63.2 %	68.8 %	33.3 %	71.4 %	85.6 %

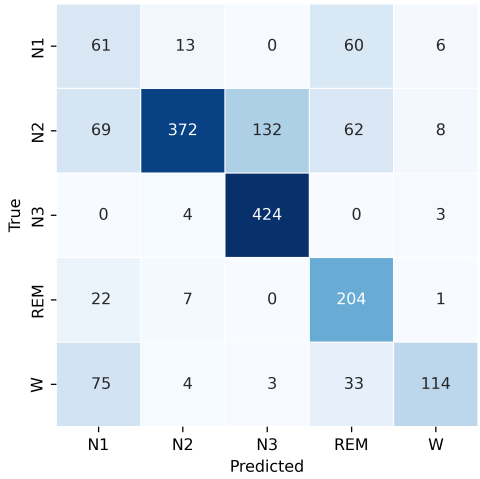


Figure 4.5: Confusion Matrix of the RUSBoost classifier, performance on the test set (80/20 hold-out).

The performance of the employed models was evaluated both in a k-fold CV and a more traditional hold-out approach, with reasonably promising metrics. However, as stated in the introductory part of this Chapter, this study relied on a public dataset, and the employed models were trained on healthy subjects data.

At the time of the study, no further investigations on sleep structure in non-healthy cohorts were conducted; however, to assess the reliability of the proposed set of features, in a single-channel configuration, the trained models were tested on a set of unseen data including subjects at risk for neurodegeneration.

For this purpose, RBD subjects in the CAP Sleep Database were adopted as external test set, and the best-performing model – i.e., the Random Forest classifier – was validated on this data batch.

Data preparation and Feature Extraction followed the procedure described in the Methods section. A total of 14583 epochs of length 30 seconds were collected in

Table 4.7: External validation of the RF classifier on RBD subjects from the CAP Sleep Database.

	W	**REM**	**N1**	**N2**	**N3**
Accuracy	88.1 %	89.5 %	94.8 %	82.6 %	86.6 %
Recall	85.2 %	58.1 %	41.1 %	65 %	76.1 %
Specificity	89.3 %	95.5 %	98.6 %	88.1 %	90.3 %
Precision	75.7 %	71.5 %	68.3 %	62 %	73.3 %
F1	80.2 %	64.1 %	51.3 %	63.5 %	74.7 %

the dataset. As expected, the data featured a high preponderance of the N2 class, with over 5000 observations. Data post-processing was carried out as previously described, and observations with missing data, or with features resulting in NaNs, were discarded.

This procedure led to a total of 10469 epochs, distributed as follows:

- **NREM**: 449 epochs for N1, 4894 for N2, 2820 for N3;
- **REM**: 1555 epochs,
- **WAKE**: 751 epochs.

The results of the 5-stage classification are reported in Table 4.7 and Figure 4.6. The results were rather promising, with an overall accuracy of 87.11%. As a general trend, the detection of N1 stage remained unsatisfactory, with F1 score slightly above 50%.

A drop in detection performance in the REM stage was also observed, achieving a Recall and F1 score of 58% and 64%, respectively. This trend might be due to the fact that subjects with RBD exhibit electroencephalographic differences in the REM stage, compared to healthy individuals. This concept was later explored and is presented in Chapter 6.

Classification Performance: REM/NREM

As previously mentioned, one of the study aims was the characterisation of REM sleep through a set of novel, class-specific predictors. Precisely, these features are meant to describe the dual nature – i.e., *tonic* or *phasic* – of the REM stage.

For this purpose, an alternative configuration was tested on the external dataset. Specifically, this envisaged a binary classification task, between REM and NREM sleep, in subjects from the CAP Sleep Database; the healthy and RBD subjects were tested separately. The choice of employing a different dataset to tackle REM detection was made to assess the cross-reliability of the proposed REM microstructure features.

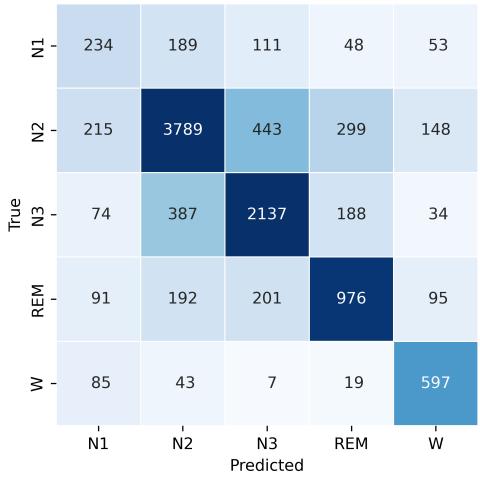


Figure 4.6: Confusion Matrix of the RF classifier, as regards the external validation on RBD subjects from the CAP Sleep Database.

To carry out this task, NREM stages N1, N2, N3 were collected in the same class, and wake epochs were discarded. The choice of only including sleep segments in this task, as stated in Paper [141], was made to assess the net contribution of the proposed novel features when targeting the direct detection of REM sleep, with respect to the commonly employed sleep-describing features. Moreover, as appreciable from the confusion matrices in Figures 4.3–4.5, in the 5-stage configuration, the detection of the REM stage was quite robust against misclassification with wake epochs; hence, these latter were not considered in this assessment.

Classification performances of three different supervised models were explored, to investigate the predictive power of the novel features related to microstructural arrangements in REM sleep. Namely, a KNN classifier, along with a SVM and DT, were employed. These two latter were chosen as they are commonly employed in binary classification tasks.

A nested LOSO-CV approach was adopted to assess detection performance (Figure 4.7). In this configuration, training is conducted on sleep epochs from $N-1$ subjects, and the model is tested on a single epoch from the left-out subject. The procedure is repeated for all subjects in the dataset. The Mean Squared Error (MSE), expressed as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(Y_{real} - Y_{pred})^2 \tag{4.7}$$

where Y_{real} and Y_{pred} are the observed and predicted values, respectively, was introduced to assess the goodness of fit of the models.

Figure 4.7: Nested Leave-One-Subject-Out approach adopted in the REM/NREM classification task.

The achieved results are shown in Tables 4.8 and 4.9, for the healthy and RBD group, respectively.

As appreciable, when testing the binary configuration REM/NREM in the healthy group, the detection performance remains promising, with Recall values above 93% in all tested classifiers, and an average Accuracy of 91.26%. The DT classifier achieved the lowest MSE, suggesting a positive behaviour of the implemented framework, while the SVM classifier performed slightly worse, featuring the highest value of the MSE. In this regard, it is important to note that model optimisation was overlooked, and likely the two classes were not linearly separable.

Regarding the RBD group, the classification performance met an observable decrease, though remaining still acceptable, with the KNN and DT attaining an overall accuracy above 75%. Though all explored models featured Recall values of at least 70%, the SVM resulted anew the worst-performing model in this configuration.

Table 4.8: Performance of the binary classification task (REM/NREM) on healthy controls in the CAP Sleep Database.

	Accuracy	**Recall**	**Specificity**	**Precision**	**MSE**
KNN	92.6 %	96.3 %	86.4 %	92.3 %	8 %
DT	94.9 %	97.9 %	89.7 %	94.1 %	5 %
SVM	86.3 %	93.5 %	74.1 %	85.9 %	13 %

Table 4.9: Performance of the binary classification (REM/NREM) task on RBD subjects in the CAP Sleep Database

	Accuracy	**Recall**	**Specificity**	**Precision**	**MSE**
KNN	75.9 %	77.9 %	74 %	74.9 %	24 %
DT	76.5 %	76.7 %	76.3 %	76.4 %	23 %
SVM	64.3 %	70.2 %	58.2 %	62.7 %	36 %

Predictive Performance of the Proposed Features

The study presented in this Chapter tackled automatic sleep staging by introducing the analysis of REM microstructure.

The classification performance achieved by the explored classifiers suggested a good predictive power of the proposed features. However, to further investigate this point, the proposed features were ranked according to their importance. This latter was assessed through Pearson's correlation coefficient. This metric evaluates the relevance of each variable to the target by expressing the strength (as magnitude) and direction (as sign) of the relationship.

Table 4.10 reports the results and the statistical significance of each variable, assessed through Mann-Whitney U test. Moderately high values of correlation were observed across all variables, which also featured a positive correlation with the task.

Table 4.10: Correlation (Pearson's ρ) with the REM class for the set of novel features implemented, along with their statistical significance (p-value).

Feature	**Pearson's ρ**	**Unpaired Test**
Approximate Entropy	0.69	$< 0.05^{*}$
Absolute Power (TREM)	0.67	$< 0.01^{**}$
Coastline (1^{st} derivative, envelope)	0.52	$< 0.05^{*}$
Complexity (envelope)	0.57	$< 0.05^{*}$
Mobility (envelope)	0.74	$< 0.05^{*}$
SEF50 (TREM)	0.6	$< 0.01^{**}$
SEF95 (TREM)	0.51	$< 0.05^{*}$
SEFd (TREM)	0.5	$< 0.05^{*}$
Shannon Entropy	0.7	$< 0.01^{**}$
Spectral Entropy	0.7	$< 0.01^{**}$

Table 4.10 reports the results and the statistical significance of each variable, assessed through Mann-Whitney U test.

Finally, to evaluate the effect of this implementation, a binary classification

task (REM vs NREM) with and without this set of features was carried out, in the RBD cohort of the CAP Sleep Dataset, to assess their applicability in a real-world scenario.

The KNN classifier, which presented with consistent detection performance in all tested scenarios, was employed for this task. The results are shown in Table 4.11. When comparing these metrics with the ones previously presented in Table 4.9, a minor, yet appreciable rise in performance is observed when the novel features are implemented. In particular, the KNN classifier in Table 4.9 presents with lower MSE, and higher values of recall, specificity, and precision.

Table 4.11: Performance of the binary classification task (REM/NREM) on RBD subjects in the CAP Sleep Database, without implementing the novel features.

	Accuracy	**Recall**	**Specificity**	**Precision**	**MSE**
KNN	74.2 %	77.6 %	70.8 %	72.7 %	26 %

4.3.4 Discussion

The research activity described in these Sections proposed an automatic sleep staging framework based on a single EEG channel, to assess the feasibility of home-based, minimally invasive, sleep studies.

For this purpose, various features were extracted from each stage, and the REM stage was further characterised through its underlying microstructure. This step represented a novelty, since, as a general rule, REM sleep had previously been considered homogeneous in feature engineering tasks.

The achieved results revealed good ability in detecting the REM stage, with a micro-averaged accuracy of almost 90% in all the three investigated models. A good trade-off in Recall and Specificity was also observed, especially in the RF and KNN models, which expressed a reasonably good Precision value (almost 75%), considering that the analysis was carried out on raw data, undergoing minimal signal processing.

The achieved results outperformed those presented in a previous study in the literature [84], based on the same dataset, where the Authors achieved an overall Accuracy of 88.5%, and Recall of 82.3%. Furthermore, the classification performance of the proposed framework appeared comparable to [42], which, however, employed a larger number of physiological signals, including multiple EEG, EOG, and EMG channels in the analysis.

The performance of the classifiers remained quite consistent when testing the trained models on an external data batch, which served as external validation, including both healthy and subjects at risk for neurodegeneration.

The lower performance recorded in the RBD group as regards REM detection suggested that patients with RBD exhibit a peculiar feature distribution when considering REM microstructure, a concept that would have been explored in future studies, and is described in Chapter 6.

The findings highlighted in this study suggested the applicability of the proposed lightweight metrics in the detection of REM sleep, and subsequent sleep staging according to the AASM criteria, with simple but efficient supervised Machine Learning models. However, some strategies to corroborate the proposed models should be adopted. First, a larger dataset could be employed in the training part of the model, thus adding in variability and allowing for better generalisation capability. Second, the healthy cohort should also envisage older subjects, to promote a more stratified training framework.

4.4 Conclusion

The classification of sleep stages is a fundamental aspect for the definition of sleep architecture and the assessment of circadian health and sleep-related disorders. The categorisation of sleep stages in adults is regulated by the AASM criteria, which envisage overnight wakefulness, three stages of NREM sleep (from light to deep sleep), and REM sleep.

Nowadays, traditional, in-hospital polysomnography is still considered the gold standard diagnostic procedure to perform sleep staging; it consists in a collection of various biosignals, including EEG, EOG, and ECG channels, which allow for the observation of the sleep evolution over multiple sources. Despite its high resolution and reliability, this technology bears some limitations in terms of patient comfort, diagnostic intrusiveness, and, lasty, high costs, and extensive manual labour.

Automatic sleep stage classification methods have emerged as a possible solution to address the issue of protracted scoring times, by leveraging Machine or Deep Learning techniques, with performance comparable to manual scoring. However, a large number of these methods rely on numerous PSG biosignals, thus offering the opportunity to only streamline the diagnostic process, without alleviating the intrusiveness of the test.

In recent years, much effort has been dedicated to the implementation of home-based sleep studies, with a simplified sensor configuration. Moving towards this perspective, the research activity reported in this Chapter aimed at performing sleep stage classification through a single EEG channel, to assess the feasibility of home-based sleep studies through lightweight instrumentation.

Precisely, the central EEG channel was selected from PSG recordings of healthy subjects in a publicly available dataset, feature engineering was carried out to characterise both overall sleep architecture and stage-specific patterns. Automatic 5-stage classification was explored through various supervised Machine Learning

models, achieving considerably accurate results. The same framework was tested on an external data batch, including both healthy and subjects with RBD, maintaining reasonable detection capability. This holds premise for future implementations of the proposed framework in populations including subjects at risk for neurodegeneration, or subjects with early-stage PD.

Furthermore, the automatic staging approach offered a significant reduction in scoring times. Upon consultation with a sleep neurologist, we concluded that manual annotation of sleep stages took on average three working days. For this work, data analysis was performed on a Windows 10 x64 machine, Intel® Core i7-6700HQ @ 2.6 GHz, RAM 16 GB, NVIDIA® GEFORCE® GTX 960M. Feature extraction required 540 $\pm$ 60 s per subject, and automatic scoring 0.034 $\pm$ 0.009 s per subject. Though being aware that the obtained results were preliminary, the proposed framework could alleviate the manual staging step.

Despite the encouraging clinical implications, several challenges persist.

First, and as briefly discussed, future developments should address a larger dataset, in order to properly validate the classification performance and assess the model robustness. In this regard, validation studies should be undertaken across diverse patient populations, including healthy subjects and subjects with sleep disorders, and possibly from different diagnostic centres. An exploratory approach to this issue has been provided in this Chapter, when evaluating the classification performance across various datasets, recorded with different instrumentation, in an attempt to simulate external validation. Furthermore, efforts should be made towards providing accurate stratification in terms of but also in terms of sex ratio, or age. Indeed, the employed datasets featured an observable imbalance in terms of age, and previous studies suggested that EEG signals express varying features when age and sex are introduced as co-factors [105, 26], with older subjects expressing reduced slow-wave sleep, and increased fragmentation. Finally, the use of larger datasets would still inevitably reflect the intrinsic class-imbalance of sleep data, as thoroughly discussed in this Chapter. While class-balancing could represent a useful strategy to provide a better generalisation for minority samples, it might eventually result in overfitting, or in the inability to represent real-world data. In this Chapter, a potential solution was be found in the adoption of sampling or weighting techniques, as seen through the RUSBoost classifier; however, the optimal solution should be tailored according to specific deployment use-case.

Second, the framework proposed in this study relied on high-quality data, recorded by expert technicians with high-resolution sensors. In view of lightweight sleep studies, the performance of portable devices should also be regarded, to ensure diagnostic accuracy. Indeed, motion artefacts or electrode detachment may occur and virtually decrease signal quality, thus providing inaccurate results which could hamper clinical utility. The verification of this point is still ongoing in cooperation with the Regional Centre for Sleep Medicine at the Molinette University Hospital (Turin, Italy), where preliminary home sleep tests are being

carried out through the portable Nox A1s PSG System (Nox Medical, Iceland, https://noxmedical.com/), which includes frontal (F), central (C), and occipital (O) channels. While various home-based and consumer-level solutions offer primarily pre-frontal (Fp) or frontal channel recordings, the work included in this Chapter relied on central-channel data. This choice was made due to the fact that central channels are regularly available in PSG recordings, and are less affected by motion artefacts, or muscle cross-talk. Future investigations will explore alternative configurations, and compare performances with frontal or pre-frontal locations, especially in consideration of the convenience of self-applied sleep monitoring headbands [8, 101].

Third, the EEG features have been evaluated on fixed-length epochs, related to the scoring guidelines provided by the AASM manual, thus allowing the direct comparison of the automatic performance with the manual annotations. However, detection capability on the transitional N1 stage remained unsatisfactory, probably due to its mixed characteristics, which fail to be represented through traditional methods. Future investigations should explore different epoch configurations, thus providing an accurate characterisation of sleep, moving towards a continuous-wise, rather than a discrete-wise, analysis.

Finally, in order to provide clinical interpretability of the employed metrics, and to allow for a proper comparison with the contemporary literature, the research activity presented in this Chapter relied on simple, hand-engineered features, combined with well-known supervised Machine Learning models. This allowed for the employment of a low-computational cost architecture and the identification of the most relevant features for each sleep stage. Nevertheless, while on the one hand future refinements could look to improvements in the feature extraction procedures, or in the classification algorithms, on the other hand fully data-driven should be explored. Indeed, in recent years, Deep Learning-based frameworks have been widely employed for sleep staging paradigms, alleviating class imbalance with data augmentation techniques [56], and providing model interpretation through explainable approaches. A preliminary investigation of deep networks on size-limited, sleep EEG data is currently underway. This experimentation seeks to explore the potential of deep architectures on tiny datasets, and its performance, compared to well-established, feature-driven shallow models. The result of this investigation have not been included in this dissertation, as they still appear embryonic. However, although expressing encouraging potential, the large-scale integration of such models is still hindered in the real-world scenario by data scarcity, and the clinical implementation necessarily requires validation through domain knowledge.

Chapter 5

Quantification of REM Sleep Without Atonia

Clinical scoring and quantification of REM Sleep Without Atonia (RSWA) is regulated by a set of rules that rely primarily on visual inspection and manual scoring of polysomnographic recordings, ofttimes leading to inter- or intra-rater variability, and protracted manual labour. In view of the development of robust diagnosis support systems, a semi-automatic framework for the correction of artefacts in the quantification of RSWA has been proposed.

This Chapter describes the Rationale and Research Activity included in the work [143], presented at the 17^{th} World Sleep Congress, in October 2023. The experimental work and study design were conducted at the Sleep Disorders Unit, Department of Neurology, Medical University of Innsbruck (Austria).

5.1 Context and Background

REM Sleep Without Atonia is a parasomnia that entails the loss of physiological muscle atonia during REM Sleep, and is considered the polysomnographic hallmark of REM Sleep Behaviour Disorder (RBD) [149]. From a physiological perspective, RSWA is a condition that entails abnormal, sustained, or periodic, muscular activity, observed through polysomnography, and confirmed through clinical interviews [81, 164]. Clinical history of injurious or disruptive behaviour, or episodes, later documented by PSG is also required to complete the diagnosis of RBD [33].

The currently available RSWA scoring criteria have been regulated by the AASM Manual for Scoring of Sleep and Associated Events [11], and require visual inspection, combined with manual scoring of EMG traces in PSG, though there is no standardised indication of *abnormal* EMG activity. More recently, and in the context of RBD assessment, the International RBD Study Group proposed an update to the recommended scoring methods included in the international guidelines

[149, 33], by adding the requirement of dream enactment as seen through video-polysomnography (vPSG) for an accurate and complete diagnosis of RBD. The recommended clinical guidelines, and the recent adjustments, bring RSWA scoring often to the stage of long and protracted, manual labour. Indeed, the scoring metrics are primarily obtained after visual assessment of PSG recordings, carried out by sleep technologists or neurologists. Specifically, visual inspection of PSG recordings is required to determine the type of EMG activation, which, is commonly defined either as tonic, i.e., sustained, elevated background activity, or phasic, i.e., rapid bursts of activity, generally identified as twitches.

In this Chapter, tonic activity will be referred to as *elevated background tone*. Over the years, various manual scoring methods have been proposed in the literature to quantify the extent of RSWA (cf. Table 5.1). These methods, which are manual and visual-based, aimed at quantifying the amount of abnormal EMG activity during REM sleep in previously set lengths of windows of observation, commonly employing pre-defined amplitude and duration thresholds. The first scoring method, commonly referred to as the Montréal Method, was proposed by Lapierre and Montplaisir [100]. The method evaluates tonic activity over 20-second length epochs, and phasic activity in epochs of 2-second length. An epoch is defined as tonic if sustained EMG activity of amplitude at least twice the background, or above 10 µV, occurs at least for 50% of its duration. Phasic activity, on the other hand, is assessed over 2-second epochs, and defined as bursts of activity of amplitude four times the background, and duration between 0.1—10 seconds.

Subsequently, other methods have been developed, mainly differing in terms of observation window length and activity duration. The most relevant two are the Mayo Clinic method [116] and the Sleep Innsbruck Barcelona (SINBAR) scoring method [68], along with others, though less common [114]. For the sake of clarity, the Montréal, Mayo and SINBAR specifications are summarized in Table 5.1.

Despite the plurality of RSWA scoring methods proposed in the literature [135, 66], none has been included as the official framework in the AASM manual for scoring sleep. Nevertheless, the most recent update of the ICSD-3 [149] highlighted the SINBAR method as the recommended scoring procedure for quantifying RSWA. As briefly mentioned, this method is based on the evaluation of EMG activity during REM sleep on various channels to allow for a reliable diagnosis [81], and it will be described in the following Section.

5.2 The SINBAR Scoring Method

The SINBAR scoring method [68] provides a reliable framework for the quantification of RSWA, and the subsequent detection and screening for RBD, as emerged from a validation study on a RBD patients cohort [67]. While the Montréal method [100] relied solely on the mylohyoid activity, and the Mayo [116] introduced the

Table 5.1: Overview of the most common visual-based scoring methods for REM Sleep Without Atonia. Bkg: background activity (baseline). Tonic Density is evaluated on the mentalis muscle.

Method	Tonic Density	Phasic Density
Montréal [100]	(%) of 20-second epochs with at least 50% of the epoch with amplitude twice the Bkg, or greater than 10 µV	(%) of 2-second mini-epochs with bursts of activity (duration: 0.1–10 s) with amplitude 4 times the Bkg, or greater than 10 µV
SINBAR [68]	(%) of 30-second epochs with at least 50% of the epoch with amplitude twice the Bkg, or greater than 10 µV	(%) of 3-second mini-epochs with bursts of activity (duration: 0.1–5 s) with amplitude twice the Bkg, or greater than 10 µV
Mayo Clinic [116]	(%) of 30-second epochs with at least 50% of the epoch with amplitude twice the Bkg, or greater than 10 µV	(%) of 3-second mini-epochs with bursts of activity (duration: 0.1–14.9 s) with amplitude twice the Bkg

analysis of the tibialis anterior (TA), the SINBAR scoring method identified the combination of the mentalis and bilateral flexor digitorum superficialis (FDS) as the optimal configuration for accurately quantifying RSWA [69]. Precisely, abnormal muscle activity is manually quantified as tonic, phasic, or *"any"* – i.e., either of the two. This latter was introduced to ensure an exhaustive evaluation, by analysing those EMG activations who would not fall within the first two conditions. The criteria for the categorisation are the following:

- **Tonic activity**: it is scored in the mentalis in 30-second epochs. Each epoch is scored as tonic if elevated background tone is present for more than 50% of the epoch, with amplitude at least twice the background, or above 10 µV.
- **Phasic activity**: it is scored in the mentalis and FDS muscles, on mini-epochs of length 3 seconds. Each mini-epoch is scored as phasic when bursts of muscle activity with amplitude twice the background and duration between 0.1—5 seconds are present.
- ***"Any"* activity**: it is scored in the mentalis muscle on 3-second mini-epochs. The epochs include "any" activity when either phasic or tonic activations are observable.

Finally, the phasic, tonic, and *"any"* indices are computed as the ratio between the epochs featuring activity and the total number of epochs. The SINBAR index [67] is also computed, as the combination of the presence of *"any"* activity in the mentalis and phasic activity in the FDS muscle. Normally, artefacts and sustained EMG tone due to arousals lead to an exclusion from the RSWA quantification.

A computerized version of this method was first presented in [69], with good scoring accuracy. The validated software (OSG, Belgium, `https://www.osg.be/`), is currently integrated in a clinical PSG system (BrainRT) at the Medical University of Innsbruck, and enables the automatic detection of EMG activity during REM sleep. Although the automatic scoring software is widely used, a validation study by the same group [69] shed light on the need for manual artefact correction to ensure robustness between the obtained score and the expected manual evaluation, although the performance on the FDS activity remained stable between the two. An important source of artefacts is found in snoring or sleep apneas, that significantly alter the submentalis EMG amplitude, resulting in observed loss of atonia, even in the absence of clinically confirmed RSWA. This is likely not observed in different muscle sources, such the tibialis anterior, or the flexor digitorum superficialis, as proposed in the SINBAR montage [67].

5.2.1 Manual Artefact Correction

Currently, no clear definition of artefact is provided in literature or in the guidelines. Generally, artefacts during REM sleep may be linked to arousals, snoring, or technical issues. The lack of a systematic approach likely translates into the proneness of visual and semi-automatic methods to inter-rater and intra-rater variability. For this reason, the SINBAR scoring method recommends manual removal of artefacts caused by snoring, technical issues, or ECG cross-talk when evaluating RSWA through the semi-automatic PSG software [69]. Recently, a study on interrater reliability, in the context of artifact correction, has been conducted by the same group, on a cohort of 25 RBD patients [31]. After manual correction of artefacts performed independently by four expert scorers (Figure 5.1), the SINBAR metrics were retrieved for each of the manually-corrected subjects and compared to the automatic scores obtained through the software.

The study highlighted the FDS as the source with higher inter-rater variability ($p < 0.001$), especially in the case of phasic activity. Although proving high reliability of the FDS muscle in representing phasic activity in RBD subjects, manual correction of artefact still requires huge, and often protracted, manual labour. Moreover, the results presented in the study show some disagreement in the choice of correction of artefacts, and the subsequent metrics assessment. This likely occurs in the mentalis muscle, as phasic activity, breathing and snoring introduce artefacts.

These premises harbour the need for automatic detection of artefacts, primarily

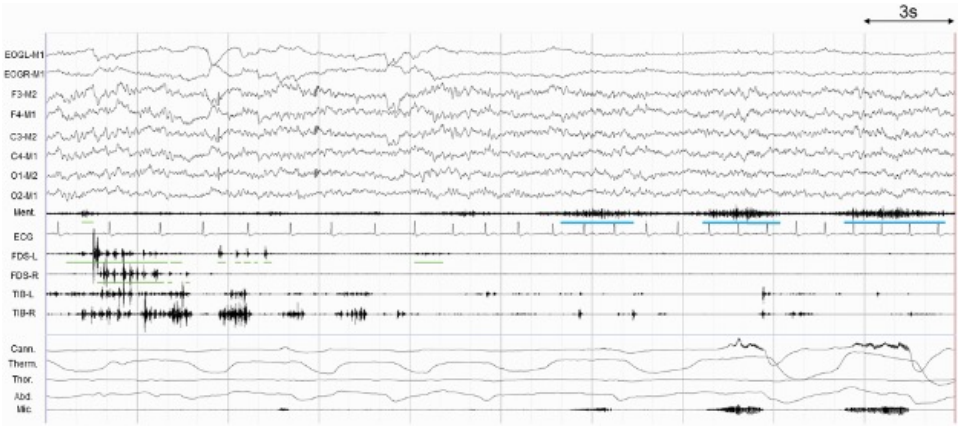


Figure 5.1: Example of considered breathing/snoring events in the *mentalis* muscle, and phasic activity in the *flexor digitorum superficialis* muscle. Portions in blue indicate manually corrected artefacts, whereas portions in green indicate the activity correctly identified by the algorithm. Adapted from [31] with permission from Oxford University Press (obtained: 19$^{\text{th}}$ February, 2024).

to: **(1)** expedite the scoring process, by easing manual labour, and **(2)** facilitate the full automatic scoring of muscle activity during REM.

Finally, it would allow for a full automated pipeline in the evaluation of REM sleep without atonia.

5.3 Research Activity

The presented research background raised the need for an objective, possibly automatic, evaluation of artefacts, to smooth the evaluation procedures for scoring REM sleep without atonia. The research activity presented in this Chapter addressed this issue.

The research pipeline, detailed in the following Sections, aimed at:

1. Objectively characterise the morphology of the EMG signal during sleep,
2. Highlight differences in muscle activations of RBD subjects *vs* controls,
3. Automatically detect muscle artefacts,
4. Validate the artefact removal method by comparing the obtained scores to the manual ones.

Automatic Detection of Artefacts in the Evaluation of REM Sleep Without Atonia

5.3.1 Materials

Subjects and Data

The preliminary investigation stage of this study included 185 subjects who underwent vPSG at the Sleep Lab in the Medical University of Innsbruck (Austria); these recordings were employed to explore the morphological characteristics of the EMG signal.

Subsequently, the second stage included a total of 25 subjects (aged 57.2 ± 14.9 years), who underwent 8-hour vPSG, scored according to the AASM criteria [11], and already included in [31]. All PSG recordings included six EEG channels (F3, F4, C4, C4, O1, O2, with reference electrodes M1, M2), the EMG channels (mentalis, submentalis, bilateral FDS, and bilateral tibialis anterior), one ECG channel, electro-oculography for eye movements, and the canonical cardiorespiratory signals. This cohort included 8 subjects with RBD, carefully selected so as to have four subjects with AHI_{REM}[1] < 15/h and four with $AHI_{REM} > 15$/h. The remainder (17 subjects) were randomly included from a cohort under study for suspected parasomnia, and were selected likewise, so has to have two groups following the same AHI criteria. No statistical differences in age or sex distribution in the two groups were observed [31].

Manual Artefact Correction and Scoring

The manual artefact correction rationale and procedure was thoroughly described in the prior study by the Innsbruck group (cf. Section 5.2.1, [31]); for the sake of clarity, this paragraph will briefly report the steps employed.

First, 3-seconds mini-epochs for the analysis were selected by an expert scorer among the available 30-s REM epochs, following the SINBAR recommendations for scoring RSWA [68] The occurrence of a sleep spindle, a K-complex, or the prolonged absence of rapid-eye movements determined the end of the mini-epoch [100, 31]. A total of 956 ± 70 mini-epochs were included. Subsequently, semi-automatic RSWA scoring was carried out through the validated OSG Software, included in the BrainRT scoring system. As mentioned previously (cf. 5.2), the algorithm was employed to score tonic, phasic, and *"any"* activity in the mentalis muscle, and phasic activity in the bilateral FDS and TA muscles. Finally, following the pipeline previously described by [31], manual artefact correction was carried

[1]The AHI is the Apnea-Hypopnea index, and represents the combined number of apnea/hypopnea events that occur in one night.

out independently on all subjects by four different scorers, blinded to the subjects' diagnoses. For the purpose of this work, probabilistic consensus of the four scorers was obtained and employed as ground-truth for the final score comparison.

5.3.2 Methods

The following Sections illustrate the Methods adopted in the study, from the preliminary signal processing to the automatic detection of artefacts in the two proposed configurations.

Signal Pre-Processing

The EMG recordings from the mentalis, submentalis and bilateral FDS were pre-processed with a Butterworth bandpass filter (order 16), in the range 50–300 Hz; powerline rejection at 50 Hz was also carried out. To ease the computational load and facilitate the subsequent procedures, the EMG recordings were down-sampled from 1000 Hz to 600 Hz. As in [31], the recordings from the bilateral TA were not included in the subsequent analysis, as previous studies highlighted that this muscular activity is not directly connected to the identification of subjects with RBD [67].

Feature Extraction and Statistical Testing

As per research objectives number **(1)** and **(2)** (cf. Section 5.3), this work first aimed at characterising the morphology of muscle twitches during REM sleep, to differentiate between RBD subjects and controls. To this end, morphological characterization was addressed through various descriptors – i.e., *features* – sorted in the following categories: temporal, spectral, time-frequency, and non-linear. The list of extracted features is displayed in Table 5.2, along with their description. These features were computed on 1-second epochs, and concatenated so as to obtain a feature array. Finally, for each feature, a set of statistics were computed (mean, median, mode, 25^{th} and 75^{th} percentiles, interquartile range, and interdecile range) and employed in the classification.

A Mann-Whitney U test [109] was applied to the extracted data to evaluate differences in distribution as regards the EMG signal characteristics in the two groups (RBD and controls, respectively).

Automatic Removal of Artefacts

The second part of the study focused on developing a method for the morphological characterisation of EMG during REM sleep to allow for the automatic removal of artefacts, through supervised ML methods. The employed pipeline is

Table 5.2: Features extracted for the morphological characterisation of EMG during REM, along with the domain and proper reference. ⋄: adapted from the cited study; †: first proposed in this work.

Category	**Feature (Name and description)**	**Reference**
Time	Amplitude metrics: mean, standard deviation, skewness, kurtosis, range, maximum and minimum value, root mean square	various
	Zero Crossing Rate	⋄[165]
	Hjorth Parameters	⋄[122]
	Percentiles (5^{th}, 10^{th}, 25^{th}, 75^{th}, 90^{th})	various
	Form, Crest and Impact Factors	various
	Event Duration	†
Frequency	Fast Fourier Transform: numerical and statistical measures (mean and median frequencies, total power, ...)	various
	Spectral Edge Frequencies (SEF25, SEF75, SEF95)	⋄[141]
	Entropy measures	⋄[141, 1]

summarized in Algorithm 2. For the sake of clarity, in this Chapter, the occurrence of RSWA will be referred to as *activity* (-ies).

As the selected REM mini-epochs included both activities and artefacts, the rationale behind this work was to accurately characterize the regions of interest, in order to proceed to the automatic detection of artefacts. Since the prior interrater reliability study from the SINBAR group [31] stated that phasic activity was the most affected by artefacts, this work first aimed at finding a metric to accurately *match* the morphology of phasic activations.

To serve the purpose, wavelet-based approaches are widely adopted in EMG processing [127, 126], and were previously proposed in the identification of EMG activations in sleep studies [54].

Following an approach similar to [28], a Continuous Wavelet Transform (CWT) was employed to match the morphology of the phasic activations. Indeed, although similar from a morphological point of view, signals and artefacts are likely based on different time-frequency scales. As wavelet transforms are commonly employed to decompose the original signal of interest into various time-varying frequency components (i.e., *scales*), this approach was adopted to differentiate the two classes.

Algorithm 2 Automatic Detection of Artefacts

```
Select 3-s REM epochs;
Divide into 0.1-s mini-epochs (M);
for epoch ← 1 to M do
    Compute Correlation Index
    Compute Energy-Based Metric
end for
Binary Classification of mini-epochs (Artefact vs Activity)
Exclude epochs marked as Artefact
Compute RSWA Metrics (according to the SINBAR assessment)
```

From a theoretical point of view, a CWT is defined, for a continuous signal $x(t)$:

$$C(a, b; x(t), \psi(t)) = \int_{-\infty}^{\infty} x(t) \frac{1}{a} \psi^*(\frac{t-b}{a})\, dt \tag{5.1}$$

Where $\psi(t)$ is the so-called *mother wavelet*. This function is the dot product of continuous signal $x(t)$ with a shifted (b) and scaled (a) version of $\psi(t)$. It yields the various constituent waveforms of the original signal. The two factors represent the morphology and the time positioning of the wavelet, respectively.

Generally, this enables the use of a wavelet transform as a feature detector, by employing a proper mother wavelet – for instance, a function resembling the *detectable* feature. As phasic activations present with a symmetrical distribution (Figure 5.2), the mother wavelet of choice was selected from the biorthogonal wavelet family [91], that includes compactly supported, symmetric wavelets. In principle, differing from Equation 5.1, a biorthogonal wavelet is formed by duality, and the presence of two wavelet functions [40], one employed in the analysis and one in the synthesis.

The selected mother wavelet (biorthogonal 3.9) is displayed in Figure 5.2; this wavelet family was selected as the shape strongly resembles the shape of a phasic activation.

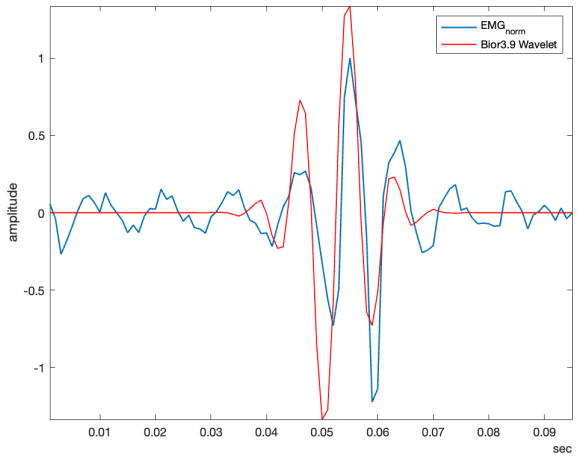


Figure 5.2: Morphology of a rapid, phasic activation observed in the mentalis muscle (blue line), and the bior3.9 mother wavelet selected for analysis and synthesis (red line).

In its discrete form, i.e., the Discrete Wavelet Transform (DWT), the scale and shift factors are integers. To obtain the various frequency components of a signal $x(t)$, the DWT can be applied to the continuous signal as a *filter bank*, working in a dyadic way (Figure 5.3), at the specified level of decomposition, to obtain

the approximate (cA) and detailed (cD) coefficients, virtually corresponding to the low-pass and high-pass version of the signal, respectively.

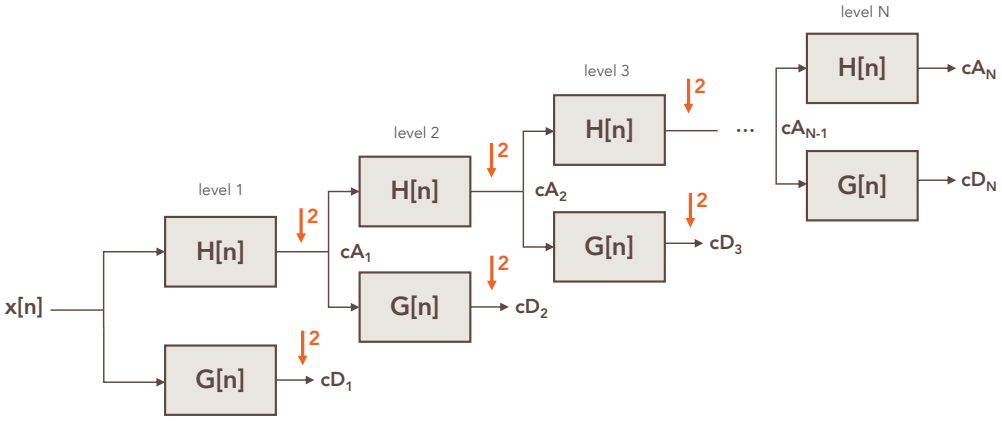


Figure 5.3: Discrete dyadic Wavelet Transform at the N^{th} level of decomposition. At each level, H[n] works as a low-pass filter and yields the approximate coefficients (cA), while G[n] works as a high-pass filter with the detail coefficients (cD) as output. At each level, the cut-off frequency is downsampled by a factor 2.

In this work, multi-level decomposition (5 levels) was applied to the selected REM mini-epochs, to obtain the cA and cD, to allow for a fine tuning of the one-dimensional signal reconstruction. The spectral distribution of the reconstructed signal (both for activity and artefacts segments) was explored through the Kernel Density Estimation (KDE), which highlighted the coefficients at the third level of detail (virtually corresponding to the frequency range 75–150 Hz) as the ones providing the best differentiation between the two signal types (Figure 5.4).

As appreciable, the selected time-frequency scale is able to decouple the different spectral contributions. Finally, the original signal was reconstructed by employing the wavelet decomposition at level 3 (Figure 5.5).

From a preliminary, visual inspection, the reconstructed signal morphologically resembled the signal with phasic activations, leaving less precise and noisy signals when matching with the artifacts.

Feature Extraction

To carry out the automatic detection of artefacts, supervised Machine Learning algorithms were employed, with features extracted from the selected EMG REM epochs, which were processed in 0.1-seconds sub-segments.

To differentiate artefacts from phasic activity, the original signal and its wavelet-reconstructed version have been exploited. Namely, an index of correlation and an energy-based metric were extracted and used in the ML pipeline.

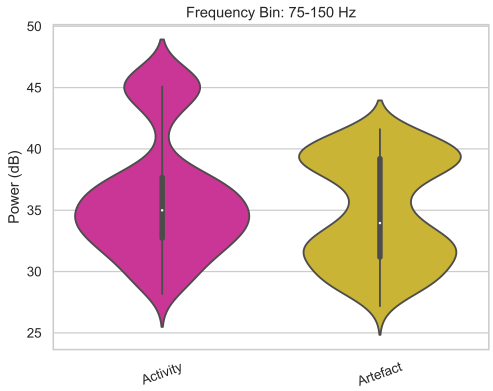
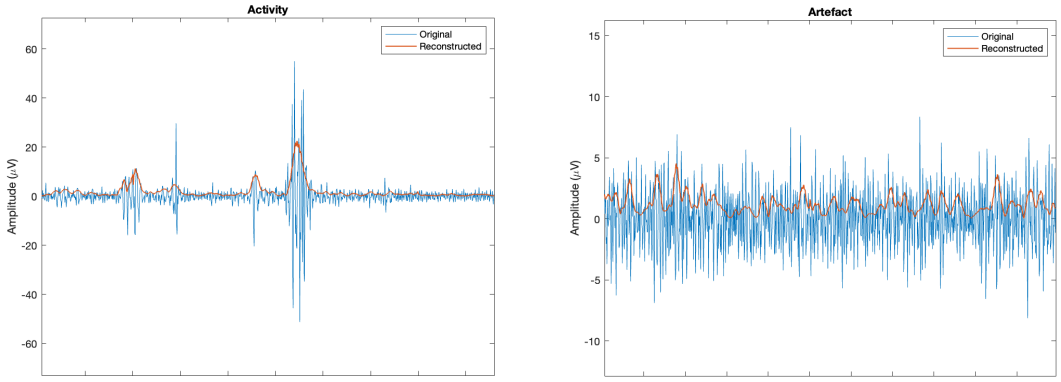


Figure 5.4: Power density distribution for activity and artefact in the frequency range 75–150 Hz (3^{rd} level of detail (sampling frequency: 600 Hz).



(a) (b)

Figure 5.5: Wavelet-based synthesis of a portion of: (a) phasic activity, and (b) breathing artefact. The reconstructed signal is shown in the solid orange line.

More in detail, to objectively capture the extent of similarity between the original and reconstructed signal, a correlation metric was proposed. This latter was defined as the ratio between the cross-correlation of the original signal $x(t)$ and the reconstruction $y(t)$ (Equation 5.2), and the auto-correlation (Equation 5.3).

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t)y(t+\tau)dt \tag{5.2}$$

$$R_{xx}(\tau) = \int_{-\infty}^{\infty} x(t)x(t+\tau)dt \tag{5.3}$$

The index of correlation (5.4) was computed in 0.1-second epochs, and expressed in the range 0 to 1, where values close to 1 represent higher similarity.

$$CI = \frac{R_{xy}}{R_{xx}} \tag{5.4}$$

This metric featured high values (above 0.7) for phasic activity areas, and low values (below 0.5) for artefacts, showing potential as screening tool to for artefacts, by excluding the low correlation segments.

However, one potential drawback lied in the fact that low EMG-activity areas – i.e., areas with no phasic activations nor artefacts – featured high CI values, significantly introducing noise in the ML detection and subsequent RSWA metric assessment.

For this reason, an additional, energy-based metric, was proposed, in order to adequately characterize the areas of interest. The metric was retrieved epoch-wise, by computing the 90^{th} percentile of the integral of the signal, as high-amplitude peaks were mostly found in the top-10% of the signal amplitude. Virtually, this metric functions as a high-pass filter for the energy of the signal, by setting a threshold to discard the low-activity areas.

On the other hand, this approach was not applied in the detection of artefacts versus background activity, due to their morphological resemblance. Instead, the features described in Section 5.3.2 were employed in the ML classification.

Automatic Detection of Artefacts

The extracted features were employed in a supervised ML framework to carry out the automatic artefact detection, in the two following configurations:

1. Artefact *vs* Phasic Activity,
2. Artefact *vs* Background Activity.

The two procedures will be described in this order.

Regarding the configuration **(1)** (Artefact *vs* Phasic activity), the feature set only consisted in the index of correlation and the energy-based metric. Therefore, no feature selection techniques were adopted, as, otherwise, the system would rely only on one feature, switching the problem to a probabilistic approach. In this binary class detection task, five different supervised models were tested. Namely, the following models were tested: SVM, KNN, RF, LDA, AdaBoost (Chapter 3).

Considering configuration **(2)** (Artefact *vs* Background activity), a larger set of features was extracted from the available data. This differentiating choice was made due to the fact that statistical testing of the morphological features yielded promising results compared to the phasic configuration, and this set of features was employed.

Specifically, to trim the initial feature set, preliminary statistical tests were conducted on the extracted features. Feature normality was first tested through the Shapiro-Wilk test [157], which highlighted the normality of the features. Second, the t-Student's independent sample test was applied. Given the high number of significant features, a feature selection approach was adopted to obtain the final

feature set. The adopted approach is the ReliefF algorithm [168], whose results were inspected, and the top 10 ranked features were employed for the analysis, following the elbow-method. The explored classifiers were: DT, SVM, KNN, RF, and LDA.

The two classification tasks were carried out independently.

For both tasks, a hold-out, combined with a Leave-One-Subject-Out cross-validation approach were adopted. Specifically, 72% of the available data (18 subjects) were employed in the training and validation steps, while the remaining 28% was set aside for external validation and RSWA metrics testing. To allow for better generalization capability of the tested models, the model hyperparameters were optimised through a Grid-Search approach, by employing the F1 score as metric for model performance comparison. For each tested model, detection performance was evaluated by means of overall Accuracy, Recall, Specificity, Precision, and F1 score.

5.3.3 Results

This subsection illustrates the results of the research activity.

Data Distribution, Statistical Testing, Feature Selection

Distribution testing of the extracted features, through the Shapiro-Wilk test, highlighted all the extracted features (both in configuration 1 and 2) to have a normal distribution ($p>0.05$).

The results of the independent sample test on the morphological features employed in the configuration **(2)** are shown in Table 5.3; the most discriminative feature ($p<0.001$) for the two types of EMG activations resulted to be SEF25, which virtually represents the 25^{th} percentile of the frequency spectra.

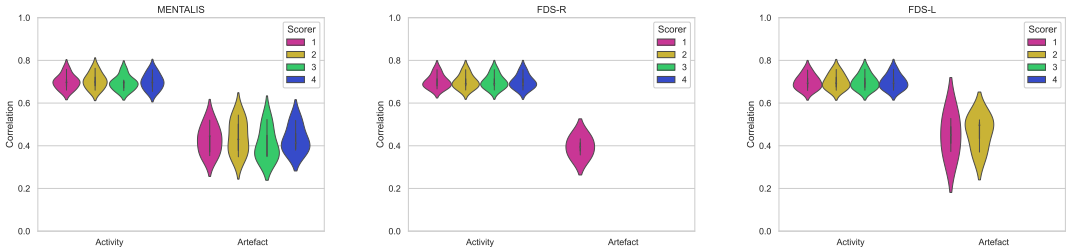
The proposed index of correlation featured a significantly different distribution the two classes, i.e., artefact and activity, respectively, as shown in Figure 5.6. The data distribution is consistent with the manual scores independently provided by the expert scorers. Phasic activity segments featured a correlation index above 0.6, whereas the average value for artefacts was of 0.42.

Performance of Automatic Detection of Artefacts

Automatic detection of artefacts was carried out through Machine Learning Algorithms. As described in the Methods paragraph, after cross-validation, the best model in each configuration was optimized through a grid search approach and later employed in the external validation set, consisting of 7 subjects. The performance metrics of the LOSO-CV are shown in Table 5.4 and Table 5.5 for the two configurations, respectively.

Table 5.3: Top 10 ranked features according to the ReliefF algorithm, in the configuration (2): Artefact *vs* Background Activity. The results of the Mann-Whitney U test are shown. Significance level: *: p<0.05, **: p<0.005, ***: p<0.001.

Feature	**Source**	**Independent Sample Test**
SEF25_{mean}	Mentalis (50–300 Hz)	$< 0.001^{***}$
Kurtosis_{mean}	Mentalis (50–300 Hz)	$< 0.001^{***}$
95pctl_{mean}	Mentalis (50–300 Hz)	$< 0.001^{***}$
75pctl_{mean}	Mentalis (5–10 Hz)	0.005^{*}
5pctl_{mean}	Mentalis (50–300 Hz)	$< 0.001^{***}$
90pctl_{mean}	Mentalis (50–300 Hz)	$< 0.001^{***}$
MEDF_{mean}	Mentalis (50–300 Hz)	$< 0.001^{***}$
AbsPower_{mean}	Mentalis (5–10 Hz)	$< 0.005^{**}$
MEDF_{mean}	Mentalis (5–10 Hz)	$< 0.001^{***}$
Entropy_{mean}	Mentalis (50–300 Hz)	$< 0.001^{***}$



(a) (b) (c)

Figure 5.6: Index of correlation (for activity and artefacts), in the (a) mentalis, (b) right flexor digitorum superficialis, and (c) left flexor digitorum superficialis. Consensus with the four scorers is shown.

Regarding the classification task **(1)**, the highest validation F1 score was obtained with a LDA classifier (89.59%). Overall, the features employed in this classification task proved high discriminative power, leading to an average validation accuracy of across the tested models. Performance metrics reach slightly lower values in the configuration **(2)**, artefact vs background activity, with an average validation accuracy of 73.44 ± 3.27 %. An optimized RF achieved a F1 score of 76,56% in the LOSO-CV approach. This modest misalignment in performance between the two configurations may find its explanation in the fact that the artefacts morphologically resemble the muscle activity under study.

Table 5.4: Classification task: Artefact vs Phasic Activity. Performance metrics of the optimised classifiers, employing a LOSO-CV.

	Accuracy	**Recall**	**Specificity**	**Precision**	**F1**
SVM	84.90 %	90.10 %	77.78 %	83.30 %	86.95 %
KNN	84 %	87.78 %	79.70 %	80 %	85.32 %
NB	81 %	85.70 %	75 %	80.35 %	82.94 %
LDA	89.04 %	91 %	88.2 %	89.18 %	89.59 %
AdaBoost	82.50 %	84.21 %	80.95 %	81 %	82.05 %

Table 5.5: Classification task: Artefact vs Background Activity. Performance metrics of the optimised classifiers, employing a LOSO-CV.

	Accuracy	**Recall**	**Specificity**	**Precision**	**F1**
DT	71.37 %	69.18 %	73.31 %	69.79 %	69.48 %
SVM	76.8 %	80.23 %	73.83%	73.21 %	76.56 %
KNN	73.15 %	74.70 %	71.76 %	70.21 %	75.67 %
RF	77.26 %	76.16 %	78.23 %	75.72 %	76.56 %
LDA	68.63 %	70.05 %	67.35 %	65.66 %	67.79 %

Metrics Validation

The two best classifiers for the two configurations, respectively, were employed in the external test set to independently carry out automatic artefact detection. The test set included 7 subjects, for whom the four scorers previously conducted manual scoring of the selected REM epochs. The trained models were employed to carry out automatic artefact detection in the selected REM mini-epochs.

Finally, the clinical RSWA metrics according to the SINBAR scoring system, namely, phasic and tonic activity, *"any"* activity and SINBAR index (%) were computed in the automatically corrected signals, and compared to the metrics obtained after manual correction by the four scorers (Figure 5.7). The accuracy of the proposed artefact removal method was assessed in terms of Pearson's correlation between the automatic and manually-corrected scores. The highest correlation coefficient (ρ) was obtained for the phasic activity index, with values of 0.96 and 0.82 for the mentalis and FDS muscles, respectively. A slight decrease was observed in the evaluation of the *"any"* activity and SINBAR index, with Pearson's ρ of 0.76 and 0.71, respectively.

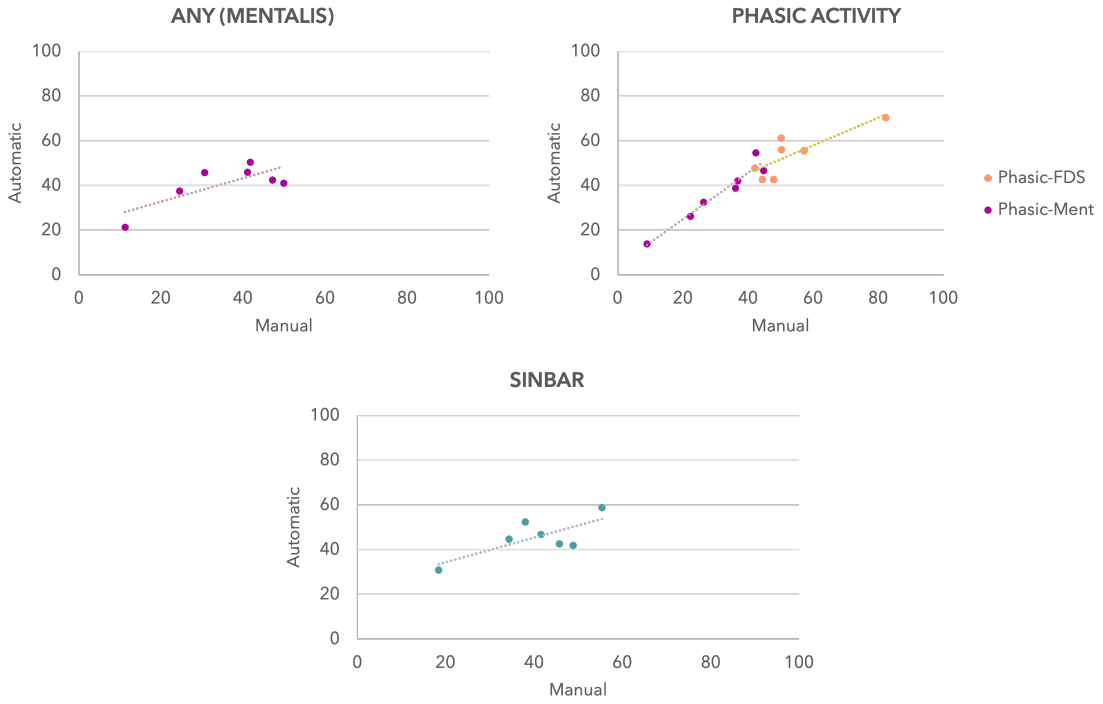


Figure 5.7: Score agreement between the automatic and manually corrected scores. Phasic (FDS, Mentalis), *"any"*, and SINBAR indices are shown.

5.3.4 Discussion

The research activity presented in this Chapter aimed at providing a possible solution for the automatic detection of artefact in the evaluation of REM Sleep Without Atonia, by exploiting the morphology of EMG activations during REM sleep and ML methods. The proposed method relied chiefly on the objective characterisation of EMG activity, to allow for a finer detection of muscle artefacts according to their type. More specifically, a metric of correlation was introduced to distinguish artefacts from phasic activity, and a set of features, characterizing the EMG signal in the temporal, spectral, and non-linear, domains was employed for the detection of artefacts against elevated background tone. The results attained through the presented configurations prove the effectiveness of a low computational cost, automatic method for artefact removal in the evaluation of RSWA. Clinical validation of the RSWA metrics, from a 4-scorer consensus, was achieved with overall Pearson's ρ of 0.81 $\pm$ 0.04. The best agreement was obtained when detecting artefacts from phasic activity, especially in the mentalis and FDS EMG sources. This result is in line with the findings in [31], which highlighted the FDS and submentalis phasic indices as the metrics with higher inter-rater reliability. A slight decrease in agreement was observed for the *"any"* mentalis and SINBAR indices,

which featured higher values of inter-rater variability in [31], even after manual correction. A possible explanation may be found in the fact that some types of artefacts, especially snoring or breathing artefacts, are ofttimes hardly detected from the elevated background tone.

Despite holding promising applicability, the limitations of this study need to be discussed. First, the study included only 25 subjects, due to the high workload required for manual scoring. Future investigations will include a larger population, with an increased number of RBD subjects. Second, being a feasibility study, train, testing, and validation were conducted on on previously selected, manually scored REM epochs. Future works should test the efficacy of the quantification of RSWA without prior epoch selection. Furthermore, the wavelet decomposition process was conducted on 0.1-second epochs. Although this allows for an efficient computational load, other window lengths should be investigated to test for better accuracy in artefact detection.

Future developments of this project should focus on the automatic differentiation of artefacts in REM sleep, due to snoring or breathing, technical interference, or ECG crosstalk. This feature could facilitate fully automatic quantification of REM without atonia without prior manual epoch selection, significantly decreasing the PSG scoring times. Indeed, the optimised framework could be applied to whole-night PSG recordings, thus enabling the fully automatic RSWA quantification according to the SINBAR method, and significantly decreasing the scoring times. Finally, an automatic analysis framework could find clinical implications in the automatic detection of RBD subjects based on the automatic scores.

5.4 Conclusion

Evidence of REM Sleep Without Atonia is largely considered the polysomnographic hallmark of REM Sleep Behaviour Disorder. RSWA manifests with sustained muscle tone during REM sleep, either in the form of elevated background tone (i.e., tonic activity), or periodic bursts, or twitches, of muscle activity. Though abnormal activations are commonly observed in submental recordings, more recently the FDS activity proved to be a reliable alternative, and prone to lower variability in scoring [31].

The current guidelines for the clinical assessment of RSWA require manual scoring of polysomnographic recordings. This entails a long and protracted procedure, ofttimes resulting in tangible inter-rater variability, especially due to the frequent occurrence of artefacts. A solid alternative to this issue has been provided through the semi-automation of the SINBAR assessment [146], with a clinical PSG-integrated software for the automatic detection of abnormal activity in REM sleep. However, activity mis-identification remains an open issue, leading to the necessity of manual correction. Indeed, artefacts observed in the EMG channels are often

misclassified as either phasic, or background activity, depending on the expressed waveform [31]. While manual post-processing provides highly precise scores, it still translates to manual labour and lengthy scoring times.

Following the findings presented in [31], which highlighted the efficacy of manual correction for inter-rater reliability, the research work described in this Chapter aimed at moving a step forward towards the automatic detection of artefacts, to expedite RSWA assessment through full-automated, reliable, and robust computer-based methods.

Based on these assumptions, the experimental activities presented in this Chapter led to the definition of two possible strategies for the automatic detection of artefacts from phasic and background activity, respectively.

While for the latter a straightforward, feature-engineering approach was adopted, with supervised Machine Learning models, the former required additional attention in the signal characterisation. For this purpose, a matched-wavelet approach was proposed and employed for the identification of phasic activity segments, through a biorthogonal mother wavelet. Two novel metrics, representing **(1)** the similarity to the selected mother wavelet, and **(2)** the amount of energy carried by the muscle activation, were proposed and employed as features to discriminate the two components (i.e., artefacts and activity).

Though based on a simple pipeline, the proposed method appeared as a valid tool for the automatic detection of artefacts and subsequent assessment of RSWA metrics, achieving for the atonia metrics a 0.81 ± 0.04 consensus with the manual scores, through Pearson's correlation. Furthermore, the obtained results corroborated the previous findings highlighted by [31], and confirmed the FDS muscle as the EMG source least affected by muscle artefacts.

The results presented in this Chapter hold premise for the implementation of a fully-automated pipeline for the evaluation of RSWA according to the SINBAR scoring system. However, as formerly discussed, the validity of this method on unlabelled, large sets of data, currently remains an open question, as the models were trained and tested on vPSG data from 25 subjects.

Furthermore, the applicability of the proposed novel features in the automatic detection of RBD subjects from all-night vPSG recordings shall still be explored, and is left for future investigations. Indeed, former studies [32] suggested the feasibility of detecting RBD subjects from the analysis of EMG patterns, thus overcoming the amplitude thresholds currently employed alongside with RSWA metrics.

Chapter 6

Automatic Detection of REM Sleep Behaviour Disorder

Within the context of minimally invasive sleep studies, automatic detection of subjects with REM Sleep Behaviour Disorder from physiological signals has been addressed. This Chapter will explore the methods developed to pursue this aim, and included in Papers [139] and [142].

6.1 Context and Background

REM Sleep Behaviour Disorder (RBD) is a parasomnia that coexists with REM Sleep Without Atonia (RSWA), and dream enactment or complex motor behaviour. However, the diagnosis is not straightforward. Among the screening criteria for the diagnosis of RBD (Section 2.2.1) the polysomnographic demonstration of RSWA is required.

As described previously (Chapter 5), the scoring rules for RSWA have require primarily a combination of visual inspection and manual scoring of EMG traces in PSG. Although a computerised version of the SINBAR method is already available [69], manual adjustments are still required, leading to long and protracted scoring procedures.

To overcome this limitation, numerous semi-automatic methods have been proposed in recent years, to assess muscle activity during sleep and aid the detection of RSWA. Generally, a combination of EMG signal processing and rule-based approaches is employed to detect and quantify RSWA, by expressing the latter through continuous indices. In more detail, surface EMG signals, recorded at the mentalis and submentalis, TA and FDS muscles are processed by means of signal processing techniques, in order to highlight the variations over time in amplitude, primarily during the REM stage. As a large majority of these approaches are threshold-based, they will be referred to as *rule-based*. The following Section provides an overview.

6.1.1 Rule-Based and Semi-Automated Methods

In the vast majority of rule-based methods, the EMG activity during REM sleep is measured within epochs of fixed length (1-, 2-, 3-seconds), and then compared to pre-defined values – i.e., *baseline* – commonly identified through the observation of the same source during quiet REM or SWS sleep. Although traditionally, basing on [100], a window length of 2 seconds was adopted to identify phasic activity in RBD, the majority of the semi-automatic methods proposed in the literature rely on a 3-second length window. Namely, these are the Supra-Threshold-REM-Activity-Metric (STREAM), the Frandsen Index (FRI) and the computerised version of the Sleep Insbruck Barcelona (SINBAR) method (cf. Section 5.2); a detailed description of these will follow, and a summary is provided in Table 6.1, in terms of approach and performance.

As mentioned in Section 2.1, in physiological conditions, the activity observed at the mentalis muscle during REM sleep is not expected to exceed the lowest tone observed during NREM sleep [89]. From this assumption, the STREAM algorithm employs as baseline reference value the 5^{th} percentile of the EMG variance in NREM sleep. The EMG signal during REM sleep is analysed in 3-second mini-epochs; the STREAM metric accounts for the percentage of REM epochs with variance exceeding the baseline [24]. A similar approach is adopted in the evaluation of the FRI; in this case, the optimal detection combination, including the baseline, activity, and threshold values, is established heuristically [65]. In this framework, the EMG signal is processed as to obtain an *activity curve* (AC), virtually related to the peak amplitude within the observation window. Motor activity is detected when the AC value is found above the detection threshold, which, in the optimal combination, represented 4 times the selected baseline. The same epoch length is employed in the computerised version of the SINBAR method. Recently, an open-source version has been proposed [146], with an AUC of 0.945 and 0.994 when detecting RBD from the mentalis and combination "mentalis + bilateral FDS", respectively.

A shorter epoch length is adopted in the computation of the REM Atonia Index (RAI), by Ferri et al. [60]. This method has been developed with the aim of quantifying the extent of RSWA, and modelled on an initial cohort of healthy subjects and people with RBD. An improved version was later proposed [61], accounting for the reduction of noise primarily due to motion artefacts, which may be very common in sleep recordings in the mentalis and submentalis muscles. More specifically, the RAI is defined as a continuous value in the range 0 to 1, with 0 representing total loss of REM muscle atonia. The value is retrieved as the ratio of 1 second mini-epochs with amplitude values below 1 µV over the total number of REM mini-epochs, excluding those with amplitude values ranging from 1 to 2 µV, as this EMG amplitude level may result ambivalent, and describing both atonia and muscle activation [60]. A value below 0.8 or 0.9 is indicative of RSWA.

A comparative study [62] was conducted to investigate the detection performance of the RAI alongside with the two commonly employed visual procedures – i.e., Montréal and SINBAR. The study, conducted on PD patients with and without RBD, highlighted the clinical validity of the RAI as a first-line, screening tool for RSWA. Finally, Kempfner et al. [93, 92] proposed a semi-automated method for the detection of RSWA based on the envelope of the EMG signal recorded at different sources, including the submental and tibialis anterior muscles. Following an approach similar to the one adopted for the computation of the FRI [65], the envelope of the EMG signal is observed over mini-epochs of 3-second length, and a feature, virtually corresponding to *on-off* muscle activity, was extracted from each mini-epoch by comparing its amplitude to the observed minimum. Afterwards, the automatic detection of RSWA is proposed in a semi-supervised fashion, through a one-class support vector machine (OC-SVM) [152], trained on data from healthy subjects. The epochs corresponding to abnormal muscle activity are detected as outliers from the algorithm, and a continuous index is computed as the ratio between outliers and total number of observation, and used as a quantitative indicator of muscle activity.

Although presenting with reasonable RBD detection performance (cf. Table 6.1), these methods are generally based on rigid sets of rules and, therefore, rely strictly on pre-defined thresholds to identify RBD-like muscle activity. However, in certain circumstances, such as sleep apneas, the EMG amplitude during REM sleep substantially increases, though without the occurrence of RSWA [29].

It has been shown that manual intervention is still required in many cases; the topic of artefact correction was discussed in Chapter 5. Besides, the presence of RSWA is a required condition in the diagnosis of RBD, but other aspects, often overlooked in rule-based and semi-automated methods – i.e., dream enactment, vocalisations – need to be taken into account [81], as well as peculiar patterns in PSG signals.

To partially overcome these limitations, without the need for human intervention, alternative methods were proposed, exploring Machine Learning approaches to detect abnormal EMG activity during REM sleep.

Table 6.1: Semi-automatic, rule-based methods for the assessment of REM Sleep Without Atonia presented in this Section. For each method, the following are included: EMG source, pre-processing specifications, window length, population under study, performance metric (in terms of area under the curve). The acronym ND indicates subjects with neurodegenerative disorders.

Index	EMG	Filter	Epoch	Sample	AUC	Reference
RAI	Mentalis	10–100 Hz	1 s	35 controls 31 RBD 10 MSA 5 OSAS	0.83	[61, 62]
STREAM	Mentalis	10–70 Hz	3 s	6 controls 23 ND (9 RBD)	0.84	[24]
FRI	Mentalis	10–70 Hz Notch 50 Hz	3 s	29 controls 21 RBD 43 PD	0.81	[65]
Automated SINBAR	Mentalis Bilateral FDS	50–300 Hz	3 s	60 controls 20 RBD	0.93	[69, 62]

6.1.2 Machine Learning for RBD Detection

Different approaches to the automatic detection of RBD through feature engineering and ML were proposed in the literature, relying on various biosignals collected during sleep.

Specifically, Cooray et al. [42] proposed a feature-driven classification pipeline based on EEG, EMG, and EOG signals, achieving a detection accuracy on the test set of 96%. In a following work [43], the Authors propose a similar approach, though disposing of the EEG source, showing significantly high accuracies for the EMG channel alone (90%), and 92% when combining the EMG with ECG features. Other works [37] explored the combination of EEG and EOG channels in detecting subjects with neurodegenerative diseases (PD and RBD), yielding a 91.4% recall in the validation stage.

Other approaches investigated the ability of EMG, or EEG channels alone, in detecting subjects with RBD. Kempfner et al. [93] trained a one-class SVM (OC-SVM) in classifying *outliers* – i.e., subjects with RBD – based on the envelope of the tibialis anterior muscle, reaching a validation AUC of 0.989. Later, Cesari et al. [32] proposed a data-driven, probabilistic method for the evaluation of RSWA based on EMG characteristics, outperforming previous approaches, with a validation accuracy of 81.52 $\pm$ 8.20 %, considering both REM and NREM sleep segments. The method was also validated on an external test set, achieving a RBD-wise accuracies of 84.17% and of 85.60% when excluding PD patients.

Lastly, as some differences were observed in the EEG waveforms when comparing RBD subjects to healthy ones [113], various studies examined the potentiality of detecting RBD through one or more EEG channels. Hansen et al. [75] employed a total of 6 EEG channels (commonly recorded during PSG) to tackle automatic classification through a Bayes probabilistic classifier and a KNN, achieving a 80% recall with the former, after a feature selection step. Likewise, Bisgard et al. [13] attempted RBD detection with the same number of channels, though on a larger cohort and with unsupervised ML models. The Authors attained a 78% accuracy on the test set, though with a slight decrease in recall, reaching a value of 63%.

The vast majority of these approaches rely on biosignals collected through PSG, exploring heterogeneous combinations – e.g., EEG, EMG, EOG – or multiple EEG channels. However, as discussed in Chapter 4, PSG offers very precise diagnostics, but requires cumbersome instrumentation.

While, from a clinical perspective, PSG is fundamental in the assessment of REM atonia, especially in consideration of the fact that higher degrees of RSWA have been highlighted as clinical predictors of RBD with higher probability [117, 115], on the other hand, RSWA does not manifest uniformly, and RBD is likely to progress differently for each subject. This aspect is noteworhy, since RBD is regarded as a precursor to most α-synucleinopathies, achieving a pivotal role in the development of neurodegenerative diseases (cf. Section 2.2.1).

Therefore, while RSWA remains a potential biomarker for risk stratification, in the perspective of providing accessible screening tools for early detection of RBD, simpler configurations shall be investigated, with the aims of **(1)** detecting RBD from a minimal set of sensors, thus assessing the feasibility of outpatient screening, possibly through lightweight technology, and **(2)** propose strategies to quantitatively monitor the progression of RBD.

6.2 Research Overview

Automatic detection of RBD from sleep biosignals provides a possible solution for expediting the diagnostic process.

The research activity described in this Chapter attempted at overcoming the possible limitations of full PSG, by tackling RBD detection from single-source biosignals (either EEG or EMG). The research work was conducted in cooperation with the Regional Centre for Sleep Medicine, Department of Neuroscience, of the Molinette University Hospital (Turin, Italy). This resulted in two main research branches, focusing on muscular activations and brain activity (**(A)** and **(B)**, respectively).

The research activities, detailed in the following Sections, aimed at:

1. Characterising EMG spectral patterns in RBD subjects **(A)**,
2. Detecting RBD from EMG recordings **(A)**,
3. Investigating REM sleep and Slow Wave Sleep in RBD **(B)**,
4. Detecting RBD from a single EEG channel **(B)**,
5. Providing an EMG-based, objective metric to assess the degree of disease progression **(A)**.

6.3 A) Automatic Detection of RBD based on EMG

The first approach tackled the characterisation of EMG patterns in RBD, and subsequent automatic detection through a set of simple metrics.

The work presented in this Section is included in Paper [139].

6.3.1 Materials

This work included both a public and a private dataset. The well-known CAP Sleep Database, publicly available on PhysioNet ([166, 72], `https://physionet.`

org, accessed: 15 March 2021), was employed. The dataset includes PSG recordings of 22 RBD subjects and 16 healthy subjects (HS); the demographics are displayed in Table 6.3.1. In the dataset, 14 subjects had a diagnosis of iRBD, while the remainder were affected by secondary: 6 with PD-RBD, 1 with Dementia with Lewy Bodies, and 1 with Multiple System Atrophy. Quality check was carried out on the recordings prior to the subsequent analysis; one subject in the HS group was discarded as it lacked any EMG recording, and three more were excluded due to the presence of ECG artefact. Subjects were included in the final processing if the recordings presented with more than 5 minutes of scored REM sleep. The private dataset (TURIN Sleep Disorders Dataset, TuSDi) was collected at the Regional Centre for Sleep Disorders in Turin, Italy; this second cohort was employed to assess the robustness of the proposed framework. It includes 18 clinically diagnosed or suspected RSWA subjects (Table 6.3.1), for which the diagnosis was confirmed after PSG. One subject was undergoing treatment for PD at the time of the study. Inclusion criteria for this dataset are provided in the Supplementary Material (Appendix A.1.1); as mentioned in Appendix A.1.1, this dataset is available on request, and was not made publicly available for privacy reasons.

All PSG recordings were manually scored by a sleep technologist, and events during sleep were manually identified. Specifically, the submental EMG channel was screened for elevated muscle tone during REM related to arousals and sleep apneas. These occurrences were labelled as artefacts and excluded from the analysis.

Table 6.2: Sample and Demographics of the datasets employed in the study: the CAP Sleep Database and the TUSDi Database.

Dataset	**Sample (Sex)**	**Age**
CAP Sleep Database	22 RBD (19 males) 16 HS (9 males)	70 ± 6 years 32.5 ± 5 years
TuSDi Database	18 RBD (11 males)	60 ± 2 years

6.3.2 Methods

The following Sections illustrate the Methods adopted in the study, from the extraction of features to the implementation of the Machine Learning pipeline.

Feature Extraction

The feature extraction process regarded two main macro-categories: polysomnographic – i.e., features describing the sleep structure – and EMG-derived features. As regards the former, clinically employed variables [63] were chosen, and computed

from the manually annotated hypnogram. Additional polysomnographic parameters were introduced, following previous work by Cesari et al. [27], covering the substructure of the sleep cycle, and providing information about sleep fragmentation. Specifically, the Sleep Transition Index (STI), REM and non-REM Fragmentation Indices (RFI and NFI), and the average length and proportion of segments classified as belonging to the same sleep stage (cf. Table 6.3).

Regarding feature extraction from EMG data, the recordings from the mentalis muscle were selected, and processed in 1-second epochs; the parameters were extracted in the time and frequency domains.

First, the RAI was computed for each subject; it accounts for the extent of atonia during REM sleep (cf. Section 6.1.1), and is computed from the amplitude of the time series. Then, various parameters from the power spectrum (PSD) of the EMG were extracted. Specifically, the Mean Frequency (an averaged measure which represents the PSD centroid), the Median Frequency, i.e., the Spectral Edge Frequency at 50%, (SEF50) representing the threshold frequency below which 50% of the total power lies, and the Spectral Edge Frequency at 95% (SEF95), i.e., the spectral 95^{th} percentile.

For the sake of clarity, the whole set of extracted features is displayed and described in Tables 6.3 and 6.4, along with their description.

Post-Processing and Feature Selection

As the final set of features encompassed different categories and varying scales, feature normalisation was applied to prevent bias from affecting the classification task. After confirming the normality of the distribution through a Shapiro-Wilk test, Z-score normalisation (Equation 6.1) was applied to all the features; this step ensured that the final feature set had a null mean (μ) and the standard deviation (σ) equalling 1. For any feature f:

$$f_{norm} = \frac{f - \mu}{\sigma} \tag{6.1}$$

Variance threshold feature selection was employed, in order to remove the low-variance predictors in the dataset, and decrease the risk for over- and underfitting. The threshold was the 25^{th} percentile, heuristically selected; all features with variance below that value were discarded.

Machine Learning Classification

This work aimed at performing a binary classification task, between healthy subjects and subjects with RBD, through the use of supervised learning models. Specifically, a KNN and a SVM were employed for this task; both models are built on distance-based algorithms (cf. Chapter 3). To ensure the best possible generalisation capability for this framework, the models hyperparameters were optimised,

Table 6.3: Polysomnographic features employed for the analysis, along with their description and proper reference (if needed). They encompass both clinically employed parameters and polysomnographic patterns.

Feature	**Description**
Sleep Onset Latency (SOL)	The amount of time required to fall asleep (minutes)
Wake After Sleep Onset (WASO)	The amount of time the subject is awake during the recording (minutes)
Total Sleep Time (TST)	Total hours of sleep
Time in Bed (TIB)	Lights-off to lights-on interval (hours)
Sleep Efficiency (SE)	The ratio between TST and TIB (%)
Arousal Index (ARI)	Frequency of occurrence of arousals
Minutes of REM Sleep (MREM)	Total duration of REM Sleep (minutes)
Proportion of N1 Sleep (PN1)	N1 sleep per TST (%)
Proportion of N2 Sleep (PN2)	N2 sleep per TST (%)
Proportion of SWS Sleep (PN3)	SWS sleep per TST (%)
Proportion of REM Sleep (PNR)	Proportion of REM sleep per TST (%)
NREM Fragmentation Index (NFI)	A measure of the number of transitions from NREM to any other NREM stage per hour of NREM sleep [27]
REM Fragmentation Index (RFI)	A measure of the number of transitions from REM to any other sleep stage per hour of REM [27]
Wake Proportion (WP)	Awake time during the night (%)
Sleep Transition Index (STI)	A measure of the number of transitions from REM to NREM (and vice versa) per hours of sleep
Average Length N1 (ALN1)	Average length of N1 segments (minutes)
Average Length N2 (ALN2)	Average length of N2 segments (minutes)
Average Length SWS (ALSWS)	Average length of SWS segments (minutes)
Average Length REM (ALREM)	Average length of REM segments (minutes)

through a Bayesian approach, with 30 iterations. Information about the employed models and the optimisation step is provided in Table 6.5. The models were trained and cross-validated on the CAP Database cohort; to ensure classification robustness, a hold-out approach was adopted. In more detail, 70% of the dataset (23 total subjects, 15 RBD) constituted the training set, and the remainder the test set (11 subjects, 7 RBD). A k-fold CV (k=5) was employed to mitigate overfitting.

Table 6.4: EMG features proposed and employed in this study for the caractherisation of RBD subjects, according to their category. A description is provided.

Domain	**Feature**	**Description**
Time	REM Sleep Atonia Index (RAI) [61]	A measure of the amount of atonia during REM Sleep (1-s mini-epochs)
Frequency	Mean Frequency of REM mini-epochs (MF)	Mean Frequency of EMG signal during REM Sleep, 1-s mini-epochs (Hz)
	Mean Frequency of REM mini-epochs (SEF50)	Median Frequency of EMG signal during REM Sleep, 1-s mini-epochs (Hz)
	Spectral Edge Frequency at 95% of REM mini-epochs (SEF95)	Frequency below which 95% of the total spectral power is found on the EMG signal during REM Sleep, computed on 1-s mini-epochs (Hz)

To assess the generalisation capability of the tested models, an additional external validation step was conducted on the TuSDi Database, so as to provide a set of unseen data to the previously trained models.

Table 6.5: Overview of the hyperparameters optimisation search range, and the optimised configuration, for the two explored classifiers.

	K-NN	**SVM**
Hyperparameters	K: range [1:1:12] *Distance Metric*: City Block, Chebyschev, Euclidean, Hamming, Mahalanobis, Minkowski, Spearman	*Kernel*: Linear, Quadratic, Gaussian, Cubic Maximum Penalty: range [0.001, 1000]
Optimised Parameters	K: 10 Distance Metric: Spearman	*Kernel*: Linear Max Penalty: 2.09

Distance-based Modelling of RBD Progress

As introduced previously, there is no uniform standard to objectify disease progression in RBD, and the latter is likely linked to a higher risk for developing a neurodegenerative disease.

For this purpose, the last part of this experimental work was devoted to propose a distance-based, continuous metric, likely correlated with the degree of dissociation in REM sleep – i.e., RBD progression. This will be referred to as Dissociation Index (DI).

Considering that no *disease* nor *healthy* model was heretofore proposed in the literature, different configurations were explored, considering both the variability of the available dataset and its boundaries, described in Paper [139]. For the sake of brevity, this Section will only describe the final configuration. Additional information on the design of the DI is provided in the Supplementary Material (Appendix A.2).

Let S be a EMG feature vector describing a healthy model – referred to as *reference*, from here after. The DI was expressed as the distance in space between each subjects' EMG features and the reference array. Lower values of this metric indicate stronger resemblance to the reference, and ideally, a null distance corresponds to the identity with the healthy model.

After a tuning phase, S was expressed as the mean of all EMG feature arrays in the HS cohort, thus comprising the muscular characteristics of the HS cohort.

Second, under the same rationale, a *neighbourhood*, defined as R, mirroring the extent of similarity within the dataset was proposed. R was computed as the Euclidean Distance (ED) between two locations in space, and served as *search range* in the next computations. Specifically, the two points represented the intra-class similarity in the HS and RBD group; more details are provided in the Supplementary Material (cf. Appendix A.2).

The obtained neighbourhood was then employed as the distance limit between each subject in the dataset and S. For this purpose, all the subjects within distance R from S were identified, and their actual ED from S was computed (Figure 6.1). This will be referred to as H_i; subjects lying outside the neighbourhood were automatically assigned the boundary value – i.e., R, as per the previous definition.

Finally, after a proper normalisation process, the DI was defined in a continuous range [0, 1], in order to provide a reproducible and easily interpretable metric, as in Equation 6.2. Increasing values of the index indicate higher dissimilarity to the reference model – i.e., virtually representing higher RBD progression; subjects outside the neighbourhood range are automatically assigned the maximum value – i.e., 1.

$$DI_i = \frac{H_i - min}{R - min} \tag{6.2}$$

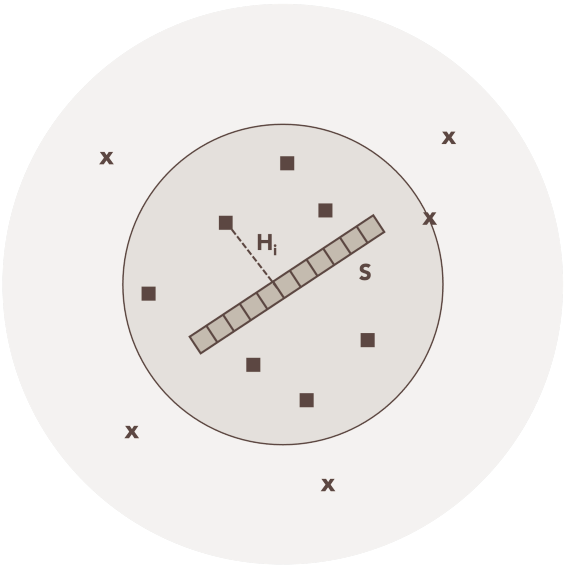


Figure 6.1: Geometry of the framework proposed for the assessment of similarity (simplification). The highlighted circle represents the neighbourhood. The Euclidean Distance H_i is computed for each subject in the neighbourhood (squares). Distant subjects (x) are assigned the maximum value.

6.3.3 Results

The Results of this activity will be arranged in the following Sections as per the Machine Learning classification and the Dissociation Index.

Muscular Features Analysis

This work introduced the spectral analysis of EMG in the automatic detection of RBD subjects, specifically through the mean and median frequencies in REM sleep, and 95^{th} spectral percentile in the same stage. The distribution of the values of each feature in the two different classes (Figure 6.2) showed significant differentiation of the two groups. Spectral features in the HS group presented with a wider interquartile range, which was assumed to reflect the intrinsic variability of the group. Indeed, the HS cohort included healthy, younger subjects, with likely a certain extent of physiological variability. Additionally, three subjects in the HS group presented with values of RAI equalling 0; as values of RAI below 0.8 are already indicative of RSWA [60], this unexpected occurrence, after post-processing and inspection of the recording, found its explanation in the presence of sleep-related breathing disorders, such as sleep apneas and snoring. As the RAI is an amplitude-based, threshold method, and although in [61] a noise-correction method was proposed, the spectrum of sleep-disordered breathing may introduce possible confounding factors that alter its value. The plots in Figure 6.2 represent only HS

subjects with a value of RAI above the cohort median (0.945). On the other hand, the RBD group showed less variability, which suggested good predictive power of this set of features for the class.

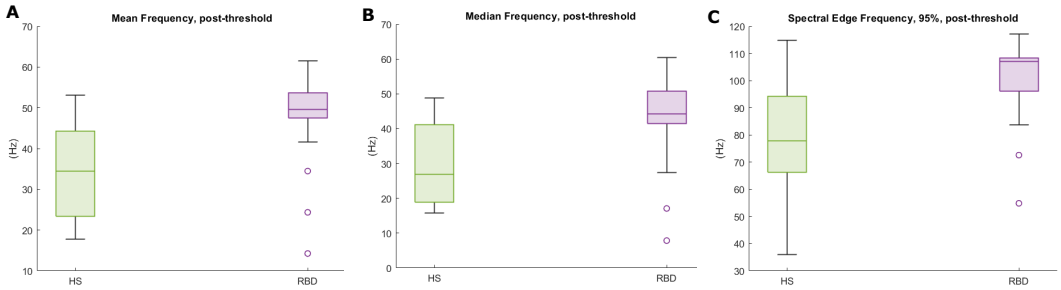


Figure 6.2: Box plot of the spectral features derived from the EMG recording in REM sleep, for the healthy (HS) and RBD participants. Reproduced with permission from [139].

Automatic Classification: Cross-Validation and Test Set Performance

The first part of this research activity tackled a binary classification (healthy *vs* RBD subjects) based on a minimal set of EMG predictors, combined with a set of polysomnographic parameters. The two explored supervised classifiers (a SVM and KNN, respectively) achieved promising training performance. The results of the 5-fold CV are displayed in Table 6.6, in terms of Accuracy, Recall, Specificity, Precision and False Detection Rate (FDR). Both classifiers are based on distance paradigms; the KNN model slightly outperformed the SVM, with an overall Accuracy approaching 87%, and Recall of 93%. Specificity and Precision values were almost comparable in the two classifiers, suggesting good robustness of the employed predictors against false positives. The metrics attained were comparable to the ones in the literature [43], though previous works relied on a higher number of predictors; therefore, the results achieved proved the feasibility of RBD detection from a limited set of muscular parameters.

Table 6.6: Cross-Validation Performance of the KNN and SVM models (5-fold)

	Accuracy	**Recall**	**Specificity**	**Precision**	**FDR**
KNN	86.96 %	93.33 %	75 %	87.50 %	12.50 %
SVM	82.61 %	86.67 %	75 %	86.67 %	13.33 %

Finally, the best model was applied to the held-out cohort (11 subjects, 7 RBD), to validate the previous findings, and attained a 81% Accuracy, with 85.71% Recall,

75% Specificity, and 85% Precision. As expected, a slight decrease in the classification Accuracy was observed; however, from an overall perspective, the model showed good generalisation capability.

Automatic Classification: External Validation Set

In the training and validation stages of the Machine Learning pipeline, the KNN emerged as the best performing model. However, to assess the effectiveness of the proposed muscular predictors and the analysis framework, both models were tested on the TuSDi Database.

Albeit following the recommended PSG montage and guidelines [10], this database included a collection of recordings conducted in a different environment, and likely with different instrumentation as the CAP Sleep Database, thus providing an effective test set. Indeed, as highlighted in [30], ML models in sleep science are seldom tested on datasets recorded at different centres. As the TuSDi database, at that time, did not include any PSG recording from healthy subjects, the classification task shifted to a one-class detection problem. For this reason, the models' performance was evaluated in terms of Average Model Accuracy (AMA) – i.e., 1 - Mean Absolute Error; the average AMA across the two models was of 97.23 %. The KNN classifier achieved almost optimal results, though this is likely related to the small numerousness of the dataset.

Dissociation Index

As mentioned in the Methods paragraph, this work explored various combinations for the R/S configuration; these are described in detail in Paper [139], and provided for completeness in the Supplementary Material (Appendix A.2). For the sake of clarity, this paragraph will only report the final combination. The more generalised neighbourhood was selected, with a value $R = 5.92$; 15 RBD subjects were found within this search radius. The ED (H_i) was then computed for each subject in the CAP dataset (5.01 ± 0.6); the 6 subjects outside the search radius were automatically assigned the maximum distance value (5.81). Then, the DI was retrieved, for both groups; the values distribution is portrayed in Figure 6.3.

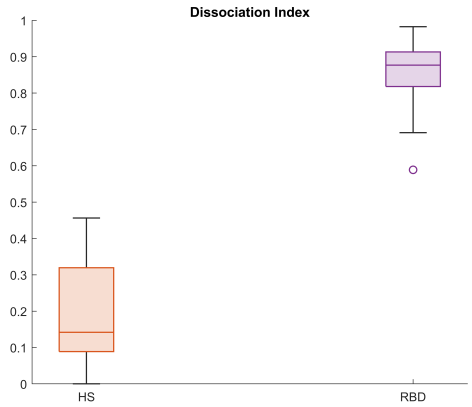


Figure 6.3: Dissociation Index in the CAP Sleep Database. Box plots for the healthy (HS) and RBD groups. Reproduced with permission from [139]

The RBD group expressed lower variability in the DI, with values above 0.7; as expected, the same did not hold for the healthy subjects in the dataset, for which the 75^{th} percentile was of 0.32, indicative of strong similarity to the healthy model.

The rationale behind the DI was attempting to provide an objective metric to quantitatively characterise disease progress in RBD. This concept aimed providing effective screening tools to facilitate longitudinal evaluations, and patient follow-up in RBD. In this perspective, and in view of future validation studies, from the analysis of the DI_{RBD} data, four areas, recalling disease progression were proposed. They are the following (Figure 6.4):

1. **Low Tier**: minimum value (0) to to 75^{th} percentile of the DI in the healthy group ($Q3_{HS}$),
2. **Moderate Tier**: from $Q3_{HS}$ to the 75^{th} percentile of DI_{RBD} ($Q3_{RBD}$),
3. **High Tier**: from $Q3_{RBD}$ to the maximum value (1),
4. **Very High Tier**: Subjects who were not found in the neighbourhood.

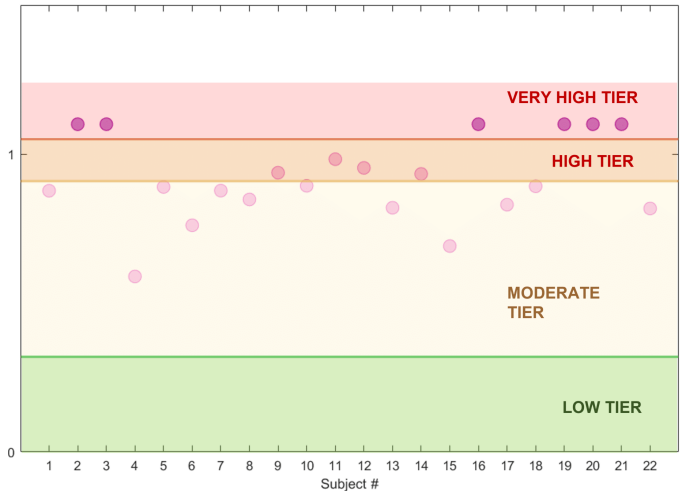


Figure 6.4: The proposed progression areas, and the Dissociation Index of the RBD subjects in the CAP Sleep Database.

6.3.4 Discussion

This research activity aimed at proposing a diagnosis support system for RBD based on simple and lightweight parameters extracted through the submental EMG during REM sleep.

First, automatic detection of subjects with RBD was tackled, by employing Machine Learning models trained on polysomnographic parameters and simple muscular predictors. Second, a distance-based metric (i.e, the Dissociation Index) was proposed, as a follow-up tool to assess RBD severity.

The achieved detection performance in the ML pipeline suggested strong predictive power of the employed parameters, reaching an overall accuracy of 87% with an optimised KNN. External validation on an additional batch of data, recorded in a different Sleep Unit was also performed, with promising performance. The obtained results bolstered the concept of screening for RBD with a minimal set of sensors; specifically, a less invasive PSG paradigm, involving a single EEG channel for sleep scoring (cf. Chapter 4) and an EMG channel to assess REM dissociation.

In this perspective, RBD monitoring might be transferred to unsupervised settings, or home care, possibly with the aid of wearable sensors and pervasive technology, and objective metrics such as the DI could significantly impact the course of follow-up.

Though achieving promising results, this feasibility study bears some limitations. First, the generalisation capability of the employed models might benefit from the inclusion of a larger, and more stratified cohort. Indeed, training on small sets of data breeds the risk of overfitting and the effect of temporal bias; in this study, an attempt to mitigate this may be found in the employment of a held-out set

and the external validation step. Second, the design of the Dissociation Index relies on a linear distance model, supporting the concept of linear disease progression. Currently, the actual trend of disease progression is unknown; this leaves space for clinical validation, with the aid of neurological scales, and possibly exploring REM Sleep Without Atonia in subjects with co-morbidities, or other neurodegenerative disorders and parasomnias.

6.4 B) Automatic Detection of RBD based on EEG

A second approach in the automatic detection of RBD was adopted. The influence of brain activity during REM and SWS in cognitive processes and cognitive longevity has been demonstrated [155]. Previous works investigated the potential of EEG-derived features in revealing RBD, either in combination with other physiological predictors [26, 42, 43] or by employing EEG alone [75, 13, 49]. Polysomnographic alternations in sleep patterns were observed in RBD subjects, who featured an increase of δ-waves density with respect to age-matched healthy subjects [113]. The research activity presented in this Section proposed a framework for automatic RBD detection based on EEG data during REM and SWS sleep collected from a single EEG channel, to investigate the predictive power of the two stages for RBD, and, once more, explore the feasibility of minimally-invasive studies.

The work presented in this Section was included in Paper [142].

6.4.1 Materials

This work included the healthy and RBD subjects from the CAP Sleep Database and additional subjects collected at the Regional Centre for Sleep Medicine (Turin, Italy), in the updated TUSDi Database; written informed consent for observational study was obtained by all participants. Inclusion criteria for the Turin cohort are provided in the Supplementary Material (Appendix A.1.2).

A total of 58 subjects (32 RBD) were included across the two datasets; the participants' demographics are provided in Table 6.4.1.

PSG recordings from the TuSDi cohort were manually scored by a sleep expert, following the AASM standards [11]; likewise, sleeps scoring according to the same standards is provided in the CAP Sleep Database.

Table 6.7: Sample and Demographics of the datasets employed in the study: the CAP Sleep Database and the TUSDi Database (this latter updated from the previous study. Healthy subjects and additional RBD subjects were included).

Dataset	Sample (Sex)	Age
CAP Sleep Database	22 RBD (19 males)	70 ± 6 years
	16 HS (9 males)	32.5 ± 5 years
TuSDi Database	10 RBD (8 males)	62 ± 6 years
	10 HS (6 males)	37 ± 16 years

6.4.2 Methods

Pre-Processing and Feature Extraction

This study envisaged RBD detection from a minimal EEG configuration; therefore, the analysis was based on data from the C3-A2 channel (or C4-A1, if the former was not available). REM and SWS segments were retrieved from manually annotated data, and exploited for feature extraction. In the perspective of implementing this classification pipeline in wearable, low-computational cost frameworks, minimal signal pre-processing was conducted on the EEG signals. Specifically, high-frequency noise – which, for this study, was identified in frequency components above 40 Hz – was discarded through an IIR Chebyschev Type 1 filter. Finally, mean amplitude removal was employed to eliminate the amplitude offset introduced by the DC, and highly present in the recordings included in the Turin cohort.

The extraction of features, to be implemented in the automatic detection pipeline, encompassed three main categories:

1. **Polysomnographic features**: a set comprising clinical parameters and variables representing sleep architecture and the degree of sleep fragmentation,
2. **Electroencephalographic features**: quantitative predictors describing the EEG signal characteristics in the time, frequency, and complex domains,
3. **Sleep substructural features**: EEG-derived variables, virtually describing the sub-architecture of the REM and SWS stages, respectively.

These will be detailed in the following paragraphs.

Polysomnographic features were extracted from whole-night, manually annotated data, and were displayed in Table 6.3 in the previous Section. An additional feature, describing the amount of minutes spent in SWS (MSWS), was included in this set.

Quantitative electroencephalographic features were extracted on the Time, Frequency, and Non-Linear domain, by processing the EEG in this fashion:

1. 30-second epochs for features in the time domain,

2. 2-second epochs for features in the spectral and non-linear domains.

The epoch lengths, widely employed in the literature, were chosen (1) to match the AASM scoring criteria for sleep stages, (2) to ensure wide-sense stationarity of the EEG signal over the observation interval.

Precisely, considering the intrinsic differences of the REM and SWS stages (2.1), specific metrics in the time domain were computed to accurately describe their waveforms, from the EEG signal and its first-order derivative.

First, normalised slope descriptors (Hjorth Parameters, [122]), were extracted; they provide a description of the underlying energy of a signal while maintaining low computational cost. Namely, they are the *activity* (ACT), *mobility* (MOB), and *complexity* (COMP). Though computed from time series, they virtually capture the statistical properties of the power spectrum – total power, standard deviation of the spectrum, and total bandwidth, respectively.

Given a 30-second epoch $y(t)$, and its derivative $y'(t)$, they are computed in this fashion:

$$ACT = var(y(t)) \tag{6.3}$$

$$MOB = \sqrt{\frac{var(\frac{dy(t)}{dt})}{var(y(t))}} \tag{6.4}$$

$$COMP = \frac{mob(\frac{dy(t)}{dt})}{MOB(y(t))} \tag{6.5}$$

Second, in an attempt to describe the regularity of the two considered stages, the *form* (FF), *crest* (CF), and *impact* (IF) factors were retrieved. These three variables are commonly employed in sound analysis, and serve to quantify the regularity of a wave by comparing its amplitude metric to a perfect sinusoidal wave.

Following the same definition as the previous paragraph, they are described as:

$$FF = \frac{y_{RMS}}{|y|_{mean}} \tag{6.6}$$

$$CF = \frac{y_{peak}}{y_{RMS}} \tag{6.7}$$

$$IF = \frac{y_{peak}}{|y|_{mean}} \tag{6.8}$$

As regards the frequency domain, the PSD was estimated from each 2-second epoch through a Welch modified periodogram, with 50% overlap with a 1-second Hamming window. These parameters were selected so as to provide an adequate

spectral resolution (1 Hz) for the EEG segments [124]. To have a proper representation of the signal spectrum, the percentiles (25^{th}, 75^{th}, and 95^{th}) and their differentials (75–25, 95–25, 95–50) were computed, so as to comprise the relevant data distribution information.

Additionally, given the intrinsic properties of the EEG signals during sleep, the Absolute and Relative power were computed for each clinically relevant band. For the sake of clarity, due to the choices in signal pre-processing, the γ band only depicted the range 30–40 Hz.

The Teager-Kaiser Energy Operator, a non-linear metric accounting for the total energy of a signal, was computed, in a similar fashion as in Paper [141].

The whole set of EEG features is provided in Table 6.8.

Table 6.8: Employed features, along with the domain and proper reference. ⋄: variables adapted from the cited study, †: variables first proposed in this work.

Category	**Feature (Name and description)**	**Reference**
Time	Amplitude metrics: mean, standard deviation, skewness, kurtosis, range, maximum and minimum value	various
	Zero Crossing Rate	[165]
	Hjorth Parameters	[122]
	Percentiles (25^{th}, 75^{th}, 95^{th})	various
	Form, Crest and Impact Factors	various
	Coastline	[177]
Frequency	Fast Fourier Transform: numerical and statistical measures (mean and median frequencies, total power, ...)	various
	Spectral Edge Frequencies (SEF25, SEF75, SEF95)	⋄[141]
	Spectral Edge Frequencies differentials ($75-25$, $95-25$, $95-50$)	†
	Absolute Power for each clinically relevant band ($\delta, \theta, \alpha, \beta, \gamma$)	various
	Relative Power for each clinically relevant band ($\delta, \theta, \alpha, \beta, \gamma$)	various
	Entropy measures	⋄[141]
Non-Linear	Teager-Kaiser Energy Operator: numerical and statistical measures	⋄[88]

Finally, a set of metrics to describe the substructure of the REM and SWS stages were extracted; they are listed in Table 6.9. Namely, Absolute and Relative power, and statistical measures on the power spectrum were retrieved from the sub-categories in each stage.

Following the rationale in Paper [141], and described in Chapter 4, the REM stage was treated as having *tonic* and a *phasic* components, depicted in the frequency bandwidths of 2–8 Hz, and 7–16 Hz, respectively.

Slow-wave sleep segments were treated in a similar fashion, and two main spectral components were highlighted; namely, the slow oscillations (SO), and the slow-wave activity (SWA), following the definitions in [14]. In more detail, the former lie below 1 Hz, and the latter account for brain activity observed in the 1–4 Hz range – therefore depicting some degree of overlap with the δ-band. This choice was made as evidence has shown the role of SWA in neurodegenerative processes related to the development of Parkinson's Disease [153].

Table 6.9: Sleep Substructure Features, extracted from the spectral properties of each considered sub-band. ◇: adapted from the cited study, †: first proposed in this study.

Sleep Stage	**Feature**	**Reference**
REM stage	Absolute and Relative Power in TREM, FREM	◇[141]
	Mean, Median Frequencies and Spectral Percentiles (SEF_x) in TREM, FREM	◇[141]
	Total Power Ratio TREM/FREM	◇[141]
Slow Wave Sleep	Absolute and Relative Power in SOs, SWA	◇[153]
	Mean, Median Frequencies, Spectral percentiles (SEF_x), statistical measures in SOs, SWA	†

To explore all possible configurations, and investigate the influence of each stage in RBD prediction, as well as the most relevant contribution to automatic detection, the presented variables were extracted separately on three separate data batches, to retrieve the following feature sets.

1. **FSet**$_1$: including polysomnographic features and EEG features extracted from the REM stage,

2. **FSet**$_2$: including polysomnographic features and EEG features extracted from the SWS stage,

3. **FSet$_3$**: including polysomnographic features and EEG features extracted from the combination of REM and SWS stages.

Post-Processing and Feature Selection

As described in the previous Section, to provide a scale-homogeneous dataset for training, z-score normalisation (Equation 6.1) was applied to the three feature sets.

Due to the high number of extracted predictors, the feature selection step was also deemed necessary; a minimal-optimal approach was adopted, through the Minimum Redundancy Maximum Relevance (mRMR) [129].

This technique aims at finding the optimal subset of features, by maximising target relevance, or dependency (D_T), and minimising inter-correlation (R). Given that the computed features are continuous, target-*dependency* is assessed through the F-statistic. On the other hand, inter-feature *redundancy* is evaluated in terms of Pearson's correlation. Finally, the importance score is computed by means of the mutual information quotient, as:

$$MIQ = \frac{D_T}{R} \tag{6.9}$$

To obtain the final feature configuration for the binary classification task, feature rankings were inspected through the elbow method, and the top-5 features were selected. These are displayed in Table 6.10.

Machine Learning Classification

Pursuing an approach analogous to the previous Research Activity ((A), Section 6.3), supervised ML models were exploited to carry out the automatic detection of RBD subjects.

The constructed feature sets were employed independently with five different classifiers; specifically a SVM, KNN, Naïve-Bayes classifier (NB), a decision tree (DT), and the Bootstrap Aggregating ensemble method (BAG).

However, to explore an alternative configuration, the validation procedure relied on a 5-fold CV approach. Contrarily to Research Activity (A), no external validation set was available; nevertheless, to mitigate the potential effect of overfitting, due to the scarce numerousness of the dataset, the CV procedure was iterated 10 times. Furthermore, to limit data leakage, this approach took into account the ID of each subject.

6.4.3 Results

This Section will illustrate the results of the proposed classification framework, according to the three explored configurations (FSet$_i$).

Table 6.10: Features employed for the classification, selected with the mRMR approach. For $FSet_3$: $\star$ REM features, $\circ$ SWS features

Feature Batch	Top-5 Features
$FSet_1$ (PSG + REM)	Relative Power (α) Minutes in REM Sleep WASO Mobility (2^{nd} order) SEF75
$FSet_2$ (PSG + SWS)	Relative Power SWA (75^{th} pctl) Relative Power (θ), STD Relative Power (α), STD Minimum Amplitude Median (75^{th} pctl)
$FSet_3$ (PSG + SWS + REM)	Relative Power SWA (75^{th} pctl), $\circ$ Relative Power (θ), STD, $\circ$ Mobility (2^{nd} order), $\star$ Relative Power (α), $\star$ Minimum Amplitude, $\circ$

Automatic classification: REM Subset

The first data batch (FS_1) comprised PSG features and features extracted from the REM segments. Table 6.11 reports the classification performance of the employed models, in terms of Accuracy, Recall, Specificity, Precision, F1 score, and Area Under the Curve (AUC). The models achieved a macro-averaged accuracy of 80.29% ± 0.03. An optimised KNN emerged as best model, with Accuracy: 83.91% ± 0.81, Sensitivity: 86.46% ± 2.95; the optimised parameters were number of neighbours (K) equalling 3 (in a search range 1–29), and Chebyschev distance as employed distance.

Automatic classification: SWS Subset

The second feature set ($FSet_2$) explored the predictive power of SWS in detecting RBD subjects in the cohort under study. Table 6.12 displays the classification performance. A macro-averaged accuracy of 81.10% was attained, showing a slight increase in performance with respect to $FSet_1$. An optimised SVM emerged as best performing model, with an accuracy of of 86.21% ± 2.11, recall of 91.23% ± 5.24; the model featured a cubic kernel and a maximum penalty (C) of 2.56 (range 0.001–1000). An increase in AUC with respect to $FSet_1$ was also observed, with a value of 0.94 ± 0.02.

Table 6.11: Performance metrics (%) of the employed classifiers as regards $FSet_1$ (PSG + REM features).

	SVM	**KNN**	**NB**	**DT**	**BAG**
Accuracy	81.03±0.5	83.91±0.81	74.14±1.72	82.18±0.81	80.17±0.86
Recall	84.38±3.13	86.46±2.95	73.44±1.56	85.42±5.31	82.81±1.56
Specificity	76.92±3.85	80.77±3.14	75±1.92	78.21±7.2	76.92±0.1
PPV	81.94±1.94	84.78±1.75	78.33±1.67	83.28±4.25	82.16±0.91
F1	83.06±0.52	85.55±0.9	75.81±1.61	84.07±0.87	82.16±0.91
AUC	0.87±0.05	0.87±0.01	0.76±0.02	0.83±0.03	0.89±0.02

Table 6.12: Classification performance (%) of the employed classifiers as regards $FSet_2$ (PSG + SWS features only).

	SVM	**KNN**	**NB**	**DT**	**BAG**
Accuracy	86.21±2.11	80.46±4.94	78.74±4.30	79.52±7.21	80.60±3.31
Recall	91.23±5.24	83.71±12.26	76.67±6.73	80.83±9.78	79.13±5.58
Specificity	83.36±1.57	80.83±2.94	81.16±3.15	79.06±6.25	82.19±1.92
PPV	76.92±2.72	74.36±7.90	76.92±5.44	74.04±6.72	77.88±3.19
F1	83.36±2.38	77.58±3.58	76.50±4.14	77.04±7.31	78.34±2.98
AUC	0.94±0.02	0.85±0.02	0.81±0.02	0.77±0.09	0.9±0.02

Automatic classification: REM+SWS Subset

Finally, in view of stage-agnostic, unsupervised, sleep studies, the combination of REM and SWS was also explored. This combination attained a significant increase in detection performance (cf. Table 6.13), reaching an average accuracy of 85.70% ± 0.04, with almost +4% rise on the two distinct datasets. The best performance was achieved through a DT, with maximum number of splits of 4, and cross-entropy as split criterion. Specifically, an accuracy of 90.80% ± 0.8, and recall of 95.83% ± 2.95 were obtained.

6.4.4 Discussion

This research activity tackled automatic detection of subjects with RBD based on quantitative metrics derived from a single-EEG channel, after minimal signal pre-processing. The proposed framework aimed at assessing the feasibility of unsupervised sleep studies, in the perspective of screening tests conducted with minimal, and possibly wearable, instrumentation.

The analysis relied on supervised ML models, trained on data from the REM and SWS stages, respectively, and a combination of the two, in an attempt to

Table 6.13: Classification performance (%) of the employed classifiers as regards $FSet_3$ (PSG + REM + SWS features).

	SVM	KNN	NB	DT	BAG
Accuracy	89.08±1.63	85.34±2.59	79.31±1.72	90.80±0.8	83.62±0.86
Recall	92.71±2.95	93.75±0.4	82.81±4.69	95.83±2.95	85.94±1.56
Specificity	86.42	75±5.77	75±1.92	84.62±3.14	80.77
PPV	88.11±0.33	82.33±3.38	80.32±0.32	88.54±1.82	84.61±0.24
F1	90.33±1.56	87.63±1.92	81.47±2.11	91.99±0.79	85.26±0.89
AUC	0.98±0.01	0.92±0.03	0.82±0.03	0.92±0.02	0.93±0.05

determine the influence of the two stages in the disease.

All explored classifiers achieved reasonably good classification performance. Specifically, the two stages presented independently with good prediction performance, achieving accuracies above 80% in both cases, as well as averaged F1 scores of 82.2% and 78.4%, for the REM and SWS sets, respectively. This metric provides a more complete explanation on the generalisation capability of the classifiers, as the original dataset presented with a slight degree of imbalance towards the RBD class, both in terms of sample size, and age.

Additionally, a considerable rise in performance was observed when considering both stages, yielding an accuracy up to 91%, and 92% F1 score (Figure 6.5). This result gains a promising implication in the perspective of implementing stage-agnostic classification tasks, and preliminary investigations on this topic are being conducted.

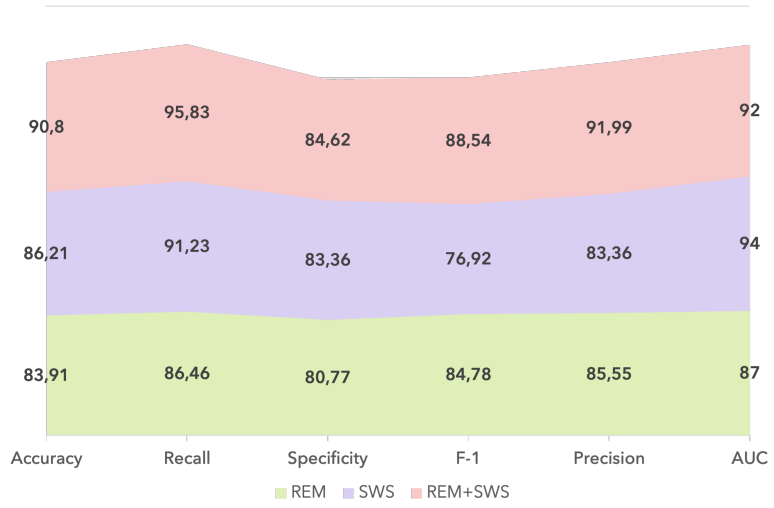


Figure 6.5: Performance comparison of the best model from each explored FSet.

Furthermore, features from deep sleep appeared to have a greater impact on the target. Indeed, the results of the feature selection procedure highlighted only EEG metrics in $FSet_2$ and $FSet_3$. Contrarily, in $FSet_1$, two PSG features were ranked among the most important. This finding found previous evidence in the literature [22], where, in a stage-agnostic feature extraction step, the δ-range emerged as the the most informative for detecting RBD subjects against healthy ones, employing the CAP Sleep Database [166]. When comparing the results of the classification framework proposed in this Section to [22], it was possible to observe with FSet3 an increase of 3% percentage points in accuracy, alongside with a considerable rise in specificity (+26 percentage points), and comparable results when limiting the dataset to $FSet_2$.

Despite the encouraging perspective suggested by the results obtained, this research activity presents with some limitations, that future developments shall address.

First, as already stated for the EMG-based detection of RBD (cf. Section 6.3), the size of the explored sample is quite limited. In particular, a publicly available, well-known resource was included, and during the project, efforts have been made to extend the participants size as regards the Turin cohort. Furthermore, the sample included a large majority of male participants; this, however, reflects the demographic prevalence of RBD. Future recruiting campaigns shall also target the inclusion of age-matched controls, to mitigate bias.

An additional weakness may be found in the fact that only fixed epoch lengths, as per the AASM standards, were investigated for the analysis. This approach relies on the premise that human sleep is largely treated in a *discrete* fashion, leaving much to be explored in the transitions between stages, or in intra-stage variability. Future work should investigate additional epoch lengths, or treat data extraction and sleep transitions in a continuous-wise approach.

6.5 Conclusion

REM Sleep Behaviour disorder is a parasomnia that is considered a prodrome to overt α-synucleinopathies, with a rate of phenoconversion of 67.5–73.5% in a range of 10 to 12 years from the initial diagnosis [181, 134].

Presently, the diagnosis of RBD is left to clinical interviews, polysomnographic evidence of REM Sleep Without Atonia and the documentation of behavioural events through video-polysomnographic recordings.

Although efforts have been made to expedite the scoring process for RSWA (cf. Chapter 5), by proposing automatic algorithms for the identification of abnormal activity during REM sleep, a large majority of these methods still require human intervention. Hence, the clinical diagnosis remains an intricate process, entailing considerable professional workload, and ofttimes prone to variability across

raters.

Most importantly, standard, in-hospital PSG is still required, leading to intrusive examinations, and de facto protracting the actual diagnosis, which is simultaneously hampered by long waiting times. These premises foster the concept that current sleep study methods are not suitable for population screening.

Recent research shed light on the feasibility of minimally-invasive strategies for monitoring sleep disorders, either relying on minimal sets of sensors, or on wearable technology, to be performed in familiar environments [97]. Accordingly, indentifying digital biomarkers for the early detection of sleep disorders, possibly through the use of simplified devices, may lead to significant opportunities in healthcare and mass screening studies.

The research activities presented in this Chapter aimed at providing possible diagnosis support tools, focusing on the automatic detection of RBD through the identification of simple, lightweight metrics to characterise the parasomnia. Polysomnography studies from both a public repository, and from subjects collected at the Centre for Sleep Disorders in Turin (Italy) were explored. Data from electromyographic and electroencephalograpic recordings were exploited independently to tackle the automatic detection of subjects with RBD through the use of supervised Machine Learning models.

The outcome of the research activities highlighted in this Chapter suggest the feasibility of detecting the presence of RBD from information retrieved by muscle activity during REM sleep, or electroencephalographic patterns during both the REM and the deepest stage of sleep.

Indeed, as regards the EMG-based automatic classification, simple metrics describing the spectral distribution of muscle activity were proposed, achieving a classification accuracy of almost 87% in a 5-fold cross-validation, and 81% on the test set. An attempt to external validation was also conducted, reaching high values of average model accuracy.

Similarly, the proposed EEG-based detection pipeline provided promising performance, indicative of the feasibility of assessing RBD from a single EEG channel. This could possibly imply a less intrusive framework. As regards the optimised models, the explored configurations attained AUC values of above 0.85, reaching up to 0.92; accordingly, reasonably high values of F1 score, with a maximum of 92%, were reported.

Finally, most semi-automatic methods are limited to the assessment of elevated muscle tone during REM sleep, but do not provide any depiction of RBD progression. Monitoring the longitudinal development of RBD might benefit follow-up procedures, and, in future perspectives, help shed light on phenoconversion mechanisms. To this purpose, a straightforward, distance-based metric was proposed, in a further attempt to explore unsupervised disease modelling. The Dissociation Index, though only prototyped, showed encouraging potential in portraying the degree of dissociation from the healthy status; however, further investigations are

imperative.

These findings indicate the feasibility of lightweight screening tools, and provide an insight into the possibility of minimally-intrusive methods for diagnosing sleep disorders. Indeed, Machine Learning assisted methods show good potential in facilitating early detection and follow-up procedures, revealing a beneficial effect on the quality of life.

Nevertheless, and as previously discussed, these studies are not without limitations.

First, the employed datasets – although one of them is widely employed in research – feature a quite limited size, which eventually led to some degree of class imbalance, in terms of age (the healthy controls were significantly younger) and sex (prevalence of male participants). Although the obtained performances compared well with the literature, further investigations and validation procedures should definitely address this issue, and include a higher number of participants, possibly with efficient stratification, to ensure the models' robustness and generalisation capability. In addition, further work is also needed to assess the clinical validity of the proposed Dissociation Index. To this aim, preliminary estimations were conducted, and additional assessments on a larger cohort, comprising PD, RBD, and PD-RBD patients, are being conducted. Currently, the validation procedure, discussed with expert neurologists and sleep physicians, relies on the manual assessment of PSG records of the mentioned subgroups and the longitudinal, retrospective analysis of the included subjects. REM Sleep Without Atonia scores are computed according to the Montréal [100] and SINBAR [68] assessment methods, as well as the RAI [61]. Then, the longitudinal scores are compared to the DI values, and their statistical correlation is explored.

Second, the employment of EMG-based data only might lead to inaccuracies when co-morbidities affecting muscle activity, or resulting in elevated background tone, are present. For this purpose, particular attention is being given to subjects with OSA. Preliminary investigations are underway, both in terms of multiclass detection (healthy, RBD, OSA, RBD+OSA), or cascade-wise detection – i.e., (1) healthy vs non-healthy clustering, (2) group-specific classification (RBD, OSA, RBD+OSA). In more detail, both EMG and ECG data collected during REM and NREM sleep are being employed, to explore the predictive power of various physiological components, and assess the detection capability of EMG alone, compared to other PSG sources. However, the experimental outcomes of this investigation are still preliminary and therefore were not included in this dissertation.

Third, the presented approaches mainly relied on a data-driven paradigm of feature engineering in supervised Machine Learning frameworks. This choice was initially made to offer an interdisciplinary understanding of the results, while maintaining the clinical interpretability of the extracted features, in view of the development of a reliable diagnosis support tool. Although a considerable number of models were tested, other opportunities, such as unsupervised learning, should be

explored, possibly to enhance patterns in data that fail to be represented in the standard feature engineering process.

Finally, although Deep Learning applications in the field of RBD are currently limited, and mostly directed at identifying phenoconversion patterns either through temporal series [147], or imaging techniques [172], future trajectories might explore this encouraging frameworks. Indeed, automated video-based analysis of RBD jerks [34], or contextual learning techniques, might offer the possibility of improving diagnostic accuracy of RBD, although, at present, their robust implementation is hampered by data scarcity.

Part II

Monitoring Systems

Chapter 7

Monitoring Sleep in Parkinson's Disease

Sleep in Parkinson's Disease entails a complex fingerprint, including motor-related disturbances which often elude clinical assessments. Besides, these latter are generally scheduled on a long-time basis, and rely on subjective evaluations and self-reports, making sleep disorders often overlooked. Pervasive health solutions offer the possibility to bridge the temporal gap between in-person assessments, and provide better quality of care.

This Chapter presents a possible framework for the continuous monitoring of motor disturbances in sleep, included in Paper [140].

7.1 Context and Background

Parkinson's Disease (PD) is the second-most prevalent chronic neurodegenerative disease, with increasing incidence (cf. Section 7). Currently, the available therapeutic approaches are chiefly devoted to alleviating motor manifestations. However, the definition of an optimal therapeutic approach remains a challenging task, due to the variety of symptoms and co-morbidities in each individual.

As discussed in Section 7, PD is characterised by both cardinal motor and non-motor symptoms. Among these latter, sleep disorders are the most prevalent, being diagnosed in up to 90% of people with PD [162], and occurring with a prevalence rate of 50–70% already in the earliest stages of the disease [12].

Sleep disorders (SD) in Parkinson's Disease entail multi-factorial manifestations, including insomnia, severe exhaustiveness, and excessive daytime sleepiness [17], eventually causing an increase in depression rates and functional disability. The disorders include also motor manifestations, such as nocturnal hypokinesia and morning akinesia, affecting the ability to turn in bed, or getting up, reflecting a significant impairment of axial movements [161]. These manifestations, though

profoundly debilitating, are hard to monitor, and often overlooked, as in-person assessments primarily rely on self-reports.

The lack of a clinically validated, objective scale introduces the need for continuous monitoring of sleep disturbances through accurate and objective parameters. Sleep actigraphy and wrist-worn inertial devices initially emerged as promising tools to record overnight motility and provide objective metrics for sleep patterns evaluation [108, 96]. However, as they are commonly placed on the non-dominant or on the least-affected wrist, they suffer from poor sensor positioning, and fail to properly describe nocturnal immobility, or difficulty in turning in bed. In literature, an alternative set-up based on inertial units positioned near the centre of mass [161, 167] or on the lower back [121] proved robust against motion artefacts, and suitable for describing disordered axial movements during sleep.

Upon these premises, this Chapter seeks to provide a possible solution for the effective and accessible monitoring of sleep disturbances in PD, in order to **(1)** assess the feasibility of remote assessments, and **(2)** provide kinematic descriptors of sleep, to facilitate the investigation of symptoms fluctuations. The research activity, detailed in the Sections below, introduces a framework for automatic detection of motor-related sleep patterns in PD, and the classification of sleep quality using objective metrics derived from wearable inertial units.

7.2 Research Overview

As discussed above, sleep disorders in PD are hugely affected by circadian fluctuations and, at present, their monitoring relies primarily on subjective assessments. The development of pervasive strategies for health offers an encouraging scenario for the objective characterisation of sleep patterns in people with neurodegenerative diseases.

Based on the promising results of a previous co-authored study, conducted on a healthy cohort [3], with a similar purpose, the research work presented in this Chapter sought to propose a lightweight and low-cost wearable set-up for monitoring sleep quality and motor manifestations in Parkinson's Disease. The activity is part of an observational study started in March 2021, in cooperation with the Parkinson's Unit, Department of Neurology, of the Molinette University Hospital (Turin, Italy).

Particularly, the study aimed at:

1. Objectively characterise motility during sleep,
2. Assess perceived sleep quality through simple metrics and validate the paradigm with clinical scales,
3. Identify a set of parameters most suitable to describe sleep in Parkinson's Disease.

7.3 Monitoring Sleep in Parkinson's Disease

The lack of a clinically validated scale introduces the need for continuous monitoring of sleep disturbances through accurate and objective parameters. Although actigraphy provides a valid solution for the assessment of diurnal and nocturnal motility over an acceptable range of time, it suffers from poor positioning. This work focused on the definition of a set of inertial-derived, simple metrics, to objectively characterise motion during sleep, and allow for continuous monitoring of sleep patterns in Parkinson's Disease when integrated in pervasive health scenarios.

7.3.1 Materials

Experimental Protocol and Instrumentation

The experimental protocol of this study was outlined with expert sleep neurologists at the Parkinson's Unit, and aimed at meeting both clinical and technological needs. Chiefly, it sought to define a minimally-intrusive set-up to ensure accurate data collection for characterising overnight motility, and capture kinematic patterns in individuals with PD.

Since data collection adopted an unsupervised approach conducted at the subjects' homes, all participants received thorough instructions on using the device and completing sleeping questionnaires. In addition, a written user manual, containing detailed information about the experiment, and the correct device usage, was provided.

The sleep test was conducted over one night. The experimental pipeline consisted in the following:

1. Completing the clinical questionnaire for sleep quality,
2. Positioning and activating the sensor before bedtime,
3. Annotating *lights-off* and *lights-on* time (these two virtually representing the bedtime onset and waking-up), and turning off the sensor,
4. Fill in the circadian rhythms survey,
5. Record vocal samples while reading a phonemically-balanced text[1].

The participants' screening was conducted by Neurology residents and attendings, among outpatient subjects at the Parkinson's Unit. The inclusion criteria defined for the study are provided in the Supplementary Material (Appendix A.1.3);

[1]For future investigations. The text is provided in the Supplementary Material (Appendix A.3.3)

the clinical diagnosis of Parkinson's Disease was a necessary condition for study inclusion. As mentioned in Appendix **??**, this dataset is available on request, and was not made publicly available for privacy reasons.

The instrumentation selected for the analysis consisted in an integrated, wireless device capable of recording motion along nine degrees of freedom. Precisely, a triaxial inertial measurement unit (IMU). The device (Shimmer3) is produced by Shimmer © (Ireland, https://shimmersensing.com/); it is commercially available for research purposes. Device calibration and data management and extraction are allowed through the complementary software ConsensysPRO. The employed IMU unit includes an accelerometer, a gyroscope, and a magnetometer, working along the three axes in space (x, y, z) to capture motion. These three sensors are commonly employed in a wide range of human motion-capture frameworks, including for human activity recognition and medical purposes, and are breifly described below.

An accelerometer quantifies proper acceleration in space, i.e., the acceleration any object acquires when subjected to freefall. Multi-axis accelerometers generally measure this physical quantity as a vector, by recording both its magnitude and direction, and express it in terms of m/s^2.

A gyroscope is a device that measures the angular velocity of an item in space, commonly expressed in dps. The sensor integrated in the Shimmer3 is a MEMS-based gyroscope, measuring through a vibrating mechanical element the rate of rotation around each axis.

Finally, the integrated MEMS-magnetometer records the variations, in terms of magnitude and orientation, of the magnetic field along the three considered axis. In the employed device, this physical quantity is expressed in µT.

The IMU unit was selected for its high-resolution data acquisition capability, as well as the low-power consumption, light weight (60 g per unit), and small form factor (Figure 7.1). Regarding the device positioning, various tests were conducted to assess the best arrangement, in terms of data quality and intrusiveness. Finally, the IMU unit was placed on the chest, at the height of the sternum, to favour proper characterisation of whole-body movements, and minimise effect of motion artefacts during the night, as previously seen in literature [161]. The selected sampling frequency was 128 Hz, and the maximal recording range was selected for all sensors to minimise information loss. Figure 7.2 shows the final experimental set-up, and the orientation of each axis.

Sleep Questionnaires

In clinical settings, one of the scales to assess overall sleep quality is provided by the Pittsburgh Sleep Quality Index (PSQI) questionnaire, validated in various cohorts [25]. It consists in a self-report, 19-item survey, that investigates subjective sleep quality, sleep latency patterns, duration and efficiency, occurrence of sleep

Figure 7.1: Dimensions of the IMU adopted in the study (Shimmer3, adapted from: https://shimmersensing.com/)

disturbances and daytime dysfunction. An additional part envisages five items to be completed by roommates or bed partners to investigate habitual sleep patterns, which, however, do not contribute to the overall score.

Since the full version of the questionnaire was at times regarded as misleading, and time-consuming, a shortened version (sPSQI) was later proposed and clinically validated. It consists in a 13-item questionnaire, adapted from the original version. The complete list of administered questions is provided in the Supplementary Material (Appendix A.3). The scoring of the sPSQI yields an integer, global score (on a scale of 0 to 15) that is employed to discriminate between adequate and poor sleep quality. In a previous co-authored investigation on an adult, healthy cohort, values above 5 appeared indicative of bad sleep quality [3], consistent with the validated values [55]. The sPSQI was administered to all participants before bedtime, to assess overall sleep quality in the month prior to the examination. Despite knowing that the obtained score does not tangibly represent the actual night of sleep, the choice of administering the questionnaire before bedtime was made so as to gather a general representation of the quality of sleep of the subjects, without introducing any bias due to the presence of the sensors, or the participation in the experiment.

Second, a sleep survey on circadian habits was employed, and filled out after the sleep test. The survey, which will be referred to as SLEEPS, was first proposed

Figure 7.2: Sensor placement adopted in the study, along with the orientation of each axis. In detail: x: medio-lateral, y: axial (longitudinal), z: antero-posterior (vertical)

Table 7.1: Demographic characteristics of the population in the study. HC: healthy controls, PD: subjects with Parkinson's Disease.

	Sample	**Age**	**sPSQI**	**SLEEPS**
HC	28 (10 females)	38 ± 10.7 years	6.55 ± 1.27	2.55 ± 1.02
PD	12 (5 females)	68 ± 4.1 years	9.42 ± 4.06	3.68 ± 2.21

and validated in a previous works on remote assessment of sleep quality [3]. It was designed to explore the correlation between circadian habits and sleep quality, by examining general health, work (or study) routine, leisure time habits, and sleep schedule. The complete set of designed questions is included in the Supplementary Material (Appendix A.3.2). For the purpose of this study, and according to the results of the previous investigation, only a subset of the most relevant questions was administered, and is included in the Supplementary Material.

Subjects and Data

An extensive part of this research activity was devoted to the recruitment of eligible subjects. The study involved a cohort of subjects with Parkinson's Disease, who will be referred to with the acronym PD, and a second cohort of healthy controls. As mentioned previously, the former were recruited at the Parkinson's Unit (Dept. Neurology, AOU Città Della Salute e della Scienza, Turin), and at a non-profit patients' association in Turin (*Associazione Amici Parkinsoniani ONLUS*). As regards the healthy group, subjects were rectruited on a voluntary basis, among the patients' spouses, or age-matched family members, and University employees without familiarity for parkinsonisms and sleep disorders.

The observational study, and data collection, is presently ongoing; however, for the research activity presented in this Chapter, the dataset included 40 subjects. Demographics for this cohort are displayed in Table 7.1.

7.3.2 Methods

Pre-Processing of Sleep Recordings

To facilitate the feature extraction procedure, minimal pre-processing was carried out on the triaxial IMU data. For the sake of maintaining low-computational load, only the data from the accelerometer and gyroscope were analysed, as they provided sufficient information to characterise motility in bed.

Data recorded through the accelerometer were filtered through a moving average filter, with sliding window of 1s. This procedure mitigated the effect of high-frequency noise, which in raw data is observable during abrupt changes in position, such as fast rolling in bed. Gyroscope data, on the other hand, were processed

through a FIR lowpass filter, with cut-off frequency of 35 Hz, and order 21, to allow for accurate detection of rolling-over.

Figure 7.3 depicts a one-night accelerometer recording of a healthy subject, and Figure 7.4 of a patient with nocturnal hypokinesia.

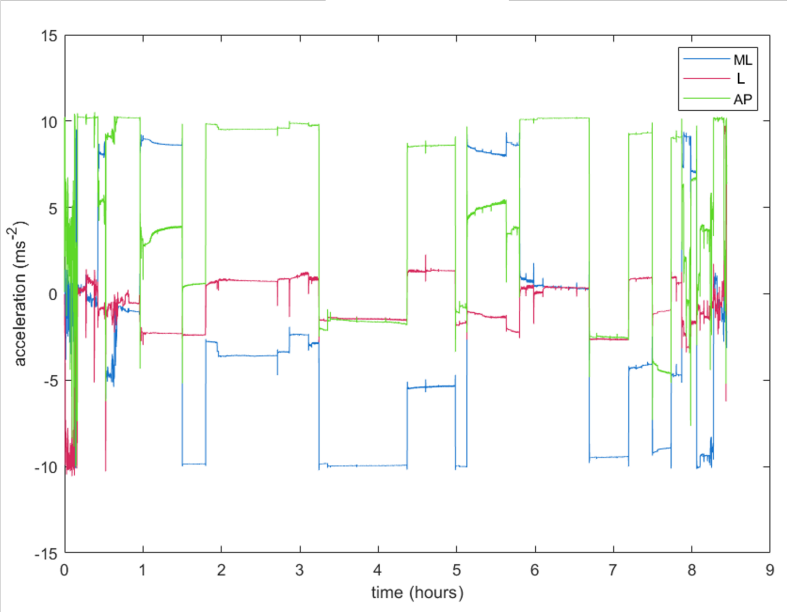


Figure 7.3: Accelerometry recording of a healthy subject. ML: Medio-lateral acceleration, L: longitudinal acceleration (body axis), AP: anterio-posterior acceleration.

Feature Extraction

As per the research objective **(1)**, a set of objective descriptors of overnight motility were extracted from inertial data. A blend of clinical (PSG-related) parameters and motility features were employed in this research activity. At present, there is no clinically validated set of parameters to characterise sleep patterns and nocturnal axial motility in Parkinson's Disease. Consequently, the proposed features encompassed both novel measures and features previously proposed in studies involving night-accelerometry or actigraphy [107, 161].

Motion-related features encompassed aspects such as reclining angle (θ), the duration of each sleeping position, and the frequency and speed of turning in bed. The definition of these quantities is detailed in the following paragraphs. The choice of employing such parameters was discussed with an expert neurologist, based on previous evidence that this set of nocturnal movements is descriptive of disease severity [121]. Indeed, it was observed that more advanced phenotypes presented with slower turning rates and more upright position, compared to controls and

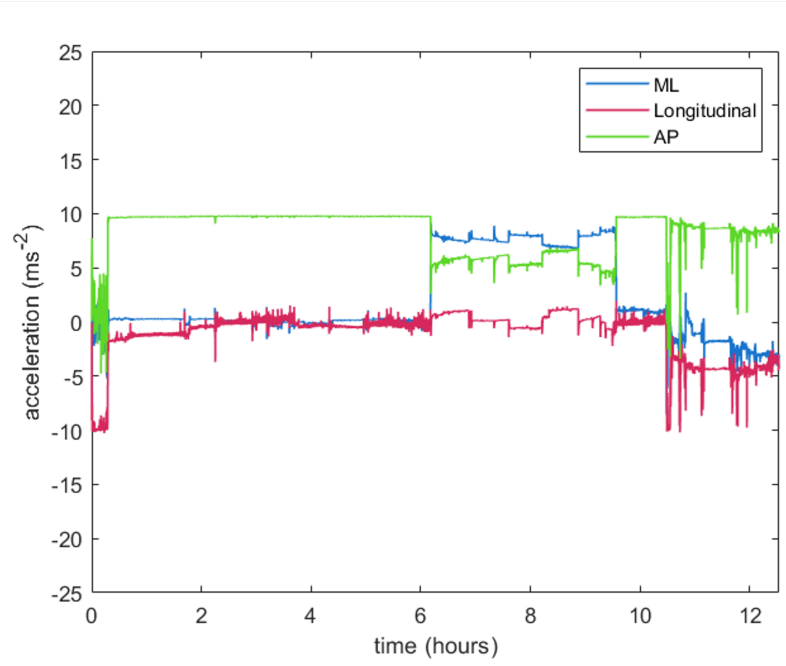


Figure 7.4: Accelerometry recording of a Parkinson's Disease subject with nocturnal hypokinesia, respectively. ML: Medio-lateral acceleration, Longitudinal acceleration (body axis), AP: anterio-posterior acceleration.

early-stage subjects. The frequency of turns, and their total number over a night of sleep has also exhibited a significant correlation with the MDS-UPDRS score [167].

The data recorded through the accelerometer were employed to retrieve information about sleeping position (Figure 7.5). In more detail, the recordings were inspected in 30s epochs, to match the international standards, and the position was estimated by virtue of a heuristically selected threshold. Precisely, values such as the reclining angle (θ) and the average acceleration for each dimension were simultaneously analysed epoch-wise to retrieve one of the five positions: *supine*, *prone*, *left-side*, *right-side*, and *standing*.

Subsequently, the occurrence of turning events was obtained by matching the sleeping position information with gyroscope data. Any observed change in sleeping position was initially marked as a potential turning event. To ensure a precise detection of axial rolling-over, each candidate event was marked as *real* only if, over a window of observation of 2 min, no change in sleeping position was observed before and after the event. From this train of data (Figure 7.5), parameters such as the number of turning events (N_{turns}) and the time intervals between events (R_{int}) were then extracted.

The features related to the velocity of turns were extracted through the analysis of the angular velocity along the longitudinal axis (y, as measured by the

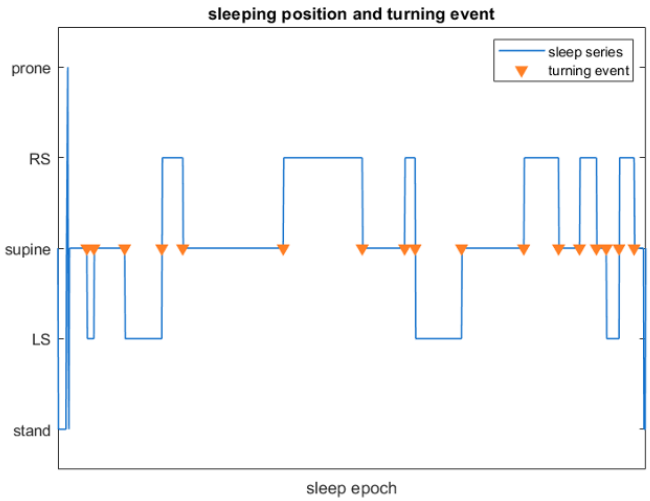


Figure 7.5: Changes in sleeping position during the night. Orange arrows represent the detected turning events.

gyroscope). In a similar fashion as previously, a preliminary peak detection step was carried out. Specifically, by inspecting the angular velocity in 30s epochs, all peaks exceeding 85% of the standard deviation of the signal in the selected epoch were marked as potential turns (Figure 7.6).

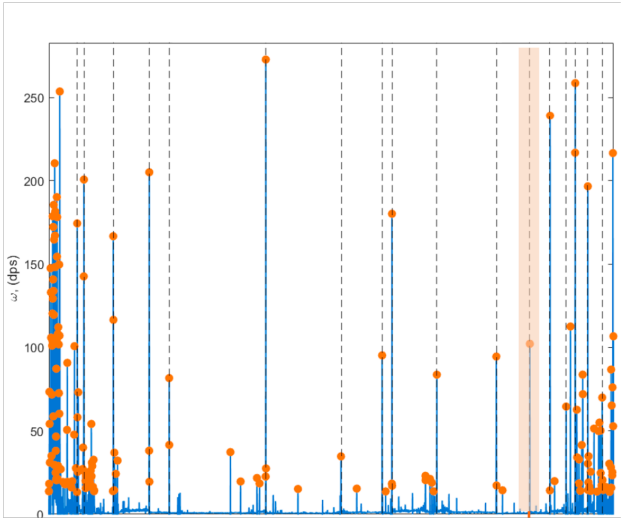


Figure 7.6: Peak detection from the longitudinal angular velocity recorded by the gyroscope. Actual turning events are marked by a vertical dashed line. The highlighted sector (orange) is a peak-search range, described in Figure 7.7.

Then, for each candidate turning event, a 50-second search range was employed,

to detect the actual peak in angular velocity that corresponded to a real, observed rolling-over (Figure 7.7). From this information, a second set of data was extracted, including parameters such as peak height (representing the extent of turning velocity, ω-turns) and peak width (indicating turning duration, T_{turns}).

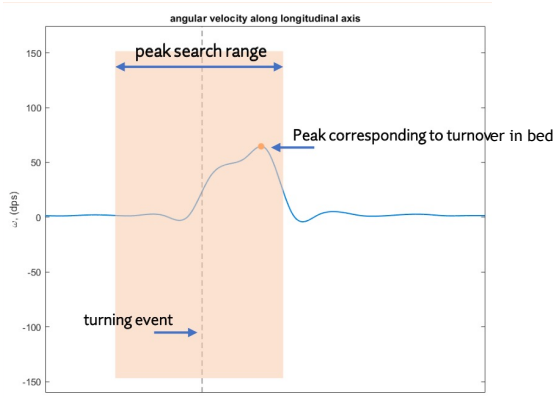


Figure 7.7: Turnover event detection from a gyroscope recording of the longitudinal axis.

Finally, a comprehensive descriptor of overnight motility was employed, namely, the Activity Index (AI) [9]. This feature was proposed in the literature as an open-source, straightforward quantifier of motion during the night, as most off-the-shelf devices do not publicly provide interpretation for their metrics. This quantity is deemed significant in the analysis of sleep patterns in Parkinson's Disease, as, on the one hand, it graphically depicts the motility trends during the night, and, on the other hand, it provides a numerical summarisation of the extent of motion. It is defined in the range $[0, 1]$, where values approaching 1 indicate higher motility.

For the purpose of this study, the AI was computed in 30 s epochs, from tri-axial accelerometry, through the following equation [9]:

$$\begin{aligned} AI &= \sqrt{\frac{1}{3}[(\sigma_x^2 - \sigma_{sys}) + (\sigma_y^2 - \sigma_{sys}) + (\sigma_z^2 - \sigma_{sys})]}, \\ \sigma_{sys} &= \sigma_x + \sigma_y + \sigma_z \end{aligned} \tag{7.1}$$

The quantity σ_{sys} represents the systematic noise of the device, a feature that characterises wide-range accelerometry, which, in this configuration, equalled 27.5 ms^{-2}. It was assessed as the variance of the three axes, measured by putting the IMU unit in a still, horizontal position.

Additionally, on the scent of expressing overnight motility as a continuous time series, the the AI *counts* – i.e., the values for each 30 s epoch – were exploited to define a novel metric, termed the Average Motility (AM). In more detail, the AM

was computed as the moving average of AI counts over a 2-minute window, a time span deemed suitable for nocturnal movements. Subsequently, the obtained time-series subsequently was scaled to a continuous value within the range $[0, 1]$. Values approaching 1 denote increased nocturnal activity, as depicted in Figure 7.8.

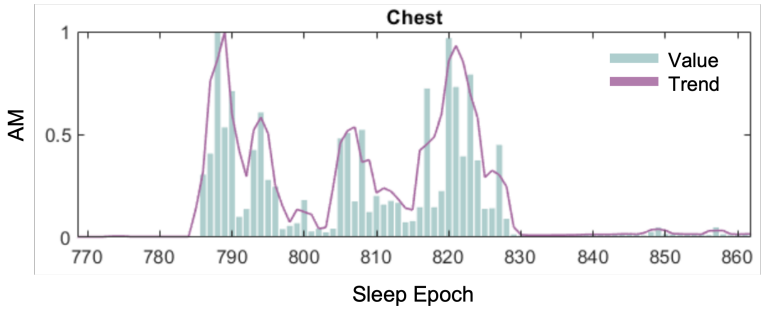


Figure 7.8: Values and trend of the Average Motility metric, for a 45 minute portion of sleep in a healthy subject, right before wake-up time.

The whole set of extracted features is shown in Table 7.2, arranged ccording to their category – i.e., *Clinical* or *Motility*. Since all features in the Motility category were computed on epochs of 30-second length, a range of summary statistics (mean, standard deviation, maximum and minimum value, 25^{th} and 75^{th} percentiles, kurtosis, and skewness) were calculated from the array, and utilised as distinct features.

Feature Analysis and Feature Selection

As described, the feature extraction step envisaged the characterisation of night motility through IMU data, through a set of parameters which included both novel metrics, and clinically-useful parameters outlined after insight from specialised neurologists.

Given the diverse nature of the extracted parameters, a statistical exploration was conducted, to investigate their importance for the application at hand. All statistical investigations were carried out through the open-source *jamovi* tool ([86], `https://jamovi.org/`). First, the distribution normality of the samples was tested through the Shapiro-Wilk test. Then, independent-samples tests were carried out, through the Student's t-test, and Mann-Whitney U test, for a parametric and non-parametric approach, respectively. In particular, the former was applied to normally distributed data, whereas the latter was the elected test for non-normally distributed data.

Notably, given the research objectives **(2)** and **(3)** – i.e., assess sleep quality (SQ), and identify parameters to describe sleep in PD, respectively – different configurations were adopted in testing.

The following two configurations were initially assessed:

Table 7.2: Features employed in the study, according to their category. Proper reference is shown, as in ⋄: adapted from cited study; ⋆: first proposed in this study.

Feature	**Description**	**Reference**
Clinical		
Sleep Onset Latency (SOL)	The amount of time required to fall asleep (min)	various
Wake After Sleep Onset (WASO)	The amount of time the subject is awake during the night (min)	various
Total Sleep Time (TST)	Total hours of sleep	various
Time in bed (TIB)	Lights-off to lights-on interval (h)	various
Sleep Efficiency (SE)	The percentage of time spent asleep while in bed (%)	various
SLEEPS score	Perceived sleep health and quality	[3]
Motility		
Tilt Angle (θ)	Reclining angle in bed	⋆
Sleeping position	Minutes spent in each sleeping position (*supine*, *prone*, *left-side*, *right-side*)	⋆
Number of turns (N_{turns})	Number of turns in bed	⋄[104]
Rotation interval (R_{int})	Interval between turning events (min)	⋄[104]
Rotation velocity (ω-turns)	Velocity of turning in bed (deg/s)	⋄[160]
Rotation acceleration (α-turns)	Acceleration of turning in bed (deg/s^2)	⋄[160]
Turning duration (T_{turns})	Duration of each turn (s)	⋄[160]
Stand/Sit Duration (SSD)	Total time spent standing or sitting during the night (min)	⋆
Activity Index (AI)	Level of activity during the night. Range: [0, 1]	[104]
Average Motility (AM)	Overnight motility trend.	⋄[9]

i) Good SQ *vs* Poor SQ,

ii) HC subjects *vs* PD subjects.

Given the wide distribution of sleep quality in the explored dataset, further

statistical investigations envisaged the following:

iii) HC with good SQ *vs* HC with poor SQ (HC_{good} versus HC_{bad}),

iv) HC with poor SQ versus PD with poor SQ (HC_{bad} versus PD_{bad})

Due to the limited number of subjects in the PD group, and the fact that sPSQI values were critically skewed towards bad sleep quality, the comparison PD with good SQ *vs* PD with bad SQ was overlooked, but will be addressed in future developments.

In addition to these comparisons, the correlation of the extracted parameters with the two research objectives was assessed. Precisely, Spearman's correlation was employed to evaluate **(a)** the correlation between the sleep features and the sPSQI score, and **(b)** the correlation with the presence of Parkinson's Disease.

Prior to the classification task, which envisaged configurations (i) and (ii), and to facilitate the interpretability of the models, and enhance data quality, feature selection was carried out. The selected approach was the ReliefF algorithm [168]; the top-k features in each configuration were deemed relevant for the subsequent classification task. For the sake of clarity, the k parameter was chosen independently for each configuration, by heuristically identifying the elbow on the retrieved feature importance scores.

Finally, z-score normalisation (Equation 6.1) was implemented to mitigate the effect of outliers and scale divergence on the subsequent analytic steps.

Automatic Classification through Supervised Machine Learning

As briefly mentioned in the previous Section, the classification task regarded only configurations (i) and (ii), primarily for numerousness reasons. This distinction identified a reasonably homogeneous group for (i) – 18 good sleepers and 22 bad sleepers – although group (ii) remained quite umbalanced, with 12 PD and 28 HC.

Due to the feature-based nature of the analysis, a binary classification task for each configuration was tackled by means of three distinct supervised models. Namely, a SVM, a KNN, and eXtreme Gradient Boosting (XGBoost). This latter is an ensemble method based on decision trees, that, at each iteration, incorporates gradient boosting to enhance classification performance, by training each learner on the residuals of the previous models.

Due to the rather scarce sample size, in order to promote the robustness of the models, and limit the risk of overfitting, hyperparameters were optimised following a Grid Search approach, on a train of 50 iterations. The set of tuned paratemers is displayed in Table 7.3.

The models were validated through a LOSO-CV approach, to allow for better generalisation of the results, as already discussed in previous Chapters of this Thesis. Finally, the performance of the explored models was assessed in terms of overall

Table 7.3: Summary of the employed classifiers and the searched hyperparameters, (parameter and range).

Model	Searched Hyperparameters
SVM	*Kernel function*: linear, polynomial, radial basis, sigmoid *Penalty (C)*: [0.1, 1, 10, 100, 1000] γ: [1, 0.1, 0.001, 0.0001]
KNN	*Minkowski Distance order (p):* [1, 2, 3, 4, 5] *Number of neighbours (K)*: [3, 5, 7] *Weights (W)*: uniform, distance-based
XGBoost	*Number of trees*: [25, 50, 100] *Depth*: [3, 5, 7] *Learning rate*: [0.001, 0.01, 0.1]

Accuracy, Recall, and F1 score.

7.3.3 Results

Statistical Analysis

Table 7.4 illustrate the outcome of the statistical independent sample tests.

For configuration (*i*) – i.e., Good vs Bad SQ – statistical significance emerged for variables describing the duration of turns ($T_{turns,std}$, $T_{turns,skew}$), and overnight motility (AI$p25$, AM$mean$). These metrics also exhibited a moderate negative correlation with the sPSQI score, as indicated by Spearman's ρ of -0.46 and -0.44, respectively.

For configuration (*ii*), – i.e., HC vs PD – the features characterising overnight body position (θ_{mean}, θ_{p75}), whole-body movements (ω-turns$_{p25}$, ω-turns$_{skew}$), and overall motility (AI$skew$, AM$_{mean}$) displayed statistical significance. Furthermore, when evaluating correlation with the presence of the disease, the listed features exhibited moderate values of ρ. In particular, θ_{mean} exhibiting a correlation coefficient of 0.54. Anew, negative correlations values were observed in features describing the velocity of turns in bed.

The final part of the statistical testing aimed at performing a preliminary stratified analysis, in an attempt to provide an exhaustive exploration of the collected data. Hence, configurations (*iii*) and (*iv*) were investigated; Table 7.5 illustrates the obtained statistics follow.

As appreciable, in the HC*good* vs HC*bad* configuration (*iii*), the features representing turning velocity emerged as significant, and exhibited Spearman's ρ values suggestive of moderately high correlation.

Different patterns emerged for HC*bad* vs PD*bad* (*iv*), where parameters related to body position, velocity of turns, and overall motility were significant, with p<0.001. These quantities also presented with moderate-to-high correlation with the sPSQI, with ρ of 0.64 for θ_{mean} (body position), and moderate, though negative correlation for the velocity of turns (ω-turns), in good agreement with previous observations.

Based on the results presented in this paragraph and summarised in Tables 7.4 and 7.5, it is possible to imply that the proposed set of metrics are appropriate for the definition of motor-related sleep patterns in the explored population. Notably, whole-body motility appeared to serve as a robust descriptor of such, both for sleep quality assessment, and the characterisation of PD.

Table 7.4: Independent Sample statistics of the features employed in the classification tasks (configurations (*i*) and (*ii*)), along with their correlation with the target (sPSQI or PD). Significance level is marked as **: $p<0.005$, ***: $p<0.001$.

Sleep Quality		
Feature	**Independent Sample Test**	**Correlation** (ρ)
ω-turns$_{p75}$	<0.05	0.32
T$_{turns,td}$	<0.005**	0.49
T$_{turns,skew}$	<0.001***	0.55
AI$_{p25}$	<0.05	-0.46
AM$_{mean}$	<0.05	-0.44
HC vs PD		
Feature	**Independent Sample Test**	**Correlation** (ρ)
θ_{mean}	<0.005**	0.54
θ_{p75}	<0.001***	0.45
R$_{int,p25}$	<0.05	-0.31
ω-turns$_{p25}$	<0.001***	-0.41
ω-turns$_{skew}$	<0.005**	-0.39
AI$_{skew}$	<0.001***	0.31
AM$_{mean}$	<0.001***	0.32

Machine Learning: Feature Selection and Binary Classification

This Section illustrates the outcome of the ML pipeline displayed in Figure 7.9.

As discussed, a feature selection step was implemented to highlight two subsets, including the most relevant features for the classification. The selected features for each task were summarised in Table 7.4, along with their statistical relevance.

Table 7.5: Statistical exploration of configurations (*iii*) and (*iv*)), along with their correlation with the sPSQI. Significance level is marked as *: $p<0.05$, **: $p<0.005$, ***: $p<0.001$.

HC_{good} *vs* HC_{bad}		
Feature	**Independent Sample Test**	**Correlation** (ρ)
ω-turns$_{p25}$	$<0.05^{*}$	-0.64
$T_{turns,std}$	$<0.005^{**}$	0.45
HC_{bad} *vs* PD_{bad}		
Feature	**Independent Sample Test**	**Correlation** (ρ)
θ_{mean}	$<0.001^{***}$	0.64
ω-turns$_{p75}$	$<0.001^{***}$	-0.41
ω-turns$_{mean}$	$<0.001^{***}$	-0.35
AI_{mean}	$<0.001^{***}$	-0.47
AM_{mean}	$<0.001^{***}$	-0.51

Evidently, 5 features were selected for the first task, and 7 for the second. The two subsets were independently given as input to the ML classifiers, to test their predictive power.

The results attained in the the Good vs Bad SQ classification task are presented in Table 7.6. The employed models presented with an overall accuracy of 79.67% ± 4.43, indicative of reasonable classification accuracy. The optimal scores are achieved through an optimised XGBoost classifier, with the following parameters: N_{trees}=50, depth=5, learning rate=0.1. This model achieved values of accuracy, Recall, and F1 score of 85.7%, 78.6%, and 82.5%. These results appeared to confirm the discriminative capability of the employed features for this task.

Likewise, Table 7.7 illustrates the performance metrics for the explored classifiers, in the HC vs PD task.

A slight increase in classification performance was observed, as all models presented with moderately high accuracy, with average value of 89.1% ± 6.39. Similarly, the average F1 score across all models reached 81.3% ± 9.57. An optimised SVM emerged as the best model, with linear kernel and penalty (C) equalling 1. The classifier achieved an an overall accuracy exceeding 96%. Similarly, it reached 95% and 93.4% accuracy F1 score, respectively.

Overall, both explored configurations demonstrated the feasibility of remotely detecting Parkinson's Disease-related sleep patterns and evaluating sleep quality through a simple and lightweight framework.

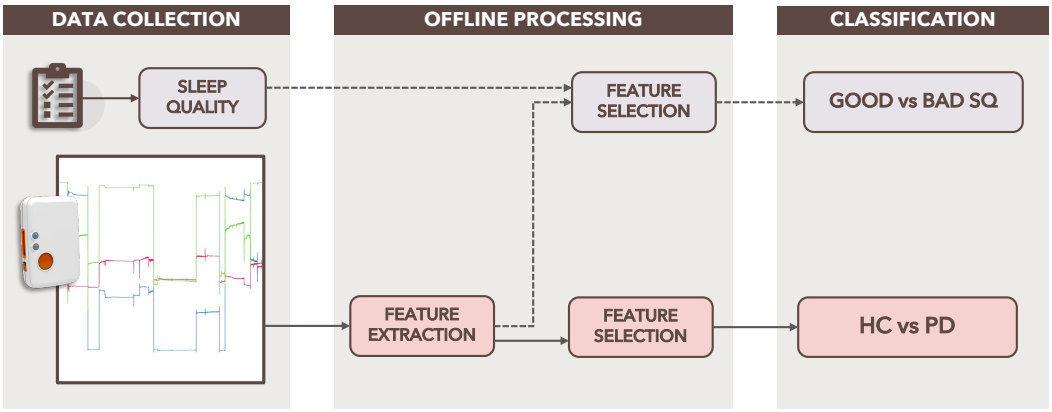


Figure 7.9: Summary of the supervised Machine Learning pipeline adopted in this study.

Table 7.6: Results of the classification task in the configuration (i): Good vs Bad Sleep Quality, in a LOSO-CV approach.

	SVM	**KNN**	**XGBoost**
Accuracy	78.1 %	75.2 %	85.7 %
Recall	74.0 %	73.3 %	78.6 %
F1	72.0 %	70.8 %	82.5 %

7.3.4 Discussion

The research activity included in this Chapter aimed at proposing a possible set-up for monitoring sleep disorders in Parkinson's Disease, to facilitate remote monitoring and efficient disease management. In light of the encouraging results of a previous study on remote assessment of sleep quality [3], this research work aimed at expanding the paradigm and translating it to the scenario of neurodegenerative diseases.

An experimental protocol, together with the recruiting campaign, was defined in accordance with neurologists, who also identified the pool of suitable participants. Following an unsupervised fashion, data collection took place in home-settings, and included both healthy subjects and PD patients.

From a technical perspective, the study aimed at capturing nighttime motility through light weight IMU units. Specifically, the recorded inertial data were processed so as to obtain a set of accurate and robust descriptors of motility in bed. Then, supervised ML models were exploited to tackle automatic classification of sleep quality (good or poor), and to detect Parkinson's Disease based on overnight motility. Hence, the proposed set-up served to assess the feasibility of remote sleep-related motor patterns via wearable sensors.

Table 7.7: Results of the classification task in the configuration (*ii*): Healthy Controls vs PD Subjects, in a LOSO-CV approach.

	SVM	**KNN**	**XGBoost**
Accuracy	96.2 %	90.4 %	80.7 %
Recall	95.0 %	72.3 %	85.2 %
F1	93.4 %	80.6 %	70.0 %

In particular, the reclining angle in bed over the night (θ) presented with moderately high correlation with the presence of PD. Values of θ approaching 90°, and describing an upright position, emerged as distinctive for PD. This outcome might be related to the higher sleep fragmentation in PD subjects and the frequent sleep interruptions, as also observed in [121]. In addition, the reduced velocity of turning in bed (ω-turns) appeared among the most informative features for the binary classification HC vs PD, as well as the overall motility in bed (activity index, AI). These results appear to reflect the clinical manifestation of nocturnal hypokinesia, in which motility is significantly reduced, and often associated with rigidity. Furthermore, these findings are consistent with previous studies [121, 167], which stated that the paradigm of turning in bed likely resembles that of upright turning in subjects with PD.

The encouraging results attained through the explored supervised models proved the feasibility of the explored task.

Indeed, automatic classification of sleep quality, with the best classifier achieving an overall accuracy of 85.7% and an F1 score of 82.5%, was in line with results in the existing literature. However, it is important to note that direct comparison is impassable, as most of the published studies with similar set-ups are conducted on young, healthy cohorts.

Regarding PD detection through overnight motility parameters, the best model attained an F1 score of 93.4%. Along with high values of accuracy, this performance suggested the suitability of whole-body motion for the characterisation of sleep in PD. However, this research activity possibly appeared to be the first to address feature-driven automatic PD detection exploiting motion during sleep, so a fair comparison with similar setups was hitherto impracticable.

Although presenting with promising performance and encouraging applicability, this preliminary study bears some limitations, which shall be addressed in future developments, and are discussed in the conclusive remarks of this Chapter.

7.4 Conclusion

Sleep disorders appear among the most common non-motor symptoms of Parkinson's Disease. Although entailing diversified phenomena, they often manifest as disturbances in motor function during sleep. Particularly, nocturnal hypokinesia and akinesia affect up to 70% of subjects already at the earliest stages of the disease. This branch of sleep-related disorders negatively affects the quality of life, significantly grazing self-care and enforcing, in certain instances, a significant burden on caregivers.

Currently, the clinical approach to motor-related sleep disturbances in PD lacks an objective point of view. Indeed, it relies primarily on subjective reports, including rating scales, interviews, or self-reports, which frequently result in symptoms under-rating [131]. Therefore, nocturnal disturbances are often overlooked, and assessing overall sleep quality and their occurrence remains a challenging task.

Hence, effective monitoring strategies would have a crucial impact on diagnostic and therapeutic pathways, by introducing objective insight in the assessment of symptoms. In this perspective, the adoption of a continuous monitoring approach would facilitate evidence-based follow-up, and highlight fluctuations over time, thus bridging the gap between in-person assessments.

Although actigraphy offers a minimally intrusive solution to monitor overnight activity over large periods of time, it often suffers from poor placement. Conversely, axial placements are able to capture rotational movements and properly describe whole-body motility.

This Chapter presented an objective approach in the characterisation of motion during of sleep, with the primary aims of **(1)** assessing overall sleep quality and **(2)** describing sleep patterns in PD, through a feature-driven approach.

The experimental protocol entailed the recruitment of subjects, and data were collected through lightweight instrumentation in an unsupervised fashion. Indeed, this activity also addressed the feasibility of home-based monitoring of sleep-related motor manifestations.

Precisely, overnight motility data were collected through IMU units placed in an axial configuration, and kinematic parameters, encompassing positional statistics, number and extent of rolling-over, were retrieved and employed in a supervised ML framework.

The results obtained appeared encouraging, albeit the investigation is currently open. The extracted parameters moderately correlated with the two study aims, suggesting their good predictive performance in the explored tasks. The highest values of correlation were observed in parameters regarding the duration of turning in bed, and average body position during the night. Accordingly, features describing the velocity of rolling-over expressed moderate correlation with the PD class, suggesting their ability to describe nocturnal hypokinesia.

These findings found confirmation through the outcome of the automatic classification. Sleep quality – in a discrete, good *vs* poor, fashion – was detected with encouraging performance, with the best model achieving a 85.7% accuracy, and 82.5% recall, in a mixed cohort including both healthy and PD subjects.

Finally, to assess the feasibility of detecting PD sleep patterns through overnight kinetic parameters, a binary task between healthy and PD subjects was addressed. The employed features proved efficient, with the explored models achieving an overall accuracy of almost 90%, and the best model exceeding 96%. Furthermore, the average F1 score was of 89% $\pm$ 5.98, suggestive of good detection capability. This metric becomes particularly important in this configuration, due to the imbalance in the class distribution.

Despite the limited size of the dataset, the proposed framework and the results obtained can be considered robust, as the feature set underwent proper feature selection, and the models' hyperparameters were optimised. Additionally, the performance was evaluated in all configuration through a Leave-One-Subject-Out cross-validation approach, to limit the influence of classification bias.

The findings described in this Chapter suggested the feasibility of remotely characterising sleep in Parkinson's Disease, and possibly implement the proposed parameters in follow-up procedures, to facilitate continuous monitoring strategies and allow for personalised therapeutic interventions.

However, as previously mentioned, the presented results are exploratory and future studies should further explore this scenario.

First, an extensive subject recruitment shall be addressed. Indeed, while the participation from healthy subjects was rather undemanding, the PD cohort presented with lower patient adherence. A co-factor, in this regard, were the clinical protocols related to the COVID-19 pandemic, which significantly hindered contact with the patients and hampered data collection. Future recruiting campaigns should inevitably target an age-matched control populations. Indeed, the current dataset envisaged a high age variability within the control group (range: 23–66 years). Introducing healthy, though older in age, subjects could provide better insight into sleep patterns in the elderly. Indeed previous studies highlighted how turning velocity and motility patterns in early-PD resembled those of controls [121]. A larger dataset, and a higher number of subjects with Parkinson's Disease could facilitate more comprehensive explorations, including longitudinal studies, or stratification according to the disease level, age, and other co-morbidities, to enhance the model generalisation and robustness.

Second, the employed devices are not commercially-available, and currently purchasable only for research purposes. Patient adeherence may benefit from the use of off-the-shelf devices, possibly user-friendly. Furthermore, pairing sleep-recording devices with smartphone apps, possibly with a playful, reward-oriented approach, could facilitate the integration of wearable setups in a real-world, pervasive scenario.

In addition, the proposed framework is applicable only for axial positioning of

the sensors. However, when testing other configurations, or sensor orientation, the predictive power of the kinematic parameters may change. Future developments will envisage the simultaneous recording of EMG, for an accurate characterisation of movement and muscle activations during sleep, and preliminary investigations are underway.

Sleep quality, withal, was assessed through a dichotomous approach. Indeed, subjects were labelled as having either *good* or *poor* sleep quality, by posing a threshold on the sPSQI value. Although the threshold was in line with the one proposed in the literature [55], future works should investigate regression approaches, therefore retrieving a continuous score, virtually corresponding to the sPSQI value, based on kinematic characterisation.

Finally, as mentioned in the introductory part of the study, vocal recordings were collected through smartphones, to provide for an integrated, multi-modal approach in the evaluation of motor and non-motor symptoms in Parkinson's Disease, and support the development of continuous monitoring strategies, thus possibly positively impacting the quality of life of individuals and caregivers.

Chapter 8

REM Sleep Parasomnias in Amyotrophic Lateral Sclerosis

Sleep disorders are present in over 70% of people living with a diagnosis of Amyotrophic Lateral Sclerosis. Among the sleep disturbances reported in ALS, REM-related parasomnias were observed, likely revaling the underlying chronic degeneration of neurons [16]. In a few cohorts, REM Sleep Without atonia was reported, and previous studies suggested its correlation with disease severity.

This Chapter illustrates the preliminary findings of a longitudinal retrospective study on the correlation between EMG-derived metrics and disease severity, conducted at the Centre for Sleep Medicine at Molinette University Hospital (Turin, Italy).

The scientific content of this Chapter was included in the paper *"Predicting Amyotrophic Lateral Sclerosis Progression: an EMG-based Survival Analysis"*, accepted for presentation at the 46$^{\text{th}}$ Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2024).

8.1 Context and Background

Amyotrophic Lateral Sclerosis (ALS) is a chronic, progressive neurodegenerative disease, affecting motor neurons that control voluntary movement, whose deterioration ultimately leads to muscle weakness and paralysis [112]. At present, the median survival time hovers from 2 to 5 years from symptoms' onset; no curative treatment is available, and treatment mainly targets symptomatic management and palliative care.

The disease presents with an heterogeneous clinical picture, including dysarthria, dysphagia, and breathing-related disorders, often culminating in respiratory failure [76]. This family of motor dysfunctions are accompanied in the vast majority of cases by non-motor manifestations, including sleep disorders [16]. Among these,

sleep-disordered breathing and sleep fragmentation are the most common, but various studies investigated the presence of RSWA, possibly with its association with RBD [106].

Although current literature on RSWA in ALS remains scarce, in previous cohort studies a potential correlation between the extent of RSWA and ALS disease severity appeared [133, 103, 180], based on the qualitative and quantitative analysis of EMG activity recorded during sleep through in-hospital PSG. In more detail, in [133] a group of 29 patients was enrolled. Then, the RAI, along with a series of clinical polysomnographic metrics, and the number of chin EMG activations during REM sleep were retrieved from full vPSG, and significant correlations were observed with the degree of functional impairment, according to the ALS-Functional Rating Scale-Revised (ALS-FRS). In [180], abnormal, elevated motor activity during REM sleep was observed, with and without association with RBD. Although the involvement of RBD in the development of ALS is still unclear, based on these findings, the investigation of EMG during sleep becomes significant, to explore the risk for development of RBD in these cohorts, and to monitor disease severity.

However, as extensively discussed in this Thesis, traditional polysomnography protocols are often considered intrusive, primarily due to the cumbersome instrumentation required. With special regard to ALS, PSG becomes a truly invasive procedure, due to its set-up and the necessity of outpatient clinic settings. Furthermore, many subjects with ALS carry a non-invasive ventilation (NIV) mask, which may add to the overall instrumentation discomfort.

This scenario implies a clinical and technological obstacle for non-invasive follow-up protocols in ALS, and highlights the need for innovative monitoring strategies, to stabilise the quality of life in the best way possible. In the perspective of continuous, remote monitoring strategies, the research activity described in this Chapter aims at **(1)** exploring lightweight EMG metrics in sleep, and **(2)** assessing their utility in personalised follow-up protocols for high-risk ALS patients.

8.2 Research Overview

As discussed in the previous Section, effective monitoring strategies are needed in ALS clinical pathways, to allow for improved patient comfort. The investigation of sleep serves as reservoir of valuable clinical insights about the progression of the disease, and may possibly help improving disease management.

The study included in this Chapter aimed at investigating the potentiality expressed by non-invasive EMG parameters collected during REM sleep in predicting survival outcomes. Previously, survival patterns in ALS cohort were explored, by stratification according to age, ALS-FRS, apnea parameters [95, 136], or needle EMG [158].

The research activity, detailed in the following Sections, was conducted in cooperation with the Centre for Sleep Medicine at Molinette University Hospital (Turin, Italy), and consisted in:

1. Characterising RSWA in the explored ALS cohort through established metrics,
2. Explore the relationship between EMG-derived metrics and disease progression, through survival model analysis,
3. Assess the predictive power of lightweight, unobtrusive EMG parameters as regards survival time.

8.3 Monitoring Disease Progression in ALS

8.3.1 Materials

Study Protocol, Subjects, and Data

This longitudinal retrospective study envisaged in-hospital PSG follow-up assessments at four time points, each within a 6-month distance from the prior, for a total of 18 months of observation. The experimental sessions and data collection were conducted at the Regional Centre for Sleep Medicine in Turin.

An initial pool of 58 subjects were selected for participating in the study; inclusion and exclusion criteria were discussed by physicians and domain experts and are provided in the Supplementary Material (Appendix A.1.4). As mentioned in Appendix A.1.4, this dataset is available on request, and was not made publicly available for privacy reasons.The subjects were undergoing melatonin treatment and were evaluated for RSWA symptoms at the time of enrollment (T0). Subsequent PSG assessments were conducted after 6 months (T1), 12 months (T2), and 18 months (T3), though only one subject reached this end-point. After the overall dataset inspection, 13 subjects were excluded from the analysis included in this work due to technical issues.

The final dataset (REMALS Database) comprised 45 ALS subjects (30 males); only 9 subjects completed the T0–T2 assessments. Drop-out prior to the end of observation time was primarily due to decease, or the necessity of invasive ventilation, which implied inability to follow-up. This category of events was therefore identified as the primary endpoint of the study, and will be referred to as *left-censoring* in the following paragraphs. At the time of actual *censorship* – i.e., the end of the observation window – a total of 36 ALS subjects reached the primary endpoint.

Finally, 35 age-matched controls (HC) without diagnosis of or familiarity for neurodegenerative disorders and with no diagnosed RBD were included in the study;

they underwent PSG at the selected time points, and employed to establish the healthy reference model.

Table 8.1 provides a summary of the demographic characteristics of the total dataset. As appreciable, at the time of the enrollment, 25 subjects in the ALS cohort screened positive for RSWA (RSWA+).

Driven by the aim of developing an unobtrusive follow-up framework, only the EMG recorded from the mylohyoid muscle was employed. Indeed, this would enable a simplified recording approach compared to PSG, thus facilitating long-term follow-up in these patients' populations. All PSG recordings were manually scored for sleep stages according to the AASM criteria [11].

Table 8.1: Summary of the demographic characteristics of the dataset under study.

	Sample	**Age** (years)	**RAI**
General ALS	45 (15 females)	65.51 ± 9.34	0.62 ± 0.34
RSWA+	25 (9 females)	65.36 ± 9.77	0.40 ± 0.30
RSWA−	20 (6 females)	65.70 ± 8.77	0.90 ± 0.04
HC	35 (15 females)	56.95 ± 12.06	0.91 ± 0.02

8.3.2 Methods

Data Processing and Feature Extraction

The selected recordings underwent minimal processing to facilitate the extraction of EMG-derived features, in line with standard diagnostic criteria, and previously discussed in Chapter 6.

Specifically, the EMG signal was bandpass filtered in the range 10–100 Hz. First, the REM Atonia Index was computed [61] on 1-second epochs. To briefly revise the concept, the RAI accounts for the number of EMG epochs in REM sleep with amplitude below 1 µV, and values close to 0 represent loss of REM atonia.

Second, an improved version of the Dissociation Index (DI) was computed. Precisely, a more comprehensive feature set was employed to characterise the healthy and non-healthy models, in an attempt to minimise information loss from the available data. Table 8.2 displays the complete set of features. For consistency with the RAI computation, each metric was computed on 1-second epochs, and various statistics were derived (median, 25^{th} and 75^{th} percentiles, IQR, kurtosis, skewness, maximum and minumum value) to obtain the subject-specific feature array. Then, in accordance with the methodology presented in Chapter 6 and detailed in the Supplementary Material (Appendix A.2), the DI was computed. For the sake of clarity, values approaching 1 indicate complete dissimilarity to the healthy model, which, in this study, was built through the EMG data from the 36 healthy controls.

Table 8.2: Features employed in the improved computation of the Dissociation Index, along with their domain and reference.

Category	Feature (Name and description)	Reference
Time	Amplitude metrics: mean, standard deviation, skewness, kurtosis, range, maximum and minimum value	various
	Hjorth Parameters: activity, mobility	[122]
	Percentiles (25^{th}, 75^{th}, 95^{th})	various
Frequency	Power Spectral Density: numerical and statistical measures (mean and median frequencies, total power)	various
	Spectral Edge Frequencies (i.e., spectral percentiles): SEF25, SEF95	◇[139]
	Average power of the whole signal band	various

Survival Analysis for Disease Modelling

As per study objective **(2)**, multiple survival models were employed in an attempt to model the disease progression and explore the survival-prediction power of the EMG metrics.

The feature set employed for this task included the following five variables: Age, Sex, Follow-up Time (days), RAI_{T0}, and DI_{T0}, which served as co-variates to model the survival functions. For the sake of clarity, RAI_{T0}, and DI_{T0} represent the RAI and DI evaluated at the time of enrollment. As appreciable, demographic variables, along with clinical and quantitative EMG metrics were included.

The general probability of survival in the explored cohort was investigated through a univariate Kaplan-Meier approach [51], relying only on the variable Follow-Up time. This latter was expressed in days from the time of enrollment until drop-out. The estimated survival function, for a general time instant t, is expressed as:

$$\hat{S(t)} = \prod_{t_i \leq t} \left(1 - \frac{o_i}{r_i}\right) \tag{8.1}$$

where o_i is the number of occurrences (i.e., decease) at time t_i, and r_i the total number of individuals in the dataset who have not experienced the event yet, and therefore have not been censored.

Second, a multivariate survival analysis was conducted, to explore the risk-prediction capability of each co-variate in the extracted feature set. This was carried out by means of a Cox' Proportional Hazards model [52], generally employed to

estimate the impact of various risk factors on survival time. For a general time instant t, the expected hazard is expressed as:

$$h(t) = h(t_0)e^{\beta_1 X_1 + \beta_2 X_2 + ... + \beta_N X_N} \tag{8.2}$$

where $X_i = (X_1, ..., X_N)$ are the values of the N covariates, β_i are the effect parameters for each predictor, and $h(t_0)$ is the baseline hazard, obtained when all predictors X_i equal zero.

Feature importance was assessed by means of the Harrell's concordance index (C-index) [77], representative of concordance probability. In more detail, this metric depicts the number of concordant pairs with respect to the total number of comparable pairs. Given a pair of observations in the dataset i and j, it is defined as [151]:

$$C = \frac{\sum_{i,j} 1(\tilde{T}_i > \tilde{T}_j) \cdot 1(\eta_j > \eta_i) \cdot \Delta_j}{\sum_{i,j} 1(\tilde{T}_i > \tilde{T}_j) \cdot \Delta_j} \tag{8.3}$$

where T represents survival times, and η_i the predictions (i.e., a one-dimensional, average score defined for each observation, which relates to the overall cumulative hazard over time). Δ is an auxiliary variable that refers to the occurrence of an event, such that Δ=0 if the event has not been observed, and Δ=1 otherwise – e.g., this latter in the case of censoring. The expressions:

- $1(\tilde{T}_i > \tilde{T}_j)$ equals 1 if $\tilde{T}_i > \tilde{T}_j$, 0 otherwise,
- $1(\tilde{\eta}_i > \tilde{\eta}_j)$ equals 1 if $\tilde{\eta}_i > \tilde{\eta}_j$, 0 otherwise.

Ideally, values of C approaching 1 represent a perfect association.

Furthermore, following previous findings in the literature that suggested a potential correlation between the presence of RSWA and the clinical severity of the disease [133], stratified survival analyses were carried out. Hence, data stratification according to the absence or presence of RSWA was conducted. For this purpose, a cut-off value of RAI=0.8 was chosen to split the dataset into two groups: subjects expressing loss of REM sleep atonia (RSWA+), and subjects that exhibited no signs of altered REM atonia (RSWA−). The RAI threshold was chosen in line with the recommendations provided in [61]; a Kaplan-Meier estimator was then modelled on the two datasets.

Lastly, given the presence of two poignant demographic co-variates in the dataset (Age and Sex) a further stratified survival analysis was tackled. However, due to the quite homogeneous Age distribution in the ALS subjects, stratification according to Age was not feasible. Hence, a stratified analysis based on Sex was conducted, and a Kaplan-Meier estimator was fit on the two obtained strata. To evaluate the risk-prediction performance of the trained models a log-rank test was adopted, employing χ^2 and p-values as statistics.

Predictive Modeling of ALS Progression

In the final part of this work, study objective **(3)** was addressed, i.e., assessing the feasibility of modeling disease progression over time through unobtrusive EMG parameters. Hence, a multivariate-oriented, Cox' Proportional Hazards model was fit on the ALS dataset. The survival analysis was carried out following the same methodology described in the previous paragraphs. Likewise, the performance of the model was assessed through Harrell's C-index.

Furthermore, to investigate the prospective applications of this method in follow-up scenarios, synthetic subjects data were employed. Specifically, feature arrays representing the entire spectrum of symptoms (including early/late onset ALS, presence/absence of RSWA, Age, Sex, and DI value) were generated and employed as input variables in a multivariate approach. Finally, predictive survival functions were obtained, over a time period of 365 days, to simulate real-world follow-up procedures.

8.3.3 Results

Preliminary investigation on the RAI

A preliminary investigation on the effect of melatonin treatment on REM Sleep Without Atonia symptomatology was conducted. Particularly, the analysis sought to find a reduction in the observed abnormal activity. However, except for very few and punctual cases, no improvement on the RAI score was observed in the population under study, as appreciable from the data distribution illustrated in Figure 8.1. On the other hand, a slight decrease in RAI emerges when considering the median value.

Hence, although some studies recommend melatonin as first-line treatment for RSWA in RBD cohorts, no statistically significant trend was identified. Indeed, the literature on this topic is rather conflicting [150, 98].

General Survival Model and Feature Importance

From the overall Kaplan-Meier model, a median survival time to the primary end-point of slightly above 6 months was observed. Figure 8.2 shows the overall survival curve of the employed sample.

As regards the multivariate analysis, feature importance, in terms of the C-index, is shown in Table 8.3. The variable DI_{T0} emerged as the best predictor, with a C-index of nearly 0.7, implying good estimation capability. The feature RAI_{T0}, exhibited a C-index value of approximately 0.6, indicating a moderate contribution in the prediction of survival. The co-variate Age appeared with a rather strong value of C-index, though with a non-statistically significant p. Finally, the variable Sex featured a C-index close to 0.5, suggestive of a non-informative prediction rule,

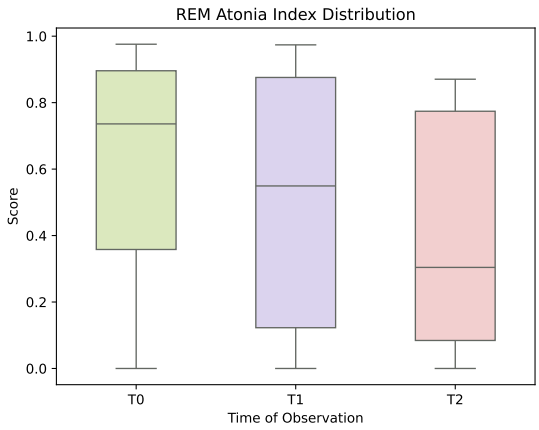


Figure 8.1: Distribution of the REM Atonia Index at time points T0, T1, T2. Due to the presence of only one subject, T3 was discarded in the representation.

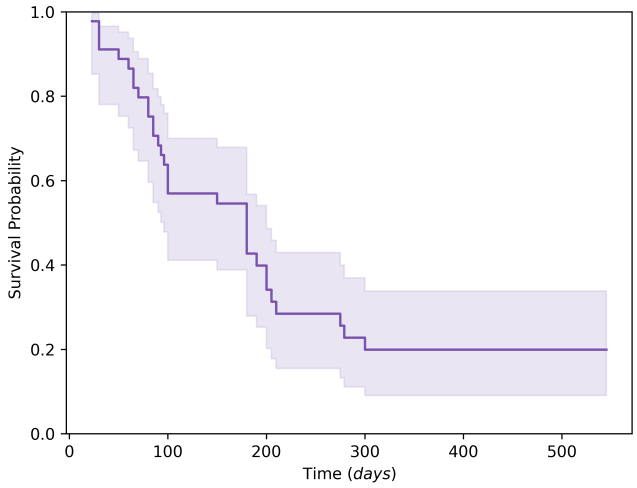


Figure 8.2: Kaplan-Meier plot of the overall survival trend. Time between study enrollment and primary endpoint (decease/inability to follow-up).

however, with a statistically significant p. This outcome, however, might be due to the class imbalance in the REMALS Database.

Stratified Analyses

The following paragraphs present the outcomes of the stratified survival explorations, based on the predictors RSWA(+/−) and Sex.

Table 8.3: Harrell's concordance index for each tested predictor. Statistical significance is shown as *, and indicates a value of $p<0.05$.

Feature	C-index	p-value
DI_{T0}	0.693	0.021*
Age	0.622	0.437
RAI_{T0}	0.572	0.045*
Sex	0.557	0.032*

Results of Stratification According to RSWA As previously described, the selected threshold (RAI=0.8) identified two groups, RSWA+, with 25 subjects, and RSWA−, counting 20 subjects. A Kaplan-Meier model was fit for both groups, and the obtained survival functions are presented in Figure 8.3. Noticeably, both groups present with a similar progression trend. Upon a preliminary visual examination, the RSWA− group appears to initially undergo a more rapid progression, albeit eventually transitioning to a plateau in the later stages of the disease. Indeed, the statistical comparison of the two groups yielded a χ^2 value of 1.817, with p-value=0.5. This outcome corroborated the similarity observed in the preliminary visual inspection of the survival curves, suggesting that stratification based on RSWA may not be informative in this cohort.

Results of Stratification According to Sex The same procedure was carried out on this second stratification, which, however, presented with a considerable group imbalance (30 males and 15 females). The survival function obtained through a Kaplan-Meier approach is displayed in Figure 8.4. From a visual inspection, the Male subgroup appeared to exhibit a more rapid degeneration. The log-rank test conducted on this stratification yielded a χ^2 value of 3.768 (p-value 0.035), indicating statistical significance. This finding confirmed the clinical implications of the disease [95], although, as previously mentioned, the different class numerosity may have had an impact.

Prediction of Disease Progression

Finally, to investigate prospective applications in follow-up strategies, a multivariate Cox Proportional Hazards model, trained on the entire dataset, was employed to make predictions on synthetic, subject-specific data.

The trained model attained a C-index of 0.734, which, as discussed, indicates good prediction performance of the extracted parameters. In Figure 8.5 the predicted survival curves for two relevant synthetic patients are shown. Precisely, these represent one young subject with early-onset ALS and no RSWA at the time of enrollment, and a second, older subject with moderate RSWA at T0. As formerly

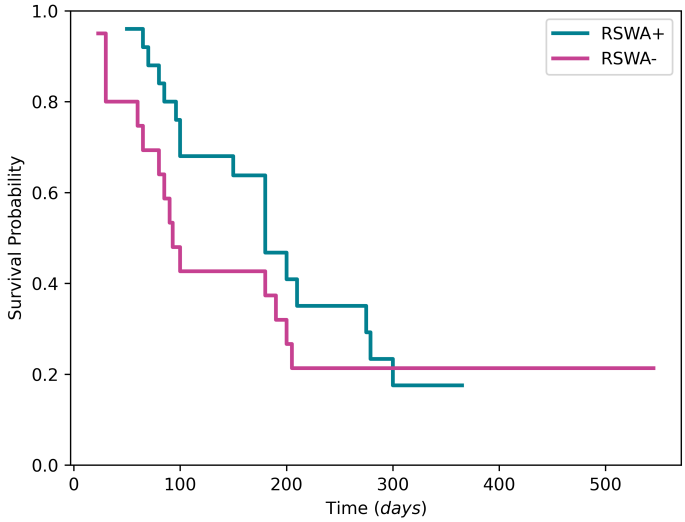


Figure 8.3: Estimated survival curves for the stratified analysis according to the RSWA co-variate, in the time between study entrollment and primary endpoint (death/inability to follow-up).

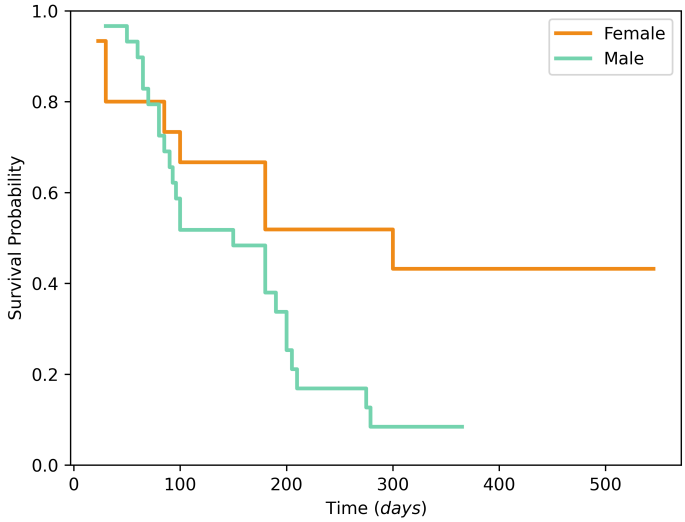


Figure 8.4: Estimated survival curves for the stratified analysis according to the Sex co-variate. in the time between study entrollment and primary endpoint (death/inability to follow-up).

discussed, these curves, previously trained on real subjects data, represent potential estimators of the disease progression rapidity and may serve as predictors of

the expected median survival time. Therefore, they may be used as support tools non-intrusive follow-up strategies.

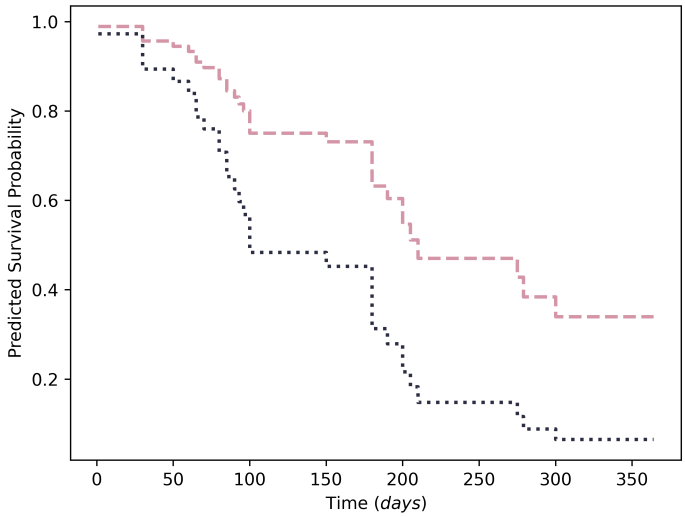


Figure 8.5: Predicted survival probability on synthetic subjects data. Dotted line: older patient, RSWA+. Dashed line: patient with young onset, RSWA–.

8.3.4 Discussion

This preliminary study explored REM sleep muscular patterns in ALS, to assess the feasibility of predicting disease progression through a minimally-intrusive framework. Indeed, lightweight, objective EMG metrics were computed from REM segments and employed as features in survival models.

The extracted predictors were adopted in a multivariate approach, and showed overall good predictive power, with the DI yielding a C-index value close to 0.7.

Although a thorough literature search has been conducted, this exploratory study appeared to be the first tackling survival analyses on unobtrusive muscular parameters on REM sleep. Indeed, prior studies employed invasive wake measurements, including needle EMG or extensive examination of the genioglossus muscle, obtaining statistically significant hazard ratios [158, 169]. Consequently, a direct comparison with studies in the literature is not feasible; however, the results obtained, along with their statistics, are in line with the findings highlighted in previous studies [95, 136, 169], though these latter employed different cohorts, and qualitative co-variates.

Finally, when exploring possible applications in follow-up strategies, predictive modelling of ALS disease progression was proposed, and tested on synthetically

generated data. The multivariate Cox' Proportional Hazards model trained on the REMALS Database achieved C-index of 0.73, indicative of encouraging goodness of fit. Furthermore, the survival curves obtained by testing the model on unseen, synthetic data, yielded different progression trends for the proposed *case studies*, highlighting their potential as digital tools in follow-up protocols.

Although presenting with promising potential, this research activity was exploratory, and presents with some challenges and limitations that should not be left unaddressed, and are discussed in the following paragraph.

8.4 Conclusion

Amyotrophic lateral sclerosis is a chronic neurodegenerative disease, that affects motor neuron. Sleep disturbances are widely present in subjects with a diagnosis of ALS, and they have a negative impact on the quality of life of both patients and caregivers.

Currently, no curative treatment for ALS is available, and pharmacological treatments primarily regards the management of symptoms or palliative care. Amelioration of sleep-disordered breathing symptomatology, such as sleep apnea, can be found through continuous positive airway pressure (CPAP) ventilation, which allows proper airflow. However, this temporary solution becomes impracticable once life-sustaining, invasive ventilation protocols are adopted.

Indeed, sleep disorders should be investigated, to exploit their potential as reservoir of clinical information, both regarding the aetiology and disease evolution.

Previous studies in literature highlighted the correlation of RSWA manifestations with the clinical severity of the disease, and suggested its role as potential risk factor in a more rapid disease progression trends. However, traditional assessments for RSWA relies on in-hospital PSG, which remains a highly intrusive diagnostic tests, and makes follow-up assessments challenging for these populations.

Given the importance of muscular activity analysis in ALS to establish the overall decline rate [180, 148], tools for unobtrusive data collection acquire pivotal importance. In particular, minimally-invasive sleep studies could offer alternative strategies to monitor and manage symptoms, and be beneficial in ALS therapeutic pathways in order to provide personalised care.

For this purpose, the research activity presented in this Chapter aimed at assessing the correlation between abnormal EMG activity during REM and the rate of survival, employing simple, lightweight, and unobtrusive metrics. Specifically, a combination of clinical and demographic data and quantitative muscular parameters were employed. Namely, these were a well-established RSWA metric (i.e., the RAI) and a previously presented EMG-derived metric (DI), this latter representing the extent of dissimilarity from a healthy model.

These parameters emerged as meaningful predictors in the explored survival

models, with the DI assessed at the time of the enrollment yielding an importance score of 0.7, in terms of Harrell's concordance index.

Furthermore, their combination with clinical and demographic parameters was exploited to train survival models and tackle predictive modelling of the disease progression, subsequently tested on synthetic, subject-specific data.

The encouraging performance of the proposed framework, discussed in the Sections above, bolsters the feasibility of lightweight follow-up protocols in these populations. Indeed, the computed muscular parameters are retrieved from surface EMG of the mylohyoid muscle. Due to the muscle positioning, this biosignal could be recorded through minimally-invasive, wearable sensors or patches [110], without the need for high-resolution instrumentation, with respect to traditional PSG.

Although the presented results promise encouraging scenarios for personalised care, this study remains preliminary and future developments should explore its limitations.

In this regard, a larger cohort should be included, to allow for a comprehensive and reliable survival analysis, to improve the robustness of the trained models and enhance the reliability of the proposed framework. Indeed, 58 patients were initially included, but 13 were discarded due to technical issues in the recordings. This occurrence, combined with the high decline rate of the disease under study, significantly affected the size of the sample.

Second, given the interesting, albeit exploratory, findings of the stratified analysis, more detailed differentiation in the dataset should be taken into account, which will necessarily be encouraged by larger cohorts. For instance, meaningful insights into progression rates could be provided by a stratification according to the type of onset (i.e., bulbar, truncal, spinal). Furthermore, the age of disease onset, the functional scale score, or the presence of non-invasive ventilation are important descriptors of self-sufficiency or disability, and should be taken into consideration, as they may emerge as relevant co-variates. However, in this first investigation, these parameters were overlooked. Additionally, the subjects in the dataset were all undergoing melatonin treatment at the time of the study. Although no significant effect of treatment was observed in this cohort, the inclusion of a placebo-controlled group could offer the possibility of evaluating the effect of this type of pharmacological treatment on survival rates.

Upon these considerations, a thorough and more detailed characterisation could be advantageous for building a robust descriptive model of disease evolution, especially in the case of neurodegenerative diseases with rapid progression rates. Eventually, the findings could be generalised to other neurodegenerative disorders with more extended time-scales.

Finally, although one of the aims of this research activity was to assess the feasibility of monitoring ALS progression through lightweight metrics, the employed data were recorded through high-resolution PSG (both through traditional and NOX instrumentation). Future data acquisition campaigns should envisage the use

of wearable set-ups, and process data acquired through different sensors, to test the reliability of real-world, remote scenarios.

Chapter 9

Conclusion

Sleep disorders represent a growing challenge in healthcare, with a global prevalence of up to 70% in older adults, and the majority of cases left undiagnosed [83]. In subjects over the age of 60 they serve as predisposing factors for the development of co-morbidities, with a negative impact on the quality of life and an increase on the healthcare burden [73].

They encompass a broad variety of conditions, including insomnia or disorders of hypersomnolence, sleep-disordered breathing, parasomnias, and motor-related symptomatic manifestations.

Emerging evidence suggested their involvement in the neuronal degeneration processes [170], and acknowledged them among the earliest markers of neurodegenerative diseases. Indeed, RBD is considered as a prodrome of α-synucleinopathies, including PD, and dysfunctions in slow-wave sleep have been associated with an increased risk for the development of AD.

Hence, sleep disorders acquire great potential as reservoir of significant clinical information. Their investigation may thus provide insightful understanding of the aetiology and progression pattern of various neurodegenerative diseases, and positively impact the therapeutic pathways.

However, state-of-the-art clinical assessment methods rely, on a large scale, on obtrusive diagnostic tests, to be performed in laboratory or outpatient settings. Polysomnographic recordings are the gold standard assessment tool for diagnosing and monitoring sleep disorders, and require cumbersome, wired instrumentation, and require extensive manual labour. Indeed, the international guidelines for PSG assessment currently require the visual inspection of overnight recordings and manual scoring following strict sets of rules, ofttimes resulting in intra- and interrater variability, and protracted diagnosis time.

This Thesis sought to investigate the feasibility of minimally-invasive sleep studies in neurodegenerative diseases, with the aims of providing possible frameworks to **(1)** support the diagnostic process, and **(2)** facilitate monitoring and follow-up.

For this purpose, data-driven, automatic approaches for processing sleep data

have been explored. Precisely, polysomnographic biosignals, such as EEG and EMG, and motility data collected through IMUs were processed to obtain objective parameters for the characterisation of sleep disorders. These parameters were leveraged in combination with supervised ML models, to tackle automatic detection of patterns, or the assessment of the disease progression. This procedure aimed at obtaining obtain simple and straightforward analysis frameworks, that could prove realiable when integrated in the clinical practice.

The research work was conducted on a multidisciplinary approach, involving in some cases a complete pathway from the study design to data analysis and interpretation, conducted in cooperation with Neurology teams. The experimental studies for RBD were conducted in close cooperation with clinical units at two facilities: Molinette University Hospital (Turin, Italy), and the Medical University of Innsbruck (Austria). The research projects on Parkinson's Disease and Amyotrophic Lateral Sclerosis were devised and conducted in collaboration with the Neurology Department and the Neuroscience Department of the University of Turin (Italy), in the *Azienda Ospedaliero-Universitaria Città della Salute e della Scienza*. The experimental activities entailed in many cases the interaction with patients, especially when regarding data collection, both in outpatient settings and at local Patients Associations (*Associazione Amici Parkinsoniani Piemonte ONLUS*).

The research works included in this Thesis explored various topics, and tackled the following challenges:

- **Automatic Sleep Staging**, to assess the feasibility of minimally-intrusive strategies in PSG, based on single-channel EEG,
- **Quantification of REM Sleep Without Atonia**, to remove the need for manual scoring of artefact, and proceed towards the automatic assessment of RSWA to facilitate diagnostic pathways,
- **Automatic Detection of RBD**, through supervised ML, employing a minimal set of sensors (single-channel EEG or EMG), to improve diagnostic strategies in RBD,
- **Assessment of RBD Progression**, through a continuous metric, presenting a possible approach to personalised follow-up procedures,
- **Home-Monitoring of Sleep Disorders**, to objectively characterise nocturnal hypokinesia and akinesia in PD through wearable, low-cost instrumentation, and assist remote monitoring frameworks,
- **Prediction of Disease Progression**, to provide insight into the progression of ALS through simple and unobtrusive EMG metrics, facilitating sleep studies in fragile subjects.

The proposed methodology, explored in Chapters 4–8, presented with encouraging outcomes, and:

- Provided effective and robust frameworks for enabling minimally-invasive sleep studies and supporting the clinical diagnosis of RBD,
- Revealed the feasibility of home-based, continuous remote monitoring solutions for sleep disorders in populations with neurodegenerative diseases.

The obtained results may represent a possible approach to the development of personalised treatment and care, for beneficial quality of care. However, as the pathophysiology of neurodegenerative disorders is still unclear, symptomatic treatment is the only available option to improve the quality of life of patients. Currently, a pivotal role is played by cognitive and motor rehabilitation, especially in PD. During the course of the 3-year PhD project, efforts have been dedicated towards the development of a personalised, adaptive Tele-Rehabilitation framework, to provide an exhaustive implementation of continuous follow-up strategies. These activities are discussed in the Appendix A.4, and were not included in the main document as they relate to *awake* biosignals.

To summarise, the research works presented in this Thesis proved the feasibility of objectively monitoring sleep through data-driven, straightforward metrics, and offered an understanding in sleep patterns in neurodegenerative diseases, presenting possible approaches for accessible and lightweight sleep monitoring.

Future trajectories shall embrace the technological challenges of the presented scenario, and tackle the limitations of the current research activity.

First, Deep Learning techniques should be explored, to identify possible underlying patterns in sleep architecture or sleep disorders. Second, future developments shall envisage the adoption of consumer-grade, wearable technology to monitor sleep in unsupervised settings, and provide accurate follow-up strategies through remote monitoring. Finally, data-driven frameworks should be employed to conduct population screening studies and enhance neurodegenerative risk prediction strategies.

Appendix A

Supplementary Material

A.1 Inclusion Criteria (Observational Studies)

The following Sections report the inclusion and exclusion criteria for the patient cohorts recruited at the Department of Neurology, Molinette University Hospital (Turin, Italy), and included in the observational studies described in this Thesis.

Data Availability

All data in the TuSDi Database are available on request and are currently not published in public repositories due to privacy and ethical reasons.

Ethical Committee

The collection of data in all presented studies was conducted in accordance with the Declaration of Helsinki. Therefore, all participants received detailed information on the study purposes and execution, and written informed consent for *observational study* was obtained. All procedures have been approved by the Ethics Committee of A.O.U. Città della Salute e della Scienza di Torino (Approval No. 00384/2020).

A.1.1 TuSDi Database: RBD Detection from EMG

For this study, inclusion criteria for RBD patients in the TuSDi Database were:

- Suspected or diagnosed RBD [1],
- Suspected narcolepsy,

[1] For the diagnosis: (1) occurrence of dream enactment confirmed by vPSG, (2) polysomnographic evidence of RSWA, (3) clinical interviews with a sleep neurologist.

- Suspected REM dissociation or other NREM parasomnias,
- Diagnosis of secondary RBD due to OSAS,
- Elevated EMG tone during REM sleep.

Exclusion criteria for RBD patients were:

- Dementia,
- Diagnosis or clinical history of psychiatric conditions that could prevent the correct execution of PSG.

A.1.2 TuSDi Database: RBD Detection from EEG

For this study, inclusion criteria for RBD patients in the TuSDi Database were:

- Suspected or diagnosed RBD[2],
- Elevated EMG tone during REM sleep,
- Absence of nocturnal epilepsy.

Exclusion criteria for RBD patients were:

- Dementia,
- Diagnosis or clinical history of psychiatric conditions that could prevent the correct execution of PSG,
- Ongoing pharmacological treatment with benzodiazepines.

A.1.3 PDSleep Database

For this study, inclusion criteria for the PD cohort were:

- Clinical diagnosis of Parkinson's Disease,
- Clinical assessment of nocturnal hypokinesia or akinesia,
- Reported or suspected nocturnal motor impairment.

Exclusion criteria for the PD cohort were:

- Dementia,

[2]Same criteria as A.1.1

- Diagnosis or clinical history of psychiatric conditions that could prevent the correct execution of PSG,
- Ongoing pharmacological treatment with benzodiazepines.

As regards the healthy cohort, inclusion criteria envisaged:

- Absence of family history for neurodegenerative disorders,
- Absence of diagnosed sleep disorders,
- Absence of reported sleep disorders.

A.1.4 REMALS Database

For this study, inclusion criteria for the ALS cohort were:

- Clinical diagnosis of Amyotrophic Lateral Sclerosis,
- Polysomnographic evidence of REM Sleep Without Atonia.

Exclusion criteria for the PD cohort were:

- Dementia,
- Diagnosis or clinical history of psychiatric conditions that could prevent the correct execution of PSG,
- Physical inability to undergo polysomnography.

As regards the healthy cohort, inclusion criteria envisaged:

- Absence of family history for neurodegenerative disorders,
- Absence of diagnosed sleep disorders,
- Absence of REM Sleep Without Atonia,
- Absence of reported sleep disorders.

A.2 Design of the Dissociation Index

In Section 6.3.2, the rationale behind the design of the Dissociation Index is presented. This metric relies on two main concepts:

- A reference (*"healthy"*) model (S),
- A neighbourhood (R), virtually mirroring the search range.

Various configurations were explored before applying the final one, and are described in the following paragraphs.

Specifically, considering that no clinically validated model of RBD progression is currently available, the concept the neighbourhood was tentatively expressed in two manners:

1. **Radius** R_1: where R expresses the ED between the HS with the best atonia score (i.e., highest RAI) and the RBD subject with the worst atonia score (i.e., lowest RAI). This virtually depicts a range between the "best" and the "worst" feature arrays in the dataset.

2. **Radius** R_2: where R expresses the ED between two points in a plane, corresponding to the whole HS and RBD groups, respectively. The two points are, in turn, obtained as the intra-group ED.

The second definition provides a more comprehensive neighbourhood, as the radius is defined by encompassing the intrinsic characteristics of each group. The values were computed on the CAP Sleep Database, as it encompassed both healthy and RBD subjects, ensuring balance. The obtained values were 6.2 and 5.92 for R_1 and R_2, respectively.

Accordingly, two reference models were explored:

1. **Reference** S_1: the feature array of the subject in the HS group with the best atonia score (i.e., highest RAI).

2. **Reference** S_2: The sample mean of all feature arrays in the healthy cohort. As seen in R_2, this provides a comprehensive measure to describe the intra-group variability.

Afterwards, the ED of each subject (expressed as feature array) from the selected reference were computed, obtaining distances H_i, and mapped to the DI as described in Section 6.3.2.

The following combinations were investigated: (1) R_1/S_1, (2) R_2/S_1, (3) R_2/S_2; Table A.1 displays the results, in terms of distance and included subjects.

Neighbourhood radius R_2 appeared to be more restrictive; finally, combination (3) appeared to be the most robust measure to evaluate the extent of similarity to an estimated healthy model, and was employed in the subsequent assessment of DI.

	$\mathbf{R}_1/\mathbf{S}_1$	$\mathbf{R}_2/\mathbf{S}_1$	$\mathbf{R}_2/\mathbf{S}_2$
Euclidean Distance ($\mu \pm \sigma$)	5.35 ± 0.63	5.24 ± 0.59	5.02 ± 0.6
Maximum Distance	6.19	5.80	5.81
Subjects Outside Radius (N)	5	7	6

Table A.1: Combinations of neighbourhood and reference model explored for the design of the Dissociation Index.

A.3 Sleep in Parkinson's Disease

This section includes the set of questions of the sleep surveys administered in Paper [140] to assess sleep quality and circadian habits, and the text used for vocal samples.

A.3.1 Shortened Pittsburgh Sleep Quality Index

The following questions were proposed in the shortened version of the Pittsburgh Sleep Quality Index [55], and relate to usual sleep habits adopted in the month prior to the examination.

1. During the past month, when have you usually gone to bed?
2. During the past month, how long (in minutes) has it taken you to fall asleep each night?
3. During the past month, when have you usually gotten up in the morning?
4. During the past month, how many actual hours of sleep did you get at night? (This may be different than the number of hours you spend in bed.)
5. During the past month, how often have you had trouble sleeping because you...
   (a) Cannot get to sleep within 30 minutes
      i. Not during the past month
      ii. Less than once a week
      iii. Once or twice a week
      iv. Three or more times a week
   (a) Wake up in the middle of the night or early morning
      i. Not during the past month

ii. Less than once a week
iii. Once or twice a week
iv. Three or more times a week

(a) Cannot breathe comfortably
i. Not during the past month
ii. Less than once a week
iii. Once or twice a week
iv. Three or more times a week

(a) Cough or snore loudly
i. Not during the past month
ii. Less than once a week
iii. Once or twice a week
iv. Three or more times a week

(a) Feel too hot
i. Not during the past month
ii. Less than once a week
iii. Once or twice a week
iv. Three or more times a week

(a) Have bad dreams
i. Not during the past month
ii. Less than once a week
iii. Once or twice a week
iv. Three or more times a week

(a) Have pain
i. Not during the past month
ii. Less than once a week
iii. Once or twice a week
iv. Three or more times a week

6. During the past month, how often have you had trouble staying awake while driving, eating meals, or engaging in social activity?

(a) Not during the past month

(b) Less than once a week

(c) Once or twice a week

(d) Three or more times a week

7. During the past month, how much of a problem has it been for you to keep up enthusiasm to get things done?

(a) Not during the past month

(b) Less than once a week

(c) Once or twice a week

(d) Three or more times a week

Scoring Instructions

The scoring instructions, provided in [55] follow.

Component 1: sleep latency

C1 —

Score of #**2** (<15 (0), 16–30 min (1), 31–60 min (2), >60 min (3)) + Score of #**5a** (if sum is equal 0 = 0; 1-2 = 1; 3-4 = 2; 5-6 = 3)

Component 2: sleep duration

C2 —

Score of #**4** (>7 (0), 6-7 (1), 5-6 (2), <5 (3))

Component 3: sleep efficiency

C3 —

(total # of hours asleep)/(total # of hours in bed) × 100; >85% = 0, 75–84% = 1, 65–74% = 2, <65% = 3

Component 4: sleep disturbances

C4 —

Examine questions **5b** to **5g** and assign scores for each questions as follows:
Response: Score

- Not during the past month: 0
- Less than once a week: 1
- Once or twice a week: 2
- Three or more times a week: 3

Get the sum of scores **5b** to **5g** (0 = 0, ≥1 ≤6 = 1; >6 ≤12 = 2; >12 = 3)

Component 5: daytime dysfunction

C5 —

Score of #**6** + Score of #**7** (0 = 0; 1-2 = 1; 3-4 = 2; 5-6 = 3)

A.3.2 SLEEPS Questionnaire

This Section includes the original set of questions included in the SLEEPS Questionnaire, proposed by our group in [3]. The items and questions were drafted in cooperation with sleep neurologists.

1. General Data

1. Diagnosis of COVID-19
   (a) Present diagnosis
   (b) Past diagnosis
   (c) Never diagnosed
2. Diagnosis of obstructive sleep apnea
   (a) Yes
   (b) Casual occurrence of sleep apnea
   (c) Never diagnosed
3. Presence of insomnia
   (a) Yes
   (b) Seldom
   (c) No
4. Presently enrolled at university
   (a) Yes
   (b) No
   (c) Not Applicable
5. Presently employed
   (a) Yes
   (b) Shift-Worker
   (c) No

2. Work and Study Routine

1. Presently attending remote class, or working from home
   (a) Yes
   (b) No
   (c) Not Applicable

2. Hours of screen time
    (a) Present diagnosis
    (b) Past diagnosis
    (c) Never diagnosed
3. Time of night indicating end of use of electronic devices (–)
4. Use of a blue light filter (Y/N)

3. Leisure Time Habits

1. Time spent outside over the working week
    (a) 0–2 hours
    (b) 3–5 hours
    (c) 6–8 hours
    (d) 9+ hours
2. Average time per day spent outdoors in the weekend
    (a) 0–2 hours
    (b) 3–5 hours
    (c) 6–8 hours
    (d) 9+ hours
3. Average time per week spent training outdoors
    (a) 0–2 hours
    (b) 3–5 hours
    (c) 6–8 hours
    (d) 9+ hours
4. Average time per week spent training indoors
    (a) 0–2 hours
    (b) 3–5 hours
    (c) 6–8 hours
    (d) 9+ hours
5. Average time per week dedicated to hobbies
    (a) 0–2 hours
    (b) 3–5 hours

(c) 6–8 hours

(d) 9+ hours

4. Sleep Habits

1. Presence of arousals

   (a) Up to once every night

   (b) More than once every night

   (c) Never occurred

2. Frequency of getting out of bed during the night

   (a) Up to once every night

   (b) More than once every night

   (c) Never occurred

3. Difficulty of getting up in the morning

   (a) Less than twice a week

   (b) More than twice a week

   (c) Never occurred

4. Morning tiredness

   (a) Less than twice a week

   (b) More than twice a week

   (c) Never occurred

5. Daytime fatigue

   (a) Less than twice a week

   (b) More than twice a week

   (c) Never occurred

6. Difficulty in falling asleep, although being tired

   (a) Less than twice a week

   (b) More than twice a week

   (c) Never occurred

7. On a scale from 0 to 4, how would define the quality of your sleep?

The preliminary analysis conducted in [3] highlighted questions **(1.3)**, **(2.3)**, **(4.1)**, **(4.3)**, **(4.4)**, **(4.5)**, **(4.6)** to be the most relevant in correlation with perceived sleep quality. Upon this premise, only this subset of questions was administered to the study participants. Scoring instructions follow.

Scoring Instructions

The final SLEEPS score is computed by summing the item-wise scores as follows.
For question **(1.3)**: 2 if (a), 1 if (b), 0 if (c).
For question **(4.1)**: 1 if (a), 2 if (b), 0 if (c).
Sum of scores for **(4.3)** and **(4.4)**, as:

- For question **(4.3)**: For question **(4.1)**: 1 if (a), 2 if (b), 0 if (c).
- For question **(4.4)**: For question **(4.1)**: 1 if (a), 2 if (b), 0 if (c).

Then, if the sub-score ≥ 3: 2, if sub-score equals 2: 1, if sub-score ≤ 1: 0

Sum of scores for **(4.5)** and **(4.6)**, as:

- For question **(4.5)**: For question **(4.1)**: 1 if (a), 2 if (b), 0 if (c).
- For question **(4.6)**: For question **(4.1)**: 1 if (a), 2 if (b), 0 if (c).

Then, if the sub-score ≥ 3: 2, if sub-score equals 2: 1, if sub-score ≤ 1: 0

A.3.3 Vocal Sample Recordings

For future investigations, the participants were asked to record through a smartphone a phonemically balanced test, as proposed in a previous investigation [3]. The text, in Italian language, was proposed in [48] and is reported below.

Il Ramarro Della Zia

Il papà (o il babbo come dice il piccolo Dado) era sul letto. Sotto di lui, accanto al lago, sedeva Gigi, detto Ciccio, cocco della mamma e della nonna. Vicino ad un sasso c'era una rosa rosso vivo e lo sciocco, vedendola, la volle per la zia. La zia Lulù cercava zanzare per il suo ramarro, ma dato che era giugno (o luglio non so bene) non ne trovava. Trovò invece una rana che saltando dalla strada finì nel lago con un grande spruzzo. Sai che fifa, la zia! Lo schizzo bagnò il suo completo rosa che divenne giallo come un taxi. Passava di lì un signore cosmopolita di nome Sardanapalo Nabucodonosor che si innamorò della zia e la portò con sé in Afghanistan.

A.4 Other Works

NeAdEx: Neuroadaptive Exergame for Telerehabilitation Purposes

With the rapid advancements in wireless technologies and video-processing techniques, Telehealth applications have rapidly grown due to their versatility, portability and accessibility. These strategies allow patients living in rural or remote areas to benefit from continuity of care, while significantly reducing healthcare costs. Motor and cognitive rehabilitation is usually carried out in hospitals, gyms or clinical facilities, but technological devices in this field offer more flexibility (i.e., the same task can be adapted to different applications and needs), high interactivity and the possibility of more immersive tasks. Moreover, these technologies offer both qualitative and quantitative feedback, available as soon as the cognitive task is completed.

As regards motor-cognitive telerehabilitation, a possible solution is offered by the so-called *Exergames*, that exploit new generation devices such as RGB-Depth cameras (e.g., Kinect) [4], balance boards (e.g., WII balance board), and virtual-reality headsets. The users can control the game through their own body and carry out goal-oriented tasks, which aim at stimulating specific motor and cognitive skills.

Within this framework, efforts have been dedicated towards the development of an interactive telerehabilitation framework, that could adapt to the subjects' needs. The project (*Neuroadaptive Exergame*, NeAdEx) envisaged the development of an exergame with a series of tasks mimicking exercises for the rehabilitation of the upper limbs, and the simultaneous recording of biosignals (EEG and electrodermal activity) through research-grade wearable devices. Specifically, the Dreem2 headset (Beacon Biosignals, Inc., USA, https://beacon.bio/) was employed to collect EEG data, and the Empatica E4 (Empatica Inc., USA, https://www.empatica.com/en-eu/) to record electrodermal activity (EDA) and inertial measures.

The project aimed at:

- Characterising subject's engagement through simple EEG-derived metrics [5],
- Assessing the stress level of participants during game-play [111].
- Investigate the effect of baseline stress on the performance.

The results of the first explorations are included in [5] and [111], and in one conference paper accepted for presentation at the 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC2024).

The experimental investigation is currently ongoing, and including healthy cohorts, subjects with PD, and subjects affected by metabolic disorders. The exploration of electrophysiological trends during Telerehabilitation would provide insights into the subjects' behaviours and identify trends in the response to the exergame. Ultimately, the investigation of these physiological parameters could support the development would allow for the development of adaptive telerehabilitation strategies without the need for expensive and invasive equipment, with beneficial effects on the quality of care.

Appendix B

Included Contributions

B.1 Automatic Sleep Staging

Title: Single-channel EEG classification of sleep stages based on REM microstructure
Published in: Healthcare Technology Letters, *John Wiley & Sons, Ltd.*, (2021)
Authors: Irene Rechichi, Maurizio Zibetti, Luigi Borzì, Gabriella Olmo, Leonardo Lopiano.

Abstract: Rapid-eye movement (REM) sleep, or paradoxical sleep, accounts for 20–25% of total night-time sleep in healthy adults and may be related, in pathological cases, to parasomnias. A large percentage of Parkinson's disease patients suffer from sleep disorders, including REM sleep behaviour disorder and hypokinesia; monitoring their sleep cycle and related activities would help to improve their quality of life. There is a need to accurately classify REM and the other stages of sleep in order to properly identify and monitor parasomnias. This study proposes a method for the identification of REM sleep from raw single-channel electroencephalogram data, employing novel features based on REM microstructures. Sleep stage classification was performed by means of random forest (RF) classifier, K-nearest neighbour (K-NN) classifier and random Under sampling boosted trees (RUSBoost); the classifiers were trained using a set of published and novel features. REM detection accuracy ranges from 89% to 92.7%, and the classifiers achieved a F-1 score (REM class) of about 0.83 (RF), 0.80 (K-NN), and 0.70 (RUSBoost). These methods provide encouraging outcomes in automatic sleep scoring and REM detection based on raw single-channel electroencephalogram, assessing the feasibility of a home sleep monitoring device with fewer channels.
DOI: https://doi.org/10.1049/htl2.12007

B.2 Automatic Detection of Artefacts

Title: Towards fully automatic quantification of REM sleep without atonia according to the Sleep Innsbruck Barcelona (SINBAR) scoring method

Published in: Abstracts from the 17th World Sleep Congress, Sleep Medicine (Supplements), *Elsevier*, (2023)

Authors: Irene Rechichi, Gabriella Olmo, Ambra Stefani, Anna Heidbreder, Evi Holzknecht, Melanie Bergmann, Abubaker Ibrahim, Elisabeth Brandauer, Birgit Högl, Matteo Cesari

Abstract:

Introduction: Rapid eye movement (REM) sleep without atonia (RWA) is the polysomnographic hallmark of REM Sleep Behavior Disorder (RBD). The state-of-the-art methods to score RWA are visual-based. Recent international guidelines recommended the Sleep Innsbruck Barcelona (SINBAR) method for scoring RWA in 3-s mini-epochs. This method calls for scoring phasic EMG activity in the flexor digitorum superficialis (FDS) and "any" (i.e., tonic and/or phasic) EMG activity in the mentalis muscle. A semi-automatic algorithm scoring RWA according to this method is currently available in a commercial polysomnographic system (BrainRT, OSG, Belgium), however it still requires manual removal of EMG artifacts from expert scorers. This work proposes a novel method that, based on morphological aspects of EMG activity and machine learning (ML), discriminates activity from artifacts in the evaluation of RWA, thus allowing automatization for artifact correction.

Materials and Methods: We included video-polysomnography studies of 25 participants (8 RBD, 17 controls, aged 57.2 ± 14.9 years). An expert scorer selected 3-s mini-epochs for RWA scoring (956 ± 70) and the BrainRT for scoring phasic and "any" EMG activity was applied. Four independent expert scorers manually removed artifacts; probabilistic consensus of the four scorers was obtained.

The algorithm for automatic removal of artifacts consisted in the following. First, wavelet transform (biorthogonal mother wavelet, matching EMG activity waveform) was applied to the selected 3-s mini-epochs in the mentalis and FDS EMG signals. Second, the coefficients at the third level of decomposition were employed in the signal reconstruction process. Third, an index of correlation, expressed as ratio between the wavelet-reconstructed signal and the original signal, was computed. Fourth, an energy-based metric, expressed as the 90th percentile of the EMG activity, was computed. The correlation index and the energy-based metric were subsequently employed as features in a binary classification task, to automatically differentiate artifacts from the phasic and "any" activity; gold standard was the consensus scoring. Supervised ML models (support vector machine, K-nearest neighbour, linear discriminant analysis, and adaptive boosting) were explored. Seventeen participants were included in the training set and eight in the test set. The models' performances were evaluated with accuracy and F1-score. Finally, RWA metrics – i.e., phasic mentalis, phasic FDS, any mentalis, SINBAR (i.e., any mentalis and/or phasic FDS – were computed on the automatically corrected EMG signals and compared with Pearson's correlation to the ones obtained by the consensus manually scoring.

Results: The explored ML models scored fairly good results. The best models yielded a F-1 score of 89.59% for phasic activity (mentalis and FDS muscles combined) and 76,56% for any. Finally, score agreement between manually and automatically corrected data had Pearson's rho of 0.96, 0.82, 0.76 and 0.71 for phasic mentalis, phasic FDS, any mentalis and SINBAR indices on the test subjects, respectively.

Conclusions: The proposed method, based on EMG activity morphology and ML, showed promising results in discriminating artifacts from real phasic and any EMG activity for RWA quantification. This method promises fully automatic RWA quantification according to the SINBAR method.

DOI: https://doi.org/10.1016/j.sleep.2023.11.834

B.3 Automatic Detection of RBD

B.3.1 EMG-Based RBD Detection

Title: Assessing rem sleep behaviour disorder: From machine learning classification to the definition of a continuous dissociation index

Published in: International Journal of Environmental Research and Public Health, *MDPI*, (2021)

Authors: Irene Rechichi, Antonella Iadarola, Maurizio Zibetti, Alessandro Cicolin, Gabriella Olmo.

Abstract:

Objectives: Rapid Eye Movement Sleep Behaviour Disorder (RBD) is regarded as a prodrome of neurodegeneration, with a high conversion rate to α-synucleinopathies such as Parkinson's Disease (PD). The clinical diagnosis of RBD co–exists with evidence of REM Sleep Without Atonia (RSWA), a parasomnia that features loss of physiological muscular atonia during REM sleep. The objectives of this study are to implement an automatic detection of RSWA from polysomnographic traces, and to propose a continuous index (the Dissociation Index) to assess the level of dissociation between REM sleep stage and atonia. This is performed using Euclidean distance in proper vector spaces. Each subject is assigned a dissociation degree based on their distance from a reference, encompassing healthy subjects and clinically diagnosed RBD patients at the two extremes.

Methods: Machine Learning models were employed to perform automatic identification of patients with RSWA through clinical polysomnographic scores, together with variables derived from electromyography. Proper distance metrics are proposed and tested to achieve a dissociation measure.

Results: The method proved efficient in classifying RSWA vs. not-RSWA subjects, achieving an overall accuracy, sensitivity and precision of 87%, 93% and 87.5%, respectively. On its part, the Dissociation Index proved to be promising in measuring the impairment level of patients.

Conclusions: The proposed method moves a step forward in the direction of automatically identifying REM sleep disorders and evaluating the impairment degree. We believe that this index may be correlated with the patients' neurodegeneration process; this assumption will undergo a robust clinical validation process involving healthy, RSWA, RBD and PD subjects.

DOI: https://doi.org/10.3390/ijerph19010248

B.3.2 EEG-Based RBD Detection

Title: Single-Channel EEG Detection of REM Sleep Behaviour Disorder: The Influence of REM and Slow Wave Sleep

Part of: International Work-Conference on Bioinformatics and Biomedical Engineering, IWBBIO 2022, Lecture Notes in Computer Science (Part I), *Springer*, (2022)

Authors: Irene Rechichi, Federica Amato, Alessandro Cicolin, Gabriella Olmo.

Abstract: Sleep Disorders have received much attention in recent years, as they are related to the risk and pathogenesis of neurodegenerative diseases. Notably, REM Sleep Behaviour Disorder (RBD) is considered an early symptom of -synucleinopathies, with a conversion rate to Parkinson's Disease (PD) up to 90%. Recent studies also highlighted the role of disturbed Non-REM Slow Wave Sleep (SWS) in neurodegenerative diseases pathogenesis and its link to cognitive outcomes in PD and Dementia. However, the diagnosis of sleep disorders is a long and cumbersome process. This study proposes a method for automatically detecting RBD from single-channel EEG data, by analysing segments recorded during both REM sleep and SWS. This paper inspects the underlying microstructure of the two stages and includes a comparison of their performance to discuss their potential as markers for RBD. Machine Learning models were employed in the binary classification between healthy and RBD subjects, with an 86% averaged accuracy on a 5-fold cross-validation when considering both stages. Besides, SWS features alone proved promising in detecting RBD, scoring a 91% sensitivity (RBD class). These findings suggest the applicability of an EEG-based, low-cost, automatic detection of RBD, leading to potential use in the early diagnosis of neurodegeneration, thus allowing for disease-modifying interventions.

DOI https://doi.org/10.1007/978-3-031-07704-3, pages 381–394.

B.4 Monitoring Sleep Disorders in PD

Title: Home Monitoring of Sleep Disturbances in Parkinson's Disease: A Wearable Solution

Published in: 2024 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops), (2024)

Authors: Irene Rechichi, Luca Di Gangi, Maurizio Zibetti, Gabriella Olmo.

Abstract: Sleep Disorders are the most common and disabling non-motor manifestations of Parkinson's Disease (PD), significantly impairing the quality of life. Monitoring sleep disturbances in PD is a complex task, given the lack of objective metrics and the infrequent neurological assessments. This study proposes a framework for the detection of PD sleep patterns from data collected from 40 subjects (12 PD) through a wearable inertial measurement unit (IMU) during sleep, as well as the automatic assessment of sleep quality. Several features describing overnight motility are proposed and employed in Machine Learning (ML) models to carry out the classification. The best model achieved a 96.2% Accuracy and 93.4% F-1 score in detecting PD subjects from controls, in a Leave-One-Subject-Out cross-validation approach. Sleep quality was assessed with an average accuracy of 79.7% $\pm$ 4.4 across the three tested classifiers, and 75% $\pm$ 5.25 F-1 score. This suggests the feasibility of characterising overnight motility in PD and effectively monitoring the symptoms' progression through lightweight technology, in a pervasive, e-Health scenario.

DOI: 10.1109/PerComWorkshops59983.2024.10502893

B.5 REM Sleep Without Atonia in ALS

Title: Predicting Amyotrophic Lateral Sclerosis Progression: an EMG-based Survival Analysis

Accepted for presentation at: 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2024)

Authors: Irene Rechichi, Gianluca Amprimo, Alessandro Cicolin, Gabriella Olmo.

Abstract: ALS is a progressive neurodegenerative disease, ultimately leading to muscle inefficiency and death. A vast majority of people with ALS also suffer from sleep disorders. Previous studies highlighted the presence of RSWA in an ALS cohort, and suggested its strong correlation with the disease severity. This study investigates the ability of EMG parameters recorded during REM sleep to predict disease progress and outcome rapidity in ALS. Survival models trained on a cohort of 45 ALS patients undergoing a longitudinal study, revealed a promising predictive power for the proposed EMG-derived metrics (c-index ≥ 0.65) and encouraging goodness of fit (through c-index and χ^2). These results suggest the possibility of employing the trained model in follow-up procedures, based on non-invasive, lightweight EMG metrics, which would significantly ease disease monitoring and help personalized symptomatic care.

Appendix C

Acronyms

AASM American Academy of Sleep Medicine

Aβ Amyloid-β

AD Alzheimer's Disease

AHI Apnea-Hypopnea Index

AI Artificial Intelligence

ALS Amyotrophic Lateral Sclerosis

ApEn approximate entropy

AUC Area Under the Curve

CSF cerebrospinal fluid

CPAP continuous positive airway pressure

CV Cross-Validation

CWT Continuous Wavelet Transform

DI Dissociation Index

DL Deep Learning

DBS deep brain stimulation

DT Decision Tree

DWT Discrete Wavelet Transform

ECG electrocardiography

ED Euclidean Distance

EMG electromyography

EEG electroencephalography

EOG electrooculography

FDS flexor digitorum superficialis

FN False Negative

FP False Positive

FREM phasic REM

ICSD-3 International Classification of Sleep Disorders

IIR infinite-impulse-response

IMU inertial measurement unit

IQR interquartile range

KDE Kernel Density Estimation

KNN K-Nearest Neighbour

LDA Linear Discriminant Analysis

LOO-CV Leave-One-Out CV

LOSO-CV Leave-One-Subject-Out CV

L-dopa Levodopa

MCRF magnocellular reticular formation

MDS-UPDRS Unified Parkinson's Disease Rating Scale

ML Machine Learning

MAE Mean Average Error

MS motor symptoms

MSE Mean Squared Error

ND neurodegenerative diseases

NMS non-motor symptoms

NREM non-rapid eye movement

PD Parkinson's Disease

PDSS Parkinson's Disease Sleep Scale

PSD power spectral density

PSG polysomnography

PSQI Pittsburgh Sleep Quality Index

RAI REM Atonia Index

RBD REM Sleep Behaviour Disorder

REM rapid eye movement

RF Random Forest

ROC Receiver Operating Characteristic

RSWA REM Sleep Without Atonia

RUSBoost Random Under Sampling Boosting

SD sleep disorders

SE sleep efficiency

SEF spectral edge frequency

SINBAR Sleep Innsbruck Barcelona

SLD sublaterodorsal nucleus

SO slow oscillations

SOL sleep onset latency

SQ sleep quality

STFT Short Time Fourier Transform

SVM Support Vector Machine

SWA slow-wave activity

SWS slow-wave sleep

TA tibialis anterior

TIB time in bed

TKEO Teager-Kaiser energy operator

TLA Three Letter Acronym

TN True Negative

TP True Positive

TREM tonic REM

TST total sleep time

XGBoost eXtreme Gradient Boosting

Bibliography

[1] U Rajendra Acharya et al. "Automated diagnosis of epileptic EEG using entropies". In: *Biomedical signal processing and control* 7.4 (2012), pp. 401–408.

[2] Abdullah Alanazi. "Using machine learning for healthcare challenges and opportunities". In: *Informatics in Medicine Unlocked* 30 (2022), p. 100924.

[3] Federica Amato et al. "Sleep Quality through Vocal Analysis: a Telemedicine Application". In: *2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. IEEE. 2022, pp. 706–711.

[4] Gianluca Amprimo et al. "Assessment tasks and virtual exergames for remote monitoring of Parkinson's disease: An integrated approach based on Azure Kinect". In: *Sensors* 22.21 (2022), p. 8173.

[5] Gianluca Amprimo et al. "Measuring Brain Activation Patterns from Raw Single-Channel EEG during Exergaming: A Pilot Study". In: *Electronics* 12.3 (2023), p. 623.

[6] Peter Anderer et al. "Overview of the hypnodensity approach to scoring sleep for polysomnography and home sleep testing". In: *Frontiers in Sleep* 2 (2023), p. 1163477.

[7] James W Antony et al. "Sleep spindles and memory reprocessing". In: *Trends in neurosciences* 42.1 (2019), pp. 1–3.

[8] Pierrick J Arnal et al. "The Dreem Headband compared to polysomnography for electroencephalographic signal acquisition and sleep staging". In: *Sleep* 43.11 (2020), zsaa097.

[9] Jiawei Bai et al. "An activity index for raw accelerometry data and its comparison with other activity metrics". In: *PloS one* 11.8 (2016), e0160644.

[10] Richard B Berry et al. *AASM scoring manual updates for 2017 (version 2.4)*. 2017.

[11] Richard B Berry et al. "The AASM manual for the scoring of sleep and associated events". In: *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine* 176.2012 (2012), p. 7.

[12] Roongroj Bhidayasiri and Claudia Trenkwalder. "Getting a good night sleep? The importance of recognizing and treating nocturnal hypokinesia in Parkinson's disease". In: *Parkinsonism & related disorders* 50 (2018), pp. 10–18.

[13] Sissel Bisgaard et al. "EEG recordings as a source for the detection of IRBD". In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2015, pp. 606–609.

[14] Rasch Björn and Born Jan. "About sleep's role in memory". In: *Physiol. Rev* 93 (2013), pp. 681–766.

[15] Mark S Blumberg and Alan M Plumeau. "A new view of "dream enactment" in REM sleep behavior disorder". In: *Sleep medicine reviews* 30 (2016), pp. 34–42.

[16] Matthias Boentert. "Sleep disturbances in patients with amyotrophic lateral sclerosis: current perspectives". In: *Nature and science of sleep* (2019), pp. 97–111.

[17] Pradeep C Bollu and Pradeep Sahota. "Sleep and Parkinson disease". In: *Missouri medicine* 114.5 (2017), p. 381.

[18] Mehdi AJ van den Bos et al. "Pathophysiology and diagnosis of ALS: insights from advances in neurophysiological techniques". In: *International journal of molecular sciences* 20.11 (2019), p. 2818.

[19] Heiko Braak et al. "Staging of brain pathology related to sporadic Parkinson's disease". In: *Neurobiology of aging* 24.2 (2003), pp. 197–211.

[20] Leo Breiman. *Classification and regression trees*. Routledge, 2017.

[21] Sara Brown. *Machine learning, explained*. https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained. [Online; Accessed 9-January-2024]. 2021.

[22] Ricardo Buettner, Annika Grimmeisen, and Anne Gotschlich. "High-performance diagnosis of sleep disorders: a novel, accurate and fast machine learning approach using electroencephalographic data". In: *Proceedings of the 53rd Hawaii International Conference on System Sciences*. 2020.

[23] Deepa Burman. "Sleep Disorders: Sleep-Related Breathing Disorders." In: *FP essentials* 460 (2017), pp. 11–21.

[24] Joseph W Burns et al. "EMG variance during polysomnography as an assessment for REM sleep behavior disorder". In: *Sleep* 30.12 (2007), pp. 1771–1778.

[25] Daniel J Buysse et al. "The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research". In: *Psychiatry research* 28.2 (1989), pp. 193–213.

[26] Matteo Cesari and Poul Jennum. "Selective polysomnographic findings in REM sleep behavior disorder (RBD) and Parkinson's disease". In: *Rapid-Eye-Movement Sleep Behavior Disorder* (2019), pp. 271–279.

[27] Matteo Cesari et al. "A data-driven system to identify REM sleep behavior disorder and to predict its progression from the prodromal stage in Parkinson's disease". In: *Sleep Medicine* 77 (2021), pp. 238–248.

[28] Matteo Cesari et al. "A new wavelet-based ECG delineator for the evaluation of the ventricular innervation". In: *IEEE journal of translational engineering in health and medicine* 5 (2017), pp. 1–15.

[29] Matteo Cesari et al. "Comparison of computerized methods for rapid eye movement sleep without atonia detection". In: *Sleep* 41.10 (2018), zsy133.

[30] Matteo Cesari et al. "External validation of a data-driven algorithm for muscular activity identification during sleep". In: *Journal of Sleep Research* 28.6 (2019), e12868.

[31] Matteo Cesari et al. "Flexor digitorum superficialis muscular activity is more reliable than mentalis muscular activity for rapid eye movement sleep without atonia quantification: a study of interrater reliability for artifact correction in the context of semiautomated scoring of rapid eye movement sleep without atonia". In: *Sleep* 44.9 (2021), zsab094.

[32] Matteo Cesari et al. "Validation of a new data-driven automated algorithm for muscular activity detection in REM sleep behavior disorder". In: *Journal of neuroscience methods* 312 (2019), pp. 53–64.

[33] Matteo Cesari et al. "Video-polysomnography procedures for diagnosis of rapid eye movement sleep behavior disorder (RBD) and the identification of its prodromal stages: guidelines from the International RBD Study Group". In: *Sleep* 45.3 (2022), zsab257.

[34] Matteo Cesari et al. "Video-polysomnography procedures for diagnosis of rapid eye movement sleep behavior disorder (RBD) and the identification of its prodromal stages: guidelines from the International RBD Study Group". In: *Sleep* 45.3 (2022), zsab257.

[35] Kallol Ray Chaudhuri et al. "The Parkinson's disease sleep scale: a new instrument for assessing sleep and nocturnal disability in Parkinson's disease". In: *Journal of Neurology, Neurosurgery & Psychiatry* 73.6 (2002), pp. 629–635.

[36] Adriano Chiò, V Silani, Italian ALS Study Group, et al. "Amyotrophic lateral sclerosis care in Italy: a nationwide study in neurological centers". In: *Journal of the neurological sciences* 191.1-2 (2001), pp. 145–150.

[37] Julie AE Christensen et al. "Data-driven modeling of sleep EEG and EOG reveals characteristics indicative of pre-Parkinson's and Parkinson's disease". In: *Journal of neuroscience methods* 235 (2014), pp. 262–276.

[38] Calogero Edoardo Cicero et al. "Prevalence of idiopathic REM behavior disorder: a systematic review and meta-analysis". In: *Sleep* 44.6 (2021), zsaa294.

[39] Chiara Cirelli and Giulio Tononi. "Is sleep essential?" In: *PLoS biology* 6.8 (2008), e216.

[40] Albert Cohen. "Ondelettes, analyses multi résolutions et traitement numérique du signal". PhD thesis. Paris 9, 1990.

[41] Jacob Cohen. "A coefficient of agreement for nominal scales". In: *Educational and psychological measurement* 20.1 (1960), pp. 37–46.

[42] Navin Cooray et al. "Detection of REM sleep behaviour disorder by automated polysomnography analysis". In: *Clinical Neurophysiology* 130.4 (2019), pp. 505–514.

[43] Navin Cooray et al. "Proof of concept: Screening for REM sleep behaviour disorder with a minimal set of sensors". In: *Clinical Neurophysiology* 132.4 (2021), pp. 904–913.

[44] Corinna Cortes and Vladimir Vapnik. "Support vector machine". In: *Machine learning* 20.3 (1995), pp. 273–297.

[45] Pedro Cruz-Vicente et al. "Recent Developments in New Therapeutic Agents against Alzheimer and Parkinson Diseases: In-Silico Approaches". In: *Molecules* 26.8 (2021).

[46] George DeMaagd and Ashok Philip. "Parkinson's disease and its management: part 1: disease entity, risk factors, pathophysiology, clinical presentation, and diagnosis". In: *Pharmacy and therapeutics* 40.8 (2015), p. 504.

[47] Stephanie Devuyst. *The DREAMS Databases and Assessment Algorithm*. May 2019. DOI: 10.5281/zenodo.2650142. URL: https://doi.org/10.5281/zenodo.2650142.

[48] Giovanni Dimauro et al. "Assessment of speech intelligibility in Parkinson's disease using a speech-to-text system". In: *IEEE Access* 5 (2017), pp. 22199–22208. ISSN: 21693536. DOI: 10.1109/ACCESS.2017.2762475.

[49] Stavros I Dimitriadis, Christos I Salis, and Dimitris Liparas. "An automatic sleep disorder detection based on EEG cross-frequency coupling and random forest model". In: *Journal of Neural Engineering* 18.4 (2021), p. 046064.

[50] Lei Ding et al. "A meta-analysis of the first-night effect in healthy individuals for the full age spectrum". In: *Sleep Medicine* 89 (2022), pp. 159–165.

[51] William N Dudley, Rita Wickham, and Nicholas Coombs. "An introduction to survival statistics: Kaplan-Meier analysis". In: *Journal of the advanced practitioner in oncology* 7.1 (2016), p. 91.

[52] Bradley Efron. "The efficiency of Cox's likelihood function for censored data". In: *Journal of the American statistical Association* 72.359 (1977), pp. 557–565.

[53] Emadeldeen Eldele et al. "An attention-based deep learning approach for sleep stage classification with single-channel EEG". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 29 (2021), pp. 809–818.

[54] Jacqueline A Fairley et al. "Wavelet analysis for detection of phasic electromyographic activity in sleep: influence of mother wavelet and dimensionality reduction". In: *Computers in biology and medicine* 48 (2014), pp. 77–84.

[55] Oluremi A Famodu et al. "Shortening of the Pittsburgh Sleep Quality Index survey using factor analysis". In: *Sleep disorders* 2018 (2018).

[56] Jiahao Fan et al. "EEG data augmentation: towards class imbalance problem in sleep staging tasks". In: *Journal of Neural Engineering* 17.5 (2020), p. 056017.

[57] Muniba Fayyaz et al. "The effect of physical activity in Parkinson's disease: a mini-review". In: *Cureus* 10.7 (2018).

[58] Eva L Feldman et al. "Amyotrophic lateral sclerosis". In: *The Lancet* 400.10360 (2022), pp. 1363–1380.

[59] Enrique Fernandez-Blanco, Daniel Rivero, and Alejandro Pazos. "Convolutional neural networks for sleep stage scoring on a two-channel EEG signal". In: *Soft Computing* 24 (2020), pp. 4067–4079.

[60] Raffaele Ferri et al. "A quantitative statistical analysis of the submentalis muscle EMG amplitude during sleep in normal controls and patients with REM sleep behavior disorder". In: *Journal of sleep research* 17.1 (2008), pp. 89–100.

[61] Raffaele Ferri et al. "Improved computation of the atonia index in normal controls and patients with REM sleep behavior disorder". In: *Sleep medicine* 11.9 (2010), pp. 947–949.

[62] Michela Figorilli et al. "Comparison between automatic and visual scorings of REM sleep without atonia for the diagnosis of REM sleep behavior disorder in Parkinson disease". In: *Sleep* 40.2 (2017), zsw060.

[63] Michela Figorilli et al. "Diagnosing REM sleep behavior disorder in Parkinson's disease without a gold standard: a latent-class model study". In: *Sleep* 43.7 (2020), zsz323.

[64] Pedro Fonseca et al. “Validation of photoplethysmography-based sleep staging compared with polysomnography in healthy middle-aged adults”. In: *Sleep* 40.7 (2017), zsx097.

[65] Rune Frandsen et al. “Analysis of automated quantification of motor activity in REM sleep behaviour disorder”. In: *Journal of sleep research* 24.5 (2015), pp. 583–590.

[66] Birgit Frauscher, Laura Ehrmann, and Birgit Högl. “Defining muscle activities for assessment of rapid eye movement sleep behavior disorder: from a qualitative to a quantitative diagnostic level”. In: *Sleep medicine* 14.8 (2013), pp. 729–733.

[67] Birgit Frauscher et al. “Normative EMG values during REM sleep for the diagnosis of REM sleep behavior disorder”. In: *Sleep* 35.6 (2012), pp. 835–847.

[68] Birgit Frauscher et al. “Quantification of electromyographic activity during REM sleep in multiple muscles in REM sleep behavior disorder”. In: *Sleep* 31.5 (2008), pp. 724–731.

[69] Birgit Frauscher et al. “Validation of an integrated software for the detection of rapid eye movement sleep behavior disorder”. In: *Sleep* 37.10 (2014), pp. 1663–1671.

[70] Ikbal Gazalba, Nurul Gayatri Indah Reza, et al. “Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification”. In: *2017 2nd international conferences on information technology, information systems and electrical engineering (ICITISEE)*. IEEE. 2017, pp. 294–298.

[71] Valentina Gnoni et al. “Hypothalamus and amyotrophic lateral sclerosis: potential implications in sleep disorders”. In: *Frontiers in Aging Neuroscience* 15 (2023), p. 1193483.

[72] Ary L Goldberger et al. “PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals”. In: *circulation* 101.23 (2000), e215–e220.

[73] Giulia Grande et al. “Multimorbidity burden and dementia risk in older adults: the role of inflammation and genetics”. In: *Alzheimer’s & Dementia* 17.5 (2021), pp. 768–776.

[74] Umesh C. Gupta and Subhas C. Gupta. “Optimizing Modifiable and Lifestyle-related Factors in the Prevention of Dementia Disorders with Special Reference to Alzheimer, Parkinson and Autism Diseases.” In: *Current Nutrition and Food Science* 16.6 (2020), pp. 900–911.

[75] Ingeborg H Hansen et al. "Detection of a sleep disorder predicting Parkinson's disease". In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2013, pp. 5793–5796.

[76] Orla Hardiman. "Management of respiratory symptoms in ALS". In: *Journal of neurology* 258.3 (2011), pp. 359–365.

[77] Frank E Harrell Jr, Kerry L Lee, and Daniel B Mark. "Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors". In: *Statistics in medicine* 15.4 (1996), pp. 361–387.

[78] Natalie L Hauglund, Chiara Pavan, and Maiken Nedergaard. "Cleaning the sleeping brain–the potential restorative function of the glymphatic system". In: *Current Opinion in Physiology* 15 (2020), pp. 1–6.

[79] Max Hirshkowitz and Amir Sharafkhaneh. "Chapter 2 - Normal sleep-recording and scoring techniques". In: *Sleep Disorders Part I*. Ed. by Pasquale Montagna and Sudhansu Chokroverty. Vol. 98. Handbook of Clinical Neurology. Elsevier, 2011, pp. 29–43. DOI: https://doi.org/10.1016/B978-0-444-52006-7.00002-2. URL: https://www.sciencedirect.com/science/article/pii/B9780444520067000022.

[80] Max Hirshkowitz and Amir Sharafkhaneh. "Chapter 47 - Polysomnography and home sleep test assessment methods in adults". In: *Atlas of Clinical Sleep Medicine (Third Edition)*. Ed. by Meir H. Kryger, Alon Y. Avidan, and Cathy Goldstein. Third Edition. Philadelphia: Elsevier, 2024, 438–461.e1. ISBN: 978-0-323-65403-6. DOI: https://doi.org/10.1016/B978-0-323-65403-6.00056-1.

[81] Birgit Högl and Ambra Stefani. "REM sleep behavior disorder (RBD): Update on diagnosis and treatment". In: *Somnologie* 21.Suppl 1 (2017), p. 1.

[82] Birgit Högl, Ambra Stefani, and Aleksandar Videnovic. "Idiopathic REM sleep behaviour disorder and neurodegeneration—an update". In: *Nature Reviews Neurology* 14.1 (2018), pp. 40–55.

[83] Stefany Cristina Claudino Idalino et al. "Association between sleep problems and multimorbidity patterns in older adults". In: *BMC Public Health* 23.1 (2023), p. 978.

[84] Syed Anas Imtiaz and Esther Rodriguez-Villegas. "A low computational cost algorithm for REM sleep detection using single channel EEG". In: *Annals of biomedical engineering* 42 (2014), pp. 2344–2359.

[85] Pankaj Jadhav et al. "Automatic sleep stage classification using time–frequency images of CWT and transfer learning using convolution neural network". In: *Biocybernetics and Biomedical Engineering* 40.1 (2020), pp. 494–504.

[86] *Jamovi - open statistical software for the desktop and cloud*. https://www.jamovi.org. Accessed: 2022-12-19.

[87] Poul Jennum, Julie AE Christensen, and Marielle Zoetmulder. "Neurophysiological basis of rapid eye movement sleep behavior disorder: informing future drug development". In: *Nature and Science of Sleep* (2016), pp. 107–120.

[88] James F Kaiser. "On a simple algorithm to calculate the'energy'of a signal". In: *International conference on acoustics, speech, and signal processing*. IEEE. 1990, pp. 381–384.

[89] Anthony Kales and Allan Rechtschaffen. "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects". In: (1968).

[90] Laura Kalevo et al. "Self-applied electrode set provides a clinically feasible solution enabling EEG recording in home sleep apnea testing". In: *IEEE Access* 10 (2022), pp. 60633–60642.

[91] A Karoui and R Vaillancourt. "Families of biorthogonal wavelets". In: *Computers & Mathematics with Applications* 28.4 (1994), pp. 25–39.

[92] Jacob Kempfner et al. "Early automatic detection of Parkinson's disease based on sleep recordings". In: *Journal of Clinical Neurophysiology* 31.5 (2014), pp. 409–415.

[93] Jacob Kempfner et al. "Rapid eye movement sleep behavior disorder as an outlier detection problem". In: *Journal of Clinical Neurophysiology* 31.1 (2014), pp. 86–93.

[94] Matthew C Kiernan et al. "Amyotrophic lateral sclerosis". In: *The lancet* 377.9769 (2011), pp. 942–955.

[95] Anne-Lene Kjældgaard, Katrine Pilely, et al. "Prediction of survival in amyotrophic lateral sclerosis: a nationwide, Danish cohort study". In: *BMC neurology* 21.1 (2021), pp. 1–8.

[96] L Klingelhoefer et al. "Night-time sleep in Parkinson's disease–the potential use of Parkinson's KinetiGraph: a prospective comparative study". In: *European journal of neurology* 23.8 (2016), pp. 1275–1288.

[97] Henri Korkalainen et al. "Self-applied home sleep recordings: the future of sleep medicine". In: *Sleep Medicine Clinics* 16.4 (2021), pp. 545–556.

[98] Dieter Kunz, Sophia Stotz, and Frederik Bes. "Treatment of isolated REM sleep behavior disorder using melatonin as a chronobiotic". In: *Journal of Pineal Research* 71.2 (2021), e12759.

[99] Tarek Lajnef et al. “Learning machines and sleeping brains: automatic sleep stage classification using decision-tree multi-class support vector machines”. In: *Journal of neuroscience methods* 250 (2015), pp. 94–105.

[100] Odile Lapierre and Jacques Montplaisir. “Polysomnographic features of REM sleep behavior disorder: development of a scoring method”. In: *Neurology* 42.7 (1992), pp. 1371–1371.

[101] Daniel J Levendowski et al. “Proof-of-concept for characterization of neurodegenerative disorders utilizing two non-REM sleep biomarkers”. In: *Frontiers in Neurology* 14 (2023), p. 1272369.

[102] Andrew SP Lim et al. “Sleep fragmentation and the risk of incident Alzheimer’s disease and cognitive decline in older persons”. In: *Sleep* 36.7 (2013), pp. 1027–1032.

[103] D Lo Coco, MONICA Puligheddu, et al. “REM sleep behavior disorder and periodic leg movements during sleep in ALS”. In: *Acta neurologica Scandinavica* 135.2 (2017), pp. 219–224.

[104] Maartje Louter, Walter Maetzler, Jos Prinzen, et al. “Accelerometer-based quantitative analysis of axial nocturnal movements differentiates patients with Parkinson’s disease, but not high-risk individuals, from controls”. In: *Journal of Neurology, Neurosurgery & Psychiatry* 86.1 (2015), pp. 32–37.

[105] Gianina Luca et al. “Age and gender variations of sleep in subjects without sleep disorders”. In: *Annals of medicine* 47.6 (2015), pp. 482–491.

[106] Diana Lucia et al. “Disorders of sleep and wakefulness in amyotrophic lateral sclerosis (ALS): a systematic review”. In: *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration* 22.3-4 (2021), pp. 161–169.

[107] Carlos Javier Madrid-Navarro et al. “Validation of a device for the ambulatory monitoring of sleep patterns: a pilot study on Parkinson’s disease”. In: *Frontiers in neurology* 10 (2019), p. 356.

[108] Jeanne E Maglione et al. “Actigraphy for the assessment of sleep measures in Parkinson’s disease”. In: *Sleep* 36.8 (2013), pp. 1209–1217.

[109] Henry B Mann and Donald R Whitney. “On a test of whether one of two random variables is stochastically larger than the other”. In: *The annals of mathematical statistics* (1947), pp. 50–60.

[110] Alessandro Manoni et al. “Long-term polygraphic monitoring through mems and charge transfer for low-power wearable applications”. In: *Sensors* 22.7 (2022), p. 2566.

[111] Giulia Masi et al. “Electrodermal Activity in the Evaluation of Engagement for Telemedicine Applications”. In: *2023 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. IEEE. 2023, pp. 130–135.

[112] Pegah Masrori and Philip Van Damme. “Amyotrophic lateral sclerosis: a clinical review”. In: *European journal of neurology* 27.10 (2020), pp. 1918–1929.

[113] Jessica Massicotte-Marquez et al. “Slow-wave sleep and delta power in rapid eye movement sleep behavior disorder”. In: *Annals of neurology* 57.2 (2005), pp. 277–282.

[114] Geert Mayer et al. “Quantification of tonic and phasic muscle activity in REM sleep behavior disorder”. In: *Journal of clinical neurophysiology* 25.1 (2008), pp. 48–55.

[115] Stuart J McCarter et al. “Autonomic dysfunction and phenoconversion in idiopathic REM sleep behavior disorder”. In: *Clinical autonomic research: official journal of the Clinical Autonomic Research Society* 30.3 (2020), p. 207.

[116] Stuart J McCarter et al. “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”. In: *Sleep* 37.10 (2014), pp. 1649–1662.

[117] Stuart J McCarter et al. “REM sleep muscle activity in idiopathic REM sleep behavior disorder predicts phenoconversion”. In: *Neurology* 93.12 (2019), e1171–e1179.

[118] Nicola Michielli, U Rajendra Acharya, and Filippo Molinari. “Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals”. In: *Computers in biology and medicine* 106 (2019), pp. 71–81.

[119] Mitchell G Miglis et al. “Biomarkers of conversion to α-synucleinopathy in isolated rapid-eye-movement sleep behaviour disorder”. In: *The Lancet Neurology* 20.8 (2021), pp. 671–684.

[120] Emmanuel Mignot. “Why we sleep: the temporal organization of recovery”. In: *PLoS biology* 6.4 (2008), e106.

[121] Anat Mirelman et al. “Tossing and turning in bed: nocturnal movements in Parkinson’s disease”. In: *Movement Disorders* 35.6 (2020), pp. 959–968.

[122] Shayan Motamedi-Fakhr et al. “Signal processing techniques applied to human sleep EEG signals—A review”. In: *Biomedical Signal Processing and Control* 10 (2014), pp. 21–33.

[123] Shayan Motamedi-Fakhr et al. “Signal processing techniques applied to human sleep EEG signals—A review”. In: *Biomedical Signal Processing and Control* 10 (2014), pp. 21–33.

[124] Marcus C Ng, Jin Jing, and M Brandon Westover. “A primer on EEG spectrograms”. In: *Journal of Clinical Neurophysiology* 39.3 (2022), pp. 177–183.

[125] Hong-Viet V Ngo, Jurgen Claassen, and Martin Dresler. “Sleep: slow wave activity predicts Amyloid-β accumulation”. In: *Current Biology* 30.22 (2020), R1371–R1373.

[126] Gabriella Olmo, F Laterza, and L Lo Presti. “Matched wavelet approach in stretching analysis of electrically evoked surface EMG signal”. In: *Signal Processing* 80.4 (2000), pp. 671–684.

[127] Gabriella Olmo and Francesco Laterza. “Matched wavelet analysis of single differential EMG signals”. In: *Wavelet Applications in Signal and Image Processing V*. Vol. 3169. SPIE. 1997, pp. 579–586.

[128] Aakash K Patel et al. “Physiology, sleep stages”. In: *StatPearls [Internet]*. StatPearls Publishing, 2022.

[129] Hanchuan Peng, Fuhui Long, and Chris Ding. “Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy”. In: *IEEE Transactions on pattern analysis and machine intelligence* 27.8 (2005), pp. 1226–1238.

[130] Steve Pincus. “Approximate entropy (ApEn) as a complexity measure”. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 5.1 (1995), pp. 110–117.

[131] Marios Politis et al. “Parkinson’s disease symptoms: the patient’s perspective”. In: *Movement Disorders* 25.11 (2010), pp. 1646–1651.

[132] John G Proakis. *Digital signal processing: principles, algorithms, and applications, 4/E*. Pearson Education India, 2007.

[133] Monica Puligheddu et al. “Isolated rapid eye movement sleep without atonia in amyotrophic lateral sclerosis”. In: *Sleep Medicine* 26 (2016), pp. 16–22.

[134] Monica Puligheddu et al. “Predictive risk factors of phenoconversion in idiopathic REM sleep behavior disorder: the Italian study “FARPRESTO””. In: *Neurological Sciences* 43.12 (2022), pp. 6919–6928.

[135] Monica Puligheddu et al. “Quantification of REM sleep without atonia: A review of study methods and meta-analysis of their performance for the diagnosis of RBD”. In: *Sleep Medicine Reviews* 68 (2023), p. 101745.

[136] Vitaliano Nicola Quaranta et al. “The prognostic role of obstructive sleep apnea at the onset of amyotrophic lateral sclerosis”. In: *Neurodegenerative Diseases* 17.1 (2016), pp. 14–21.

[137] Kannan Ramar et al. “Sleep is essential to health: an American Academy of Sleep Medicine position statement”. In: *Journal of Clinical Sleep Medicine* 17.10 (2021), pp. 2115–2119.

[138] Movement Disorder Society Task Force on Rating Scales for Parkinson’s Disease. “The unified Parkinson’s disease rating scale (UPDRS): status and recommendations”. In: *Movement Disorders* 18.7 (2003), pp. 738–750.

[139] Irene Rechichi et al. “Assessing rem sleep behaviour disorder: From machine learning classification to the definition of a continuous dissociation index”. In: *International Journal of Environmental Research and Public Health* 19.1 (2021), p. 248.

[140] Irene Rechichi et al. “Home Monitoring of Sleep Disturbances in Parkinson’s Disease: A Wearable Solution”. In: *2024 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. IEEE. 2024, pp. 106–111. DOI: 10.1109/PerComWorkshops59983.2024.10502893.

[141] Irene Rechichi et al. “Single-channel EEG classification of sleep stages based on REM microstructure”. In: *Healthcare Technology Letters* 8.3 (2021), p. 58.

[142] Irene Rechichi et al. “Single-Channel EEG Detection of REM Sleep Behaviour Disorder: The Influence of REM and Slow Wave Sleep”. In: *International Work-Conference on Bioinformatics and Biomedical Engineering*. Springer. 2022, pp. 381–394.

[143] Irene Rechichi et al. “Towards fully automatic quantification of REM sleep without atonia according to the Sleep Innsbruck Barcelona (SINBAR) scoring method”. In: *Sleep Medicine* 115 (2024). Abstracts from the 17th World Sleep Congress, p. 307. ISSN: 1389-9457.

[144] Jose Luis Rodríguez-Sotelo et al. “Automatic sleep stages classification using EEG entropy features and unsupervised pattern analysis techniques”. In: *Entropy* 16.12 (2014), pp. 6573–6589.

[145] Esther Rodriguez-Villegas et al. “A pilot study of a wearable apnoea detection device”. In: *BMJ open* 4.10 (2014), e005299.

[146] Annika Röthenbacher et al. “RBDtector: an open-source software to detect REM sleep without atonia according to visual scoring criteria”. In: *Scientific Reports* 12.1 (2022), p. 20886.

[147] Giulio Ruffini et al. “Deep learning with EEG spectrograms in rapid eye movement behavior disorder”. In: *Frontiers in neurology* 10 (2019), p. 806.

[148] David J Rushton et al. “Patients with ALS show highly correlated progression rates in left and right limb muscles”. In: *Neurology* 89.2 (2017), pp. 196–206.

[149] Michael J Sateia. "International classification of sleep disorders". In: *Chest* 146.5 (2014), pp. 1387–1394.

[150] Carolin Schaefer, Dieter Kunz, and Frederik Bes. "Melatonin effects in REM sleep behavior disorder associated with obstructive sleep apnea syndrome: a case series". In: *Current Alzheimer Research* 14.10 (2017), pp. 1084–1089.

[151] Matthias Schmid, Marvin N Wright, and Andreas Ziegler. "On the use of Harrell's C for clinical risk prediction via random survival forests". In: *Expert Systems with Applications* 63 (2016), pp. 450–459.

[152] Bernhard Schölkopf et al. "Estimating the support of a high-dimensional distribution". In: *Neural computation* 13.7 (2001), pp. 1443–1471.

[153] Simon J Schreiner et al. "Reduced Regional NREM Sleep Slow-Wave Activity Is Associated With Cognitive Impairment in Parkinson Disease". In: *Frontiers in Neurology* 12 (2021), p. 156.

[154] scikit-learn. *Feature Selection: Removing features with low variance*. https://scikit-learn.org/stable/modules/feature_selection.html#variance-threshold.

[155] Michael K Scullin and Chenlu Gao. "Dynamic contributions of slow wave sleep and REM sleep to cognitive longevity". In: *Current sleep medicine reports* 4 (2018), pp. 284–293.

[156] Chris Seiffert et al. "RUSBoost: Improving classification performance when training data is skewed". In: *2008 19th international conference on pattern recognition*. IEEE. 2008, pp. 1–4.

[157] Samuel Sanford Shapiro and Martin B Wilk. "An analysis of variance test for normality (complete samples)". In: *Biometrika* 52.3-4 (1965), pp. 591–611.

[158] Toshio Shimizu et al. "Complex fasciculation potentials and survival in amyotrophic lateral sclerosis". In: *Clinical Neurophysiology* 125.5 (2014), pp. 1059–1064.

[159] Péter Simor et al. "The microstructure of REM sleep: Why phasic and tonic?" In: *Sleep medicine reviews* 52 (2020), p. 101305.

[160] Jirada Sringean et al. "How well do Parkinson's disease patients turn in bed? Quantitative analysis of nocturnal hypokinesia using multisite wearable inertial sensors". In: *Parkinsonism & Related Disorders* 23 (2016), pp. 10–16.

[161] Jirada Sringean et al. "How well do Parkinson's disease patients turn in bed? Quantitative analysis of nocturnal hypokinesia using multisite wearable inertial sensors". In: *Parkinsonism & Related Disorders* 23 (2016), pp. 10–16.

[162] Ambra Stefani and Birgit Högl. "Sleep in Parkinson's disease". In: *Neuropsychopharmacology* 45.1 (2020), pp. 121–128.

[163] Jens B Stephansen et al. "Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy". In: *Nature communications* 9.1 (2018), p. 5229.

[164] Karin Stiasny-Kolster et al. "The REM sleep behavior disorder screening questionnaire—a new diagnostic instrument". In: *Movement disorders* 22.16 (2007), pp. 2386–2393.

[165] Kristína Šušmáková and Anna Krakovská. "Discrimination ability of individual measures used in sleep stages classification". In: *Artificial intelligence in medicine* 44.3 (2008), pp. 261–277.

[166] Mario Giovanni Terzano et al. "Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep." In: *Sleep medicine* 2.6 (2001), pp. 537–553.

[167] Kenji Uchino et al. "Impact of inability to turn in bed assessed by a wearable three-axis accelerometer on patients with Parkinson's disease". In: *PloS one* 12.11 (2017), e0187616.

[168] Ryan J Urbanowicz et al. "Relief-based feature selection: Introduction and review". In: *Journal of biomedical informatics* 85 (2018), pp. 189–203.

[169] Veria Vacchiano et al. "Prognostic value of EMG genioglossus involvement in amyotrophic lateral sclerosis". In: *Clinical Neurophysiology* 132.10 (2021), pp. 2416–2421.

[170] Jeroen Van Schependom and Miguel D'haeseleer. *Advances in neurodegenerative diseases*. 2023.

[171] Gabriel E Vázquez-Vélez and Huda Y Zoghbi. "Parkinson's disease genetics and pathophysiology". In: *Annual review of neuroscience* 44 (2021), pp. 87–108.

[172] Rick van Veen et al. "FDG-PET combined with learning vector quantization allows classification of neurodegenerative diseases and reveals the trajectory of idiopathic REM sleep behavior disorder". In: *Computer Methods and Programs in Biomedicine* 225 (2022), p. 107042.

[173] Jussi Virkkala et al. "Automatic detection of slow wave sleep using two channel electro-oculography". In: *Journal of neuroscience methods* 160.1 (2007), pp. 171–177.

[174] Matthew P Walker. "The role of sleep in cognition and emotion". In: *Annals of the New York Academy of Sciences* 1156.1 (2009), pp. 168–197.

[175] Lulu Xie et al. "Sleep drives metabolite clearance from the adult brain". In: *science* 342.6156 (2013), pp. 373–377.

[176] Fei Xue et al. “Analysis of nocturnal hypokinesia and sleep quality in Parkinson’s disease”. In: *Journal of Clinical Neuroscience* 54 (2018), pp. 96–101.

[177] Benjamin D Yetton et al. “Automatic detection of rapid eye movements (REMs): A machine learning approach”. In: *Journal of neuroscience methods* 259 (2016), pp. 72–82.

[178] Ozal Yildirim, Ulas Baran Baloglu, and U Rajendra Acharya. “A deep learning model for automated sleep stages classification using PSG signals”. In: *International journal of environmental research and public health* 16.4 (2019), p. 599.

[179] Magdy Younes, Jill Raneri, and Patrick Hanly. “Staging sleep in polysomnograms: analysis of inter-scorer variability”. In: *Journal of Clinical Sleep Medicine* 12.6 (2016), pp. 885–894.

[180] C Zenner et al. “P95. REM sleep without atonia and REM sleep behavioral disorder in patients with amytrophic lateral sclerosis”. In: *Clinical Neurophysiology* 126.8 (2015), e149.

[181] Hui Zhang et al. “Risk factors for phenoconversion in rapid eye movement sleep behavior disorder”. In: *Annals of neurology* 91.3 (2022), pp. 404–416.

[182] Ranqi Zhao, Yi Xia, and Qiuyang Wang. “Dual-modal and multi-scale deep neural networks for sleep staging using EEG and ECG signals”. In: *Biomedical Signal Processing and Control* 66 (2021), p. 102455.

[183] Tianqi Zhu, Wei Luo, and Feng Yu. “Convolution-and attention-based neural network for automated sleep stage classification”. In: *International Journal of Environmental Research and Public Health* 17.11 (2020), p. 4152.

[184] Lukáš Zoubek et al. “Feature selection for sleep/wake stages classification using data driven methods”. In: *Biomedical Signal Processing and Control* 2.3 (2007), pp. 171–179.

This Ph.D. thesis has been typeset by means of the TEX-system facilities. The typesetting engine was pdfLATEX. The document class was `toptesi`, by Claudio Beccari, with option `tipotesi=scudo`. This class is available in every up-to-date and complete TEX-system installation.