Machine Learning-Based Sleep Analysis Using a CW Radar

Bachelor Thesis in Medical Engineering

submitted
by

Paul Zwießele

born 04.07.1999 in Ravensburg

Written at

Machine Learning and Data Analytics Lab
Department Artificial Intelligence in Biomedical Engineering
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)

Advisors: Prof. Dr. Bjoern Eskofier, Daniel Krauss (M. Sc.), Luca Abel (M. Sc.)

Started: 28.08.2023

Finished: 28.01.2024

Ich versichere, dass ich die Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.
Die Richtlinien des Lehrstuhls für Bachelor- und Masterarbeiten habe ich gelesen und anerkannt, insbesondere die Regelung des Nutzungsrechts.

Erlangen, den 28.01.2024

Übersicht

Schlaf ist ein lebenswichtiger physiologischer Prozess, der für das allgemeine Wohlbefinden unerlässlich ist. Durch schlafmangel kann es zu einer Vielzahl von gesundheitlichen Folgen, sowie zu einer erhöhten Sterblichkeit kommen. Umgekehrt können verschiedene Krankheiten den Schlaf beeinträchtigen. Dies unterstreicht die Bedeutung einer sorgfältigen Schlafüberwachung für die frühzeitige Erkennung und langfristige Kontrolle von Erkrankungen.

Traditionelle Schlafüberwachungsmethoden, wie der derzeitige Goldstandard, die Polysomnographie PSG sind unkomfprtabel, arbeitsintensiv und kostspielig. Da viele Elektroden direkt am Patienten angebracht werden müssen. Zudem kann die klinische Umgebung, die Schlafqualität beeinträchtigen und somit die Dauer, potenzieller Studien reduzieren.

Um diese Nachteile zu umgehen, wird in dieser Studie ein auf Radar basierender, komfortabeler Ansatz zur Klassifizierung von Schlafstadien verfolgt. In dieser Arbeit wird sich ausschließlich auf die Bewegung während des Schlafes als Information zur Klassifizierung der Schlafphasen konzentriert. Dabei wird entweder zwischen Wach und Schlaf, oder Wach, non-rapid eye movement (NREM) und rapid eye movement (REM) unterschieden.

Für die Optimierung des Ansatzes werden folgende machine learning (ML) Algorithmen zur Klassifizierung eingesetzt: multi-layer perceptron (MLP), support vector machines (SVM), und extreme gradient boosting (XGB).

Der Bewegungserkennungsalgorithmus erzielte eine Leistung etwas unter der Erwartung und konnte Bewegungen mit einem F1-Wert von 66% erkennen. Die Algorithmen zur Klassifikation der Schlafphasen zeigten vielversprechende Ergebnisse mit F1-Werten von 81% für die binäre und 63% für die dreistufige Klassifikation.

Zusammenfassend wurde in dieser Arbeit eine robuste Bewegungserkennungsmethode entwickelt. Die extrahierten Bewegungsmerkmale wurden anschließend für die maschinellen Lern-basierte Schlafstadiumsbestimmung eingesetzt. Diese Klassifikation erreichte vielversprechende Leistungskennzahlen für die binäre und dreistufige Schlafstadiumsklassifikation.

Künftige Forschungsarbeiten könnten die Genauigkeit der Bewegungserkennung verbessern und zusätzliche Merkmale, wie z. B. Vitalparameter, in den Schlafklassifizierungsprozess integrieren. Durch die Erweiterung des Umfangs der Datenanalyse könnte eine umfassendere Bewertung der Schlafmuster angestrebt werden, was die Gesamtklassifizierungsleistung verbessern würde.

Abstract

Sleep is a vital physiological process that plays a crucial role in maintaining overall well-being. Its deprivation can have detrimental health consequences, including an increased risk of various diseases and a higher mortality rate. Conversely, various health conditions can also disrupt sleep patterns. This highlights the importance of careful sleep monitoring for both disease detection and monitoring.

Traditional sleep tracking methods, such as the current gold standard, polysomnography (PSG), are invasive, inconvenient, and costly. The use of multiple sensors in a clinical setting can disrupt sleep quality and prevent longitudinal data acquisition.

This research presents a non-invasive sleep stage classification approach using radar signals and machine learning (ML), offering an alternative to conventional PSG. For the classification of the sleep phases, this work relies solely on movement information extracted from the radar data. The work performed within this thesis involves data collection, preprocessing, movement detection, feature extraction, and sleep stage classification using three ML algorithms: multi-layer perceptron (MLP), support vector machines (SVM), and extreme gradient boosting (XGB).

The movement detection algorithm achieved a performance slightly lower than expected, demonstrating its ability to detect movements with an F1-score of 66%. The sleep stage classification algorithms exhibited promising results, with F1-scores of 81% for Wake/Sleep and 63% for Wake/non-rapid eye movement (NREM)/rapid eye movement (REM) classification, suggesting the potential of radar data for sleep staging.

In summary, this study crafted a robust movement detection pipeline. The extracted movement features were subsequently employed for ML-based sleep staging. This classification demonstrated promising performance, for binary and three-stage sleep phase classification, respectively. Future research will aim to enhance the performance of movement detection and incorporate additional physiological signals, such as cardio or respiratory information, into the sleep staging framework. By broadening the scope of data analysis, the goal could be, to achieve a more holistic assessment of sleep patterns, potentially improving overall classification performance.

Contents

Chapter 1

Introduction

Sleep is a major aspect of life, it constitutes approximately one-third of the human lifespan [K P19; Imt21]. It also plays a critical role in maintaining overall health and well-being. Inadequate sleep can lead to adverse health consequences, including cardiovascular diseases, cognitive impairment, and an increased mortality rate [Gra17].
Moreover, various illnesses can lead to sleep disruptions and other nocturnal symptoms. Parkinson's disease (PD), the second most common neurodegenerative condition, is among those diseases [Gar03]. Its symptoms include sleeplessness and sleep fragmentation of various etiologies such as REM sleep behavior disorder (RBD), restless-legs-syndrome (RLS), and periodic limb movement (PLMS) [Com07]. These sleep disturbances lead to insomnia and daytime sleepiness, of which over 90% of PD patients are impacted [Zha20].
Affecting approximately 50% of PD patients, RBD is one of the most common parasomnias [Zuz20]. RBD involves a loss of paralysis during the rapid eye movement (REM) sleep stage [Coo19]. This absence of paralysis can manifest itself in a variety of behaviors during REM sleep, ranging from mild muscular spasms to vigorous movements that may result in injury to the patient themself, or a bed partner [Zuz20; Mer12].
Sleep-related symptoms progressively intensify with the development of PD, establishing them as a clinical marker to evaluate the condition's progression [Zha20]. Furthermore, considering that sleep-related side effects may precede motor symptoms by several years [Ira06], the integration of sleep staging emerges as a potent tool for the nuanced tracking and early detection of the disease.
Sleep-related symptoms can vary significantly across different phases of sleep, necessitating careful differentiation between these stages for accurate diagnosis and effective treatment [Gar03]. According to Willemen et al. [Wil14], sleep stages can be divided into REM and non-rapid eye

movement (NREM), which alternate in sleep cycles of approximately 90 minutes. The REM sleep stage, is characterized by special brainwave characteristics, and swift eye movements, distinguishing it from the preceding NREM stages [Col08]. The NREM phase can be subdivided into three distinct categories, namely N1, N2, and N3 sleep phases. Within this classification, N1 and N2 correspond to stages of light sleep, while N3 marks a deep sleep stage. Each of these stages is characterized by specific vital signs and movement parameters [Wil14].

The prevailing gold standard for analyzing sleep and monitoring sleep stages is polysomnography (PSG) because of its detailed recording capabilities and reliability [Sad15; Goe21]. PSG uses various sensors to collect vital signs and movement data. Examples of these sensors include electroencephalography (EEG) to measure brain waves, electrooculography (EOG) to track eye movements, and electrocardiography (ECG) for heart rate detection [Sad15; Hon19]. PSG is discussed in further detail in Chapter 2.

The identification of sleep-related symptoms through sleep analysis has traditionally relied on techniques such as PSG or comparable wearable devices. However, this approach presents challenges. The multitude of sensors leads to a time-consuming setup and evaluation process. Furthermore, the artificial sleep setting can lead to irregular sleep patterns of study participants [Goe21]. Therefore, a promising alternative could be radar-based contactless sleep tracking [Rah15]. Unlike traditional PSG, radar-based sleep monitoring is non-invasive and can be employed for extended periods [Lee23]. Moreover, it offers cost savings over PSG, as it does not necessitate the presence of a trained technician for setup and supervision. Additionally, radar-based sleep monitoring holds the potential for home-based utilization, potentially alleviating sleep quality impairments stemming from the artificial sleep environment of a clinical setting [Lau20].

While radar sensors cannot detect the same physiological signals as the PSG this study attempts to approximate the results with secondary side effects. In this bachelor thesis, the primary emphasis is directed towards the quantification of movement. This involves using an integrated actigraph to measure the patient's movement and utilizing this data for the validation of the radar data [Leo20]. This thesis constructs a sleep analysis pipeline consisting of the implementation of pre-processing and feature extraction procedures for movement data obtained from a radar system. The extracted movement information is then compared to ground truth derived from PSG. In addition, various already established machine learning (ML) algorithms were used to apply sleep stage classification.

Chapter 2

Related Work

Many people suffer from some kind of sleep disorder or illness that causes sleep-related symptoms [Imt21]. These symptoms give detailed insights into the type and severity of the patient's condition. Therefore, one important diagnostic tool for the detection and monitoring of disorders is sleep analysis [Sad15].

The current diagnostic gold standard is PSG, a procedure that collects data from multiple sensors. These sensors measure brain waves via EEG, muscle activity through electromyography (EMG), eye movements by employing EOG, and the heart rate with ECG [Imt21]. In addition to those sensors, a typical PSG setup often includes photoplethysmography (PPG) to measure pulse and oxygen levels, an accelerometer for tracking movement and sleep position, respiratory inductance plethysmography (RIP) to monitor breathing patterns, a microphone for detecting sounds like snoring, and a thermal airflow sensor to measure nasal airflow [Imt21].

EEG, EOG, and EMG data are typically used to distinguish different stages of sleep. Eye movements and variable activity in frontal, central, and occipital brain regions are combined with chin EMG data to differentiate between sleep phases [Run19]. These stages consist of wake (W), REM, stage 1 (N1), stage 2 (N2), and stage 3 (N3) [Wil14]. The phases of sleep, namely N1, N2, N3, and REM, can be identified by observing abrupt changes in the EEG frequency. These shifts manifest in the theta range (5 – 12 Hz), the alpha range (8 – 13 Hz), or frequencies exceeding 16 Hz and with a duration over 3 s [Gir21]. In addition to a shift in the EEG signals, the REM sleep phase is marked with an at least 1 s long increase of the chin EMG signals [Run19]. These arousals are tracked overnight and then scored as an arousal index to measure the sleep disruption event.

This index provides a very detailed and reliable insight into the patient's sleep phases [Rah15].

While PSG remains the gold standard for sleep disorder diagnosis, it is not without limitations. The process is particularly characterized by a significant amount of manual work in the initial setup phase, which requires careful positioning and the correct application of electrodes to the patient [Gai22]. In a commercial sleep laboratory, the considerable effort needed for data analysis contributes to the high cost of recording [Rah15; Fab21]. However, the primary limitation of sleep recording using PSG is an intrusive and unnatural sleep environment [Goe21; Hon19]. This requires adjustment by the patient, often leading to the first-night-effect (FNE), resulting in sleep patterns that are not representative of the individual's usual sleep behaviors [Sad15]. The FNE is a phenomenon that can result in diminished sleep quality when individuals sleep in an unfamiliar environment for the first time [Agn66].
Patients with sleep disorders that cause daytime sleepiness, such as behavioral insomnia, schedule disorder, or regular night walks, can be challenging to record due to the need for planning and monitoring of PSG recordings [Sad15]. Taken together, these factors imply that widespread, longitudinal, and realistic sleep recording with PSG is highly impractical [Hon19].

These limitations are driving interest in alternative sleep analysis systems. Several techniques have been investigated to obtain the results of PSG, but with a less obtrusive setup that would allow prolonged measurements. These include Ballistocardiography, Doppler Laser, or Actigraphy [Goe21]. Many systems, primarily based on actigraphy, are commercially available and allow long-term sleep tracking. Due to their reduced sensor count and reliance solely on patient movement data, which can be very similar across different sleep phases, these sleep monitoring systems present limitations in accuracy and reliability relative to PSG [Rah15].

Radar technology provides a non-invasive solution for sleep analysis that is easy to use once installed [Tof20]. Its unobtrusive nature also reduces the mentioned FNE, leading to more realistic sleeping patterns in the resulting data. As outlined in the Chapter Radar Background 3, radar can detect both minor vital signs and major body movements.
Hong et al. [Hon18] used this capability and employed a 2.4 GHz radar system, located above the patient. In contrast, Kagawa et al. [Kag16] and Rahman et al. [Rah15] utilized a 24 GHz radar positioned under the mattress.
In the following section, this thesis will compare these three studies regarding their study structure, the pre-processing pipeline they used, the ML algorithms they employed and, the results that were achieved with this. This comparison is also depicted in the Table 2.1.

All three studies compared the results from the continuous wave (CW) Doppler radar system with traditional sleep staging by PSG. Hong et al. [Hon18] recorded 11 overnight PSG sessions, while Rahman et al. [Rah15] conducted multiple recordings of the same participants, resulting in a total of sixteen sleep sessions. Kagawa et al. [Kag16] recorded every participant one time, amounting to a total of 10 nights. The number of participants in the studies was 13, 8, and 10, respectively.

The complexity of data from radar-based sleep analysis systems necessitates careful processing. This involves two main steps: pre-processing the data for movement event detection, and incorporating artificial intelligence (AI).
Pre-processing includes filtering, noise suppression, synchronization, and handling data irregularities. This step is crucial to ensure that the subsequent sleep stage classification accurately detects the patterns of interest. Rahman et al. [Rah15] used a minimum-order Butterworth band-pass filter to isolate the breathing rate frequency between 9 and 20 breaths per minute. This filter applied stop-band frequencies at 0.1 Hz, 0.8 Hz, and pass-band frequencies between 0.3 Hz and 0.7 Hz. A similar filter was used to detect the patient's heart rate between 40 and 80 beats per minute. In this case, the stop-band frequencies were at 1 Hz and 3 Hz, respectively, and the pass-band was between 1.5 Hz and 2.5 Hz. To smooth the resulting signal, a moving average filter (MAF) with a window size of 30 s and a shift of 5 Hz was used to minimize heart and respiratory rate estimation error.

The study by Hong et al. [Hon18] also used band-pass filters with similar cutoff and pass-band frequencies to extract respiration and heart rate. Furthermore, the authors defined 11 different sleep features including a body movement feature, five respiration features, four heartbeat features, and a time feature. The feature set comprised various signal parameters including entropy and variance.
Rahman et al. [Rah15] extracted body movement using a low pass filter with a cutoff frequency of 3 Hz. This served the purpose of eliminating high-frequency periodic noise caused by environmental influences. Furthermore, the root mean square (RMS) energy and zero crossings of the baseband signal data were used as features for each 30 s frame.

A different approach was performed by Kagawa et al. [Kag16]. Their study used the standard deviation (SD) of the respiratory intervals to calculate a time domain index of fluctuation. The SD σ is defined as:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(R(i) - \tilde{R})^2} \tag{2.1}$$

where i is the interval, N the total number of intervals, $R(i)$ (i=1,2,..., N) is the radar output signal and $\tilde{R}$ is the average of the $R(i)$ data-set.

In contrast, Hong et al. [Hon18] assumed that movements during sleep primarily occur in the light sleep phases, approximately every five to 10 minutes. Therefore, they aggregated the effective energy of the signal over a 6-minute duration to ascertain whether the patient was in a light or deep sleep phase.
The used formula was defined as:

$$BI(k) = \sum_{i=1}^{6}(A_k(i) - A_k(i)_{\text{min}}) \tag{2.2}$$

where BI is the body movement index, k is the index, $A_k(i)_{\text{min}}$ the minimum value of the amplitude $A_k(i)$ in 60 second-intervals.

ML, a subset of AI, is particularly useful for efficiently handling large data sets. There are different ML algorithms designed for different data sets and specific needs [Shi18].
The mentioned studies utilized various ML algorithms for the final sleep stage classification. In the work of Hong et al. [Hon18], 18 different ML models were used, falling into four categories: support vector machines (SVM), decision tree, k-nearest neighbor classifier (KNN), and ensemble classifiers [Hon18]. Rahman et al. [Rah15] employed algorithms, such as naive bayes, logistic regression, SVM, and random forest, for their study [Rah15].
To improve sleep analysis using ML algorithms, Hong et al. [Hon18] demonstrated notable achievements, attaining an 89.6% accuracy in Sleep/Wake classification and an 85.8% accuracy in Wake/NREM/REM detection. The study by Rahman et al. [Rah15] also analyzed Sleep/Wake and Wake/NREM/REM detection. Their methodology incorporated data from an array of sensors, specifically those monitoring movement, breathing, and heart rate. Incorporating all essential physiological parameters and employing the most precise ML algorithm, random forest, they attained an F1-Score of 89.1% for the Wake/Sleep detection and an F1-score of 80.2% for the Wake/NREM/REM classification.

Additionally, they executed the scoring process utilizing solely the movement data. With this restriction, the random forest algorithm yielded F1-scores of 86.2% and 75.8%, respectively [Rah15].
The study by Kagawa et al. [Kag16], achieved an accuracy of 66.4% for classifying between wake and sleep, and 57.1% for classifying between wake, NREM, and REM sleep stages. For this Kagawa et al. [Kag16] used the body movement index and the heart rate, to assess sleep stages and compare the findings to the PSG classification. The lower metrics may have occurred because this study used canonical discriminant analysis (CDA) as the ML model [Kag16].
The resulting performance metrics can be seen in Table 2.1.

Research Team	**Sleep Stage Classification Performance**	**Features Used**	**Metric**
Hong et al. (2018)	89.6% (Sleep/Wake) 85.8% (Wake/NREM/REM)	Body Movement, Respiration, Heart Rate	Accuracy
Rahman et al. (2015)	89.1% (Wake/Sleep) 80.2% (Wake/NREM/REM)	Body Movement, Respiration, Heart Rate	F1-score
Rahman et al. (2015)	86.2% (Wake/Sleep) 75.8% (Wake/NREM/REM)	Body Movement	F1-score
Kagawa et al. (2016)	66.4% (Wake/Sleep) 57.1% (Wake/NREM/REM)	Body Movement, Respiration, Heart Rate	Accuracy

Table 2.1: Findings of the different studies.

Recent studies have shown that radar-based sleep analysis systems can monitor sleep stages effectively. These systems have the potential to overcome the limitations of traditional PSG, which is an intrusive and expensive procedure. Radar-based systems are non-invasive and can be used for long-term sleep monitoring.
Several studies have compared the accuracy of radar-based sleep analysis systems with PSG. The studies have yielded promising results, with radar systems achieving accuracy rates of up to 89.6% for Sleep/Wake classification and 85.8% for Wake/NREM/REM detection.
The success of radar-based sleep analysis systems lies in the use of ML algorithms to classify sleep stages. ML algorithms can learn to identify patterns in radar signals that are associated with different sleep stages.

Chapter 3

Radar Background

Contactless radar-based sleep monitoring offers the primary advantage of continuous monitoring over extended periods, without the drawbacks of the FNE that can lead to the generation of unrealistic sleep patterns [Rah15; Daw19].

The basic principle of the radar technology is the transmission of a radio signal. The signal then interacts with objects in its trajectory, resulting in reflection. Since electromagnetic waves travel at the speed of light in free space, it is possible to measure the time it takes the wave to travel the distance between the transmitter, the target, and the receiver. This time is also known as the roundtrip time of flight (RTOF) [Bor15]. One type of system is CW radar, also called Doppler radar, which uses a sinusoidal signal emitted by an oscillator. The reflected signal has the same shape as the transmitted signal, but with a delay and lower amplitude. These changes in the signal are caused by the RTOF and the reflectivity of the target. Calculations based on these changes enable the detection of very small distances [Wil17].
The majority of radar systems for biomedical sensing operate between $5.8\,\text{GHz}$ and $244\,\text{GHz}$ in the industrial, scientific, and medical (IMS) radar bands [Wil17; Li18; Alb23]. With these frequencies, it is possible to measure changes in distance in the µm range. In a $60\,\text{GHz}$ CW radar system, it is possible to detect movements as little as $7\,\mu\text{m}$ [Wil17]. This very precise measurement of the distance between the object and the radar system can be used to detect small movements caused by vital signs [Che22; Liu14]. These include unobtrusively measuring the patient's respiration, heart rate, and body movements. Studies using radar-based sleep analysis have shown that vital signs can be detected with high accuracy [Goe21].

In addition, it is possible to detect larger body movements, in the context of biomedical sensing, examples are the motion of different limbs or body parts. In a system using a $60\,\text{GHz}$ CW radar, the maximum unambiguous movement distance is limited to $2.5\,\text{mm}$.
To mitigate this problem, a CW radar system requires a sufficiently high sampling rate to ensure that the displacement between consecutive samples does not exceed the $2.5\,\text{mm}$ threshold [Kao13]. In radar systems, the modulation of the sinusoidal wave signal conventionally involves utilizing both in-phase (I) and quadrature (Q) components [Bor15]. However, significant movements can be observed in both components. Consequently, a detailed exploration of this modulation intricacy was deemed unnecessary for the scope of this thesis.

Radar-based sleep monitoring systems are emerging as a non-invasive and contactless alternative to traditional PSG, offering several advantages.
One type of these systems is CW radar technology, which can track sleep patterns by emitting a sinusoidal signal that reflects off the body. This reflected signal is then used to measure the distance between the body and the radar system, allowing detection of even subtle movements as small as $7\,\mu\text{m}$, in a $60\,\text{GHz}$ system. This sensitivity makes CW radar systems well-suited to monitoring vital signs such as respiration and heart rate as well as body movements.

Chapter 4

Methodology

Advancements in computational processing and ML have enabled the development of new systems for sleep recording. This research utilizes a CW radar system to analyze sleep patterns. The primary objective is to extract movement data from the processed radar signals and approximate these findings to the data obtained from a SOMNOMEDICS PSG system.
This research is conducted within the framework of PD early detection through sleep analysis. As part of this research, data acquisition was undertaken, particularly capturing sleep parameters of a control cohort in a sleep laboratory.
The methodology used in this research involves pre-processing, feature extraction, and the incorporation of ML algorithms. These were developed by the machine learning and data analytics (MaD) Lab under Daniel Krauss, the supervisor of this thesis.
The subsequent sections will provide a detailed description of the pre-processing and feature extraction techniques applied to the radar data. These sections will also discuss the challenges encountered during this process and the strategies implemented to address these challenges. This research aims to establish a reliable, non-invasive method for sleep analysis using movement information extracted from a CW radar system.

4.1 Data Collection

4.1.1 Recording Systems

To verify the accuracy of contactless sleep tracking, using radar sensing technology, sleep data were simultaneously recorded using two distinct recording systems: a radar-based system and a PSG recording device.
A 60 GHz CW radar system was used for the non-invasive recording. It consisted of four radar nodes placed in a horizontal line at the level of the torso underneath the mattress. The radar hardware as well as the software, consisting of the radar server and graphical user interface (GUI) was provided by the research center "Empatho-Kinaesthetic Sensor Technology Sensor Techniques and Data Analysis Methods for Empatho-Kinaesthetic Modeling and Condition monitoring (EmpKinS)".
A SOMNOMEDICS HD eco PSG system, served as the ground truth for evaluating the calculated movement and sleep stage information in this thesis.
The data in the following Table 4.1 were collected by the PSG.

Data Type	Description
EEG	Measures brain activity
EOG	Measures eye movements
ECG	Measures heart activity
EMG	Measures muscle activity
External Activity	Measures general body activity
Body Position	Monitors the position of the body during sleep
SP02	Measures oxygen saturation in the blood
Pulse	Measures the heart rate
Microphone	Records sounds, such as snoring
Thermistor	Measures temperature
Pneumotachograph	Measures airflow
RIP-Effort Abdomen	Measures respiratory effort in the abdomen
RIP-Effort Thorax	Measures respiratory effort in the thorax
Night vision camera	Confirming movement and identifying night onset and offset

Table 4.1: Data recorded by the PSG.

To enable synchronization between the systems, an m-sequence was recorded on both of them. This was also the only cable from the PSG that was connected to the stationary setup.

4.1.2 Prior to the Recording

This study involved the acquisition of data from a healthy control cohort comprising 37 individuals, which will serve as a baseline for the subsequent sleep monitoring of PD patients. Clinical recordings of PD patients will be conducted after the completion of this thesis and are not going to be included in the evaluation.

The starting time was determined based on the participant's normal sleep schedule, with a 90-minute preparation window. For example, if a participant typically went to bed at 11:00 PM, they were invited to the sleep lab at 9:30 PM. The participant was requested to provide their date of birth, height, and weight, and to fill out an electronic questionnaire.
Upon arrival, participants are required to complete a consent form indicating that the resulting data will be used for the study and shared with medical professionals for manual sleep stage classification. Additionally, participants were informed and provided consent to be recorded with the night vision camera. The camera was utilized to identify the start and end of the sleep phase, as well as to verify movements during sleep.

Questionnaire

The questionnaire[1] consisted of 58 questions, divided into the Pittsburgh sleep quality index (PSQI) [Buy89], snoring, tiredness, observed apnea, pressure, BMI, age, neck circumference, gender (STOP-Bang) Questionnaire [Chu08], RBD Screening Questionnaire [Sti07], and the SF-12 Health Survey [Jen97]. The questionnaire focused on the participant's past sleeping experiences, habits, and sleep settings. It also assessed specific sleep incidents such as vivid dreams or sudden limb movements. In addition, the study aimed to evaluate the participant's sleeping experience by asking about their sporting activities, daily life, and health. The questions ranged from simple yes/no answers, statements of the usual sleep duration, and grading of health from excellent to poor.
An example of the questionnaire can be seen in Appendix A.

[1] The full questionnaire is included in the Appendix at the end of the thesis.

4.1.3 Sleep Recording

The PSG system was set up using the participant's provided birthday, height, and weight data. The PSG and night vision camera recordings were initiated, and the layout of the PSG sensors was selected.
The sensors were attached in the following sequence: EMG, ECG, RIP thorax and abdomen, EEG, EOG, chin EMG, microphone, nose capsule for the Pneumotachograph, and the SP02 sensor in the form of a finger clamp. The camera was positioned on the leg side of the sleep laboratory to capture a comprehensive view of the entire bed.
A depiction of the sleep laboratory is provided in Figure 4.1.

Figure 4.1: The sleep laboratory setup used in this thesis.

4.1.4 Dataset

This section provides a thorough description of the data utilized in the study, including its format, and cleaning procedure. The dataset primarily consists of the radar and the PSG recordings, which are further supplemented with the night vision camera footage.

The radar data were archived in h5 file format and transferred using an exporter supplied by the EmpKinS research center. The PSG data were disseminated into multiple files, and stored as both a european data format (EDF) file and raw data. Additionally, the footage from the night vision camera was archived in 20 minute-intervals.

The PSG data extracted from the EDF file was organized into 57 distinct DataStreams, each corresponding to a specific sensor modality. For example, the Actigraph signal was stored under the label “Activity”, while limb movement data were compartmentalized into separate channels for the arms and legs. In contrast, the radar data were divided into the signals of the four radar nodes. These channels each contained both the I and Q parts of the radar signal. However, since the movement data is represented in both components, as can be seen in Chapter Radar Background 3, this thesis adopted the I channel as the standard data source for movement analysis.

Data Cleaning

Unfortunately, some data were lost, due to different reasons:

- **Missing Radar Signal:** In three recordings, the radar signal was either partially or completely lost.
- **Missing EEG Data:** In four recordings the EEG signal was completely lost or insufficient, in one additional recording the signal from multiple EEG sensors were of bad quality.
- **Missing Signals of other Sensors:** In three recordings the signals of other sensors like the EOG were lost.

While recordings with insufficient data were present, not all of them were excluded from the evaluation. Since the estimation of movement does not necessitate the utilization of all recorded PSG signals, a loss of some data were deemed tolerable. Nevertheless, insufficient radar or EEG data did prompt the exclusion of five recordings. Additionally, excessive signal loss compromised the automated sleep scoring of the PSG recordings. This shortcoming hampered the accuracy of the sleep scoring ground truth, a critical aspect for this thesis that relied on automated sleep analysis.
As a result of signal loss, five recordings were deemed unsuitable for inclusion in this thesis.
Tables 4.2 and 4.3 provide a comprehensive overview of the demographic characteristics of the recorded cohort.

Metric	**Mean**	**Range**	**Unit**
Age	39.38 ± 16.91	18 - 77	years
Weight	71.60 ± 14.61	48 - 105	kg
Height	172.76 ± 9.31	152 - 197	cm

Table 4.2: Descriptive statistics of age, weight, and height, of the recorded cohort.

Gender	**Count**
Male	14
Female	23

Table 4.3: Gender distribution of all recorded participants.

4.2 Pre-Processing

This section describes the data processing pipeline from the raw radar data to the calculation of movement events.

4.2.1 Data Import

As stated in Chapter 4.1.4 the radar and PSG data were stored in different formats. To import these files, a dataset “D04MainStudy”, provided by the supervisor of this thesis, Daniel Krauss was utilized. This dataset combined the data from the radar and PSG systems, providing easy access to various information channels in a pandas DataFrame[2] format. The index of this dataset could be selected in the form of the number of samples, time, and the date and local time as “local_datetime”. The radar data, PSG data, and if needed, EMG data of the extremities were imported. The activity channel from the PSG data were primarily used as a reference for the calculated movement features in this thesis. Additionally, the sync channel of the PSG was imported to synchronize the radar and movement data. The EMG data were used as a reference for the calculated movement localization, as it differentiates which limb was in motion.

4.2.2 Data Synchronization

As explained in the Radar Background Chapter 3, a high sampling rate is required for the CW radar system to accurately record large movements. In this thesis, a radar sampling rate of 1953.125 Hz was used. In contrast, the PSG does not require such a high sampling rate to record vital signs. To enable a comparative use of the data from both systems, the signals must be synchronized. First, this thesis employed a method that finds a common start and end timestamp for both signals, ensuring that both signals have valid values throughout the entire duration. The “local_ datetime” index was introduced to solve this problem. After determining the timestamps, the signals were cut accordingly.
The method for synchronizing the radar and PSG data were imported from the python package empkins-io[3], which is used to load and convert data from EmpKinS sensors. To ensure synchronization and improve performance, the sampling rate of both signals was increased to a standardized value of 2 kHz.

[2] https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.html
[3] This method was provided by the EmpKinS project of the Friedrich-Alexander-Universität (FAU) Erlangen-Nürnberg.

For synchronization, a technique known as m-sequence synchronization was employed. An m-sequence, also known as a maximum-length sequence, is a type of pseudo-random binary sequence that is useful for aligning and synchronizing data. Its unique properties, such as maximum length and balanced spectral characteristics, make it ideal for accurately aligning and synchronizing datasets. The m-sequence serves as a reference signal for cross-correlation-based synchronization, contributing to precise temporal alignment between different datasets.

At the time of completion of this thesis, the synchronization algorithm employed in the empkins-io package was not fully functional, leading to a misalignment of the radar and PSG signals. While precise synchronization is crucial for detecting vital signs like heart rate and heart rate variability (HRV), which are characterized by low amplitude and high-frequency fluctuations, the lack of perfect alignment is less significant for movement detection. This is due to the prominence of amplitude and duration in movement signals, which can be effectively identified even with minor timing discrepancies.
The segmentation of the signal into 30-second intervals further supports this notion, as the primary goal of movement detection is to identify the occurrence of movement events, rather than pinpointing their exact timing. This approach, while less accurate for detecting subtle movements, is sufficient for assessing sleep patterns and identifying periods of increased movement activity. Considering the focus of this thesis on large body movements, precise synchronization was not critical, as aligning the "local_datetime" index was sufficient for determining whether a movement occurred.

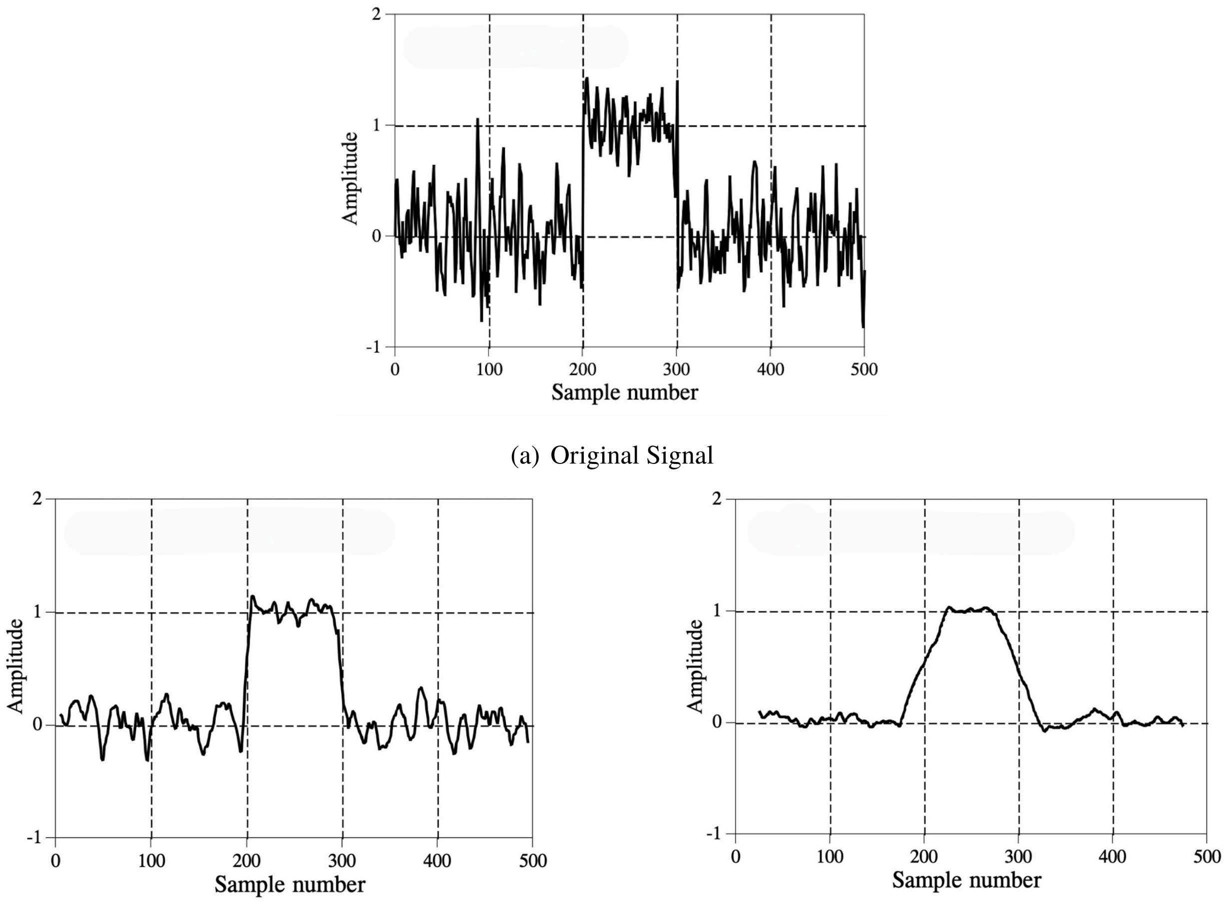
4.2.3 Moving Average Filter

A MAF is a signal smoothing technique that averages a set number of data points, known as the window size, at each point in time. This technique reduces noise and reveals underlying trends in the data. The MAF can be described by:

$$y[k] = \frac{1}{M} \sum_{i=-\lceil \frac{M-1}{2} \rceil}^{\lceil \frac{M-1}{2} \rceil} x[k+i] \tag{4.1}$$

$y[k]$ represents the output signal at the time index k and $x[t]$ is the input signal at the same time interval. M is the number of samples in the moving average window. The summation is performed over the range of i samples.

This thesis employs a centered MAF with a window size of 19 001 samples. The filter utilizes 8 500 samples before and after the current time index, resulting in an average of approximately $8.5\,\mathrm{s}$ of data if synchronized with a sampling rate of $2\,\mathrm{kHz}$. The applied MAF can be seen in Figure 4.3. The window size was determined via grid-search and selected based on its ability to maximize the resulting F1-score.
The centered MAF exhibits minimal phase shift in the transformed signal relative to the original. Nevertheless, the centered MAF incurs data loss during the sampling of the first and last 8 500 signal samples. To resolve this limitation, half the window size was padded to both ends of the signal. The value of the added samples was determined by averaging the corresponding existing data points.



(a) Original Signal

(b) 11 Point Moving Average

(c) 51 Point Moving Average

Figure 4.2: This graphic illustrates the functional operation of a MAF on a prototypical signal. 4.2(a) shows the original signal, while 4.2(b) and 4.2(c) show the results of different filter sizes, modified from [Smi99].

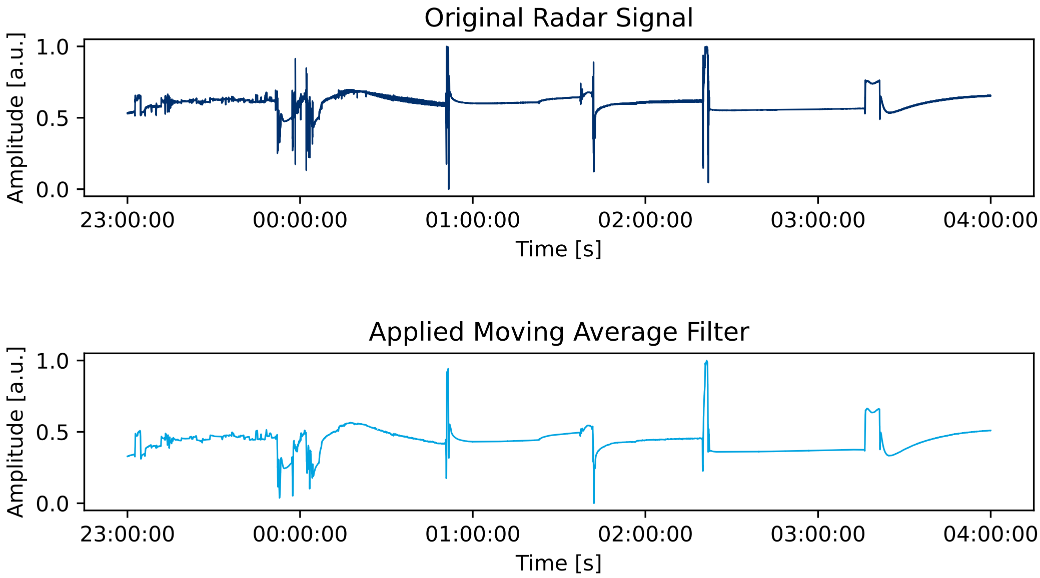


Figure 4.3: This figure shows the applied MAF with a window size of 19 001 samples on the raw radar signal.

4.2.4 First-Order Derivative

The amplitude and RTOF of a radar signal, can provide information about the person's position and distance relative to the system itself. For motion detection, changes in radar signal amplitude are crucial indicators of movement. The rate of change of the radar signal can be determined using the first-order derivative (FOD).

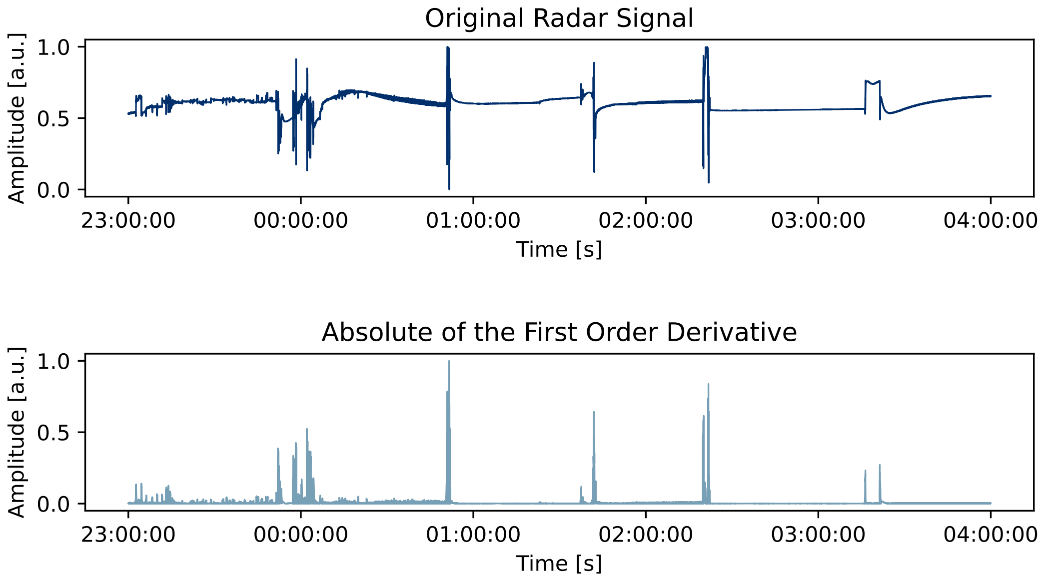


Figure 4.4: The Figure illustrates the application of a FOD algorithm to a radar signal that has undergone a MAF preprocessing step. For easier visualization, the absolute value of the derived FOD, was taken.

4.2.5 Data Adjustment

To adjust all data on the same axis, the index had to be synchronized as described in the Chapter Data Synchronization 4.2.2. The amplitude of raw radar data were normalized to the range of 1 by dividing the entire signal by the maximum value. This way, the ratios between the individual data points remained the same, and only the amplitude was reduced. For easier evaluation of the derivative data, the absolute value of the signal was taken according to the following function:

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases} \tag{4.2}$$

This allows for easier detection of movement and differentiation between signal amplitudes.

4.2.6 Energy calculation

Although the FOD provides insight into the timing of a movement, it does not indicate its duration. To solve this issue a method was implemented, that used the time-based index of the data to group the signal into 30-second intervals using the pandas floor[4] method. To quantify the overall movement activity within each 30-second interval, the average of the derivative signal over that period was calculated. This approach enabled the assessment of the total magnitude of movement within each time segment. The 30-second intervals were chosen because the ML algorithms also performed in intervals of this duration.

The output of this function can be represented as:

$$y[k] = \frac{1}{N} \sum_{i=1}^{N} x[k_i] \tag{4.3}$$

where k is a 30-second interval, k_i the frames in the interval and N is the total number of data points in an interval.

[4]https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.Series.dt.floor.html

4.2.7 Movement Differentiation

Upon analyzing the movement's quantity and length, its intensity and position remain unclear. This thesis differentiates the amplitudes of the four radar nodes to determine the movement's location. If one region displays a particularly high amplitude, the movement is likely concentrated around that radar node. Conversely, if all locations exhibit a high amplitude, it suggests an overall large motion.
For further explanation of the different thresholds used to localize and detect large movements, it is axiomatic that the largest movement in one recording is set to 1, and all thresholds are fractions of this amplitude.
To distinguish between the different radar nodes, the data from all nodes were consolidated into a single DataFrame and categorized as "left", "mid_left", "mid_right", and "right". Additionally, an extra variable named "no_movement" with a value of 0.001 was included. The value of the included "no_movement" label was not selected based on any specific rationale but rather serves as a benchmark for subsequent identification of "no movement" periods.
The radar nodes were arranged from left to right in ascending order, with node 1 on the left and node 4 on the right, of the top view of the mattress.
To eliminate minor movements from the different DataStreams, all data points below a threshold of 0.2 were removed. The area with the largest movement was calculated using the pandas method pd.idxmax[5]. This method returns the column with the largest value for each row. If no movement was detected in the DataStreams of the different radar nodes, the "no_ movement" variable was selected as the index with the maximum value.
To differentiate between general movements with a local maximum, the large movements were subtracted from the location data.

To calculate large movements, all radar nodes had to measure substantial movement. This thesis distinguishes between large and very large movements using different thresholds, which can be seen in Table 4.4. It is assumed that the participant sleeps approximately in the middle of the bed, which leads to artificially increased amplitudes of movement in radar nodes 2 and 3. To counteract this, different thresholds were used for the periphery and center nodes.

[5]https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.idxmax.html

Movement Type	Periphery Value	Center Value
Large	0.30	0.50
Very Large	0.60	0.75

Table 4.4: Thresholds for detecting large and very large movements.

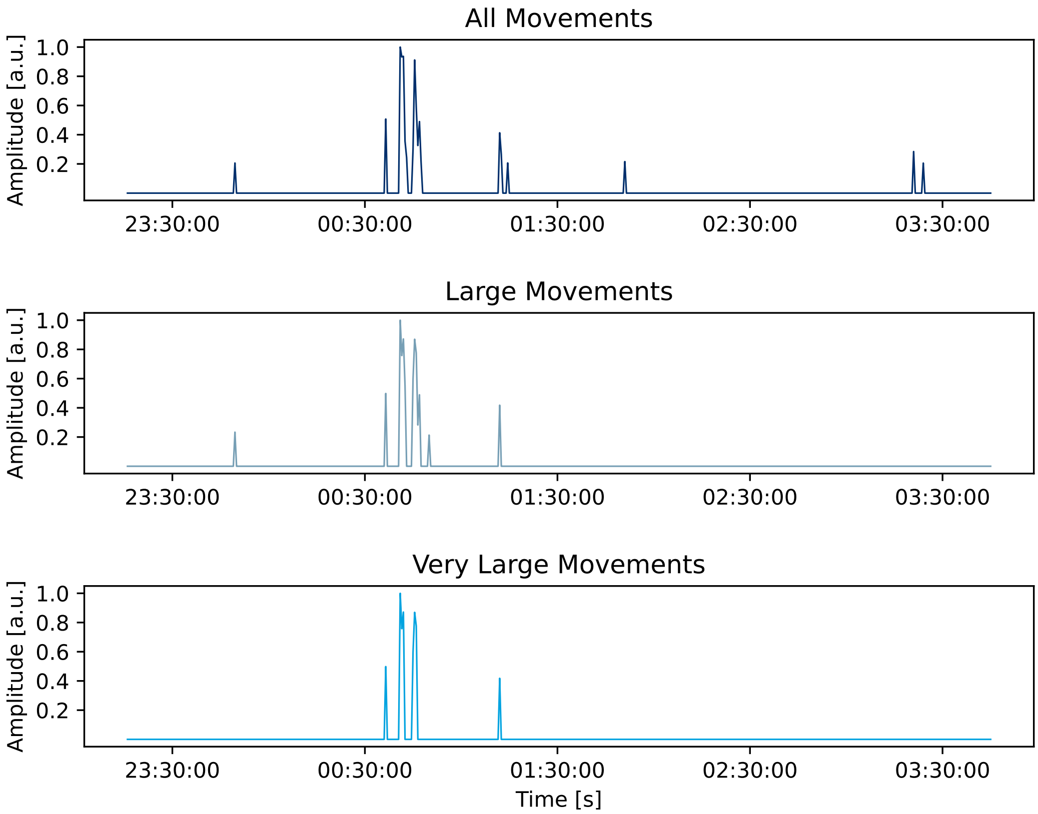


Figure 4.5: Example of a recording of four hours with the differentiation between large and very large movements.

4.2.8 Thresholds

Before metric calculation, a threshold was applied to both the PSG movement data and the processed radar data. In the PSG signal, all data points smaller than 0.05 were set to 0.0 to reduce noise. The radar data threshold and MAF window size were progressively evaluated to maximize the F1-score of the signal in relation to the PSG. Varied window sizes demonstrated a positive impact on precision and a corresponding negative impact on recall. Smaller window sizes revealed more noise, leading to false positive (FP) datapoints.

Conversely, excessively wide window sizes could inadvertently suppress and overlook smaller movements, resulting in false negative (FN) samples. A balanced approach between these two phenomena was important.
For this, the different window sizes between 10 001 and 20 001 have been tested for the maximization of the F1-score and a balance between precision and recall. Additionally, a review of the effectiveness of thresholds ranging from 0.01 to 0.34 has been conducted.
This optimization process yielded optimal performance with a threshold of 0.2 and a window size of 19 001 corresponding to approximately 8.5 s, with a sampling rate of 2 kHz.

4.3 Feature Extraction

Feature extraction plays a pivotal role in ML algorithms by transforming raw data into meaningful representations that serve as inputs for ML models. The selection of relevant and high-quality features is essential as it directly impacts the performance of these models in learning and making accurate predictions. The extracted features are listed in Table 4.5.

Feature Name	**Description**
Mean	Calculates the average value of the radar movement data
Median	Finds the median value of the radar movement data
SD	Standard deviation of the radar movement data
Maximum	Finds the highest value in the radar movement data
Minimum	Finds the lowest value in the radar movement data
Variance	Measures how spread out the radar movement data is
Sum	Adds up all movement events of the radar movement data
Skewness	Measures the skewness of the radar movement data
MLM	Calculates the mean of large movements (MLM)
MVLM	Calculates the mean of very large movements (MVLM)
Kurtosis	Measures the kurtosis of the radar movement data
25th Percentile	Only considers the values below 25% of the complete radar movement data
50th Percentile	Only considers the values below 50% of the complete radar movement data
75th Percentile	Only considers the values below 75% of the complete radar movement data
Autocorrelation	Measures the similarity between observations as a function of the time lag between them

Table 4.5: Feature set extracted from the radar data.

Time-series features were extracted by utilizing centered moving windows with window sizes varying from $n = 2, \ldots, 19$, encompassing a progressive expansion of temporal intervals within the feature extraction process. For each window size, n represents the number of 30-second intervals. The total number of extracted features amounted to 272. The decision to initiate the window size at 2 reflects the redundancy of using a window size of 1, which merely replicates the original signal without contributing any additional temporal information.
The original signal was incorporated as an additional feature and appended to the feature DataFrame. The ground truth movement data were then trimmed to align with the features extracted from the radar data in terms of length and corresponding "local_datetime" index. Finally, both the extracted features and ground truth data were stored in a comma-separated values (CSV) file for utilization with the use of the ML algorithms.

4.4 Machine Learning Algorithms

Inside the ML algorithm scripts, the features in combination with the ground truth were utilized for training and subsequently testing of the algorithms.
The study compared the performance of three ML algorithms in classifying sleep stages using radar data. The algorithms were modified to handle the new modality called "radar_movement" and to interface with the dataset.

- **Support Vector Machine (SVM):** SVMs are supervised ML algorithms, primarily used for classification and regression tasks. By maximizing the distance between two classes, SVMs identify the optimal hyperplane separating the data. This hyperplane, a n-dimensional plane, acts as a decision boundary that separates the data into two classes, such that the maximum possible distance is maintained between the instances on either side.
- **Multi Layer Perceptron (MLP):** MLPs are artificial neural networks that consist of interconnected layers of nodes, making them versatile tools for tackling a wide range of tasks. The input layer receives raw data, while hidden layers progressively refine its representation, extracting higher-level features and patterns. The output layer generates the final prediction or classification based on the acquired knowledge.

- **Extreme Gradient Boosting (XGB):** XGB, is an ensemble ML algorithm, that leverages gradient boosting to enhance decision tree performance. Gradient boosting, an iterative approach, constructs a superior model by building upon the predictions of preceding models.

Minor modifications were necessary to make the existing algorithms compatible with the collected data from this study. The main change was the incorporation of a new modality called "radar_movement". Another modification involved adding an option to select between the different datasets and to fit the new dataset "D04MainStudy" to the algorithms. To facilitate interaction with ML algorithms, two properties, "ground_truth", and "features", were added to the dataset structure. Additionally, three interface methods were implemented to provide a seamless link between the algorithms and the data.

The ground truth data were then modified to align with the intended sleep stage classification categories: Wake/Sleep or Wake/NREM/REM.
For the Wake/Sleep classification, the labels for N1, N2, N3, and REM were all replaced with Sleep. The Wake/NREM/REM classification involved merging the N1, N2, and N3 phases into a single NREM category. These modified classifications were stored in separate columns of the "ground_truth" data structure, labeled "binary" and "3stage", respectively.

4.5 Summary

This Chapter 4 explained the systematic approach to sleep analysis using CW radar technology. The methodology includes pre-processing, feature extraction, and the application of ML algorithms. Pre-processing consisted of a centered MAF that smooths the radar data, the FOD detects instances of amplitude variation and pre-determined thresholds mitigate noise and suppress unwanted artifacts. Movement differentiation involves analyzing the amplitudes of individual radar nodes to identify movement patterns. Afterwards, feature extraction involves selecting and calculating a set of time-series features from the radar data. Furthermore, ML algorithms, such as SVM, MLP, and XGB, were employed for the classification of the sleep stages.

Chapter 5

Evaluation

5.1 Performance Metrics

To assess the effectiveness of the processing pipeline and the outcomes obtained in this research, a diverse range of performance metrics were employed. These metrics encompass a broad spectrum of aspects pertaining to the resulting data. The movement detection process was evaluated using precision, recall, and the F1-score, while the performance of the ML algorithms was evaluated using all metrics listed in Table 5.1.

Metrics	**Definition**	**Formula**
Accuracy	The proportion of correctly classified instances	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	The proportion of positive predictions that are actually positive	$\frac{TP}{TP+FP}$
Recall	The proportion of positive instances that are correctly classified as positive	$\frac{TP}{TP+FN}$
F1-Score	The harmonic mean of precision and recall	$2 \cdot \frac{\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}}$
Kappa	A measure of agreement between predicted and ground truth labels, considering chance	$\frac{(\text{Observed Agreement}-\text{Expected Agreement})}{1-\text{Expected Agreement}}$
Specificity	The proportion of negative instances that are correctly classified as negative	$\frac{TN}{TN+FP}$
MCC	Matthews correlation coefficient	$\frac{(TP\cdot TN)-(FP\cdot FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$

Table 5.1: Performance metrics for evaluating the effectiveness of the movement detection pipeline and ML algorithms.

5.2 Evaluation of the Study Dataset

First, the data were recorded, and then the performance metrics, which can be seen in Table 5.1 were calculated.
This section will analyze the dataset used in this study, assessing its strengths and limitations.

- **Data collection and processing:** This thesis incorporates the recording of 37 participants in overnight sessions. All recordings were part of the reference group to the cohort with PD. For more information on the recording process, see Chapter 4.
- **Data characteristics:** The data were synchronized using m-sequence synchronization and then pre-processed using a MAF, a FOD, and energy calculation. Movement differentiation was performed by analyzing the amplitudes of the four radar nodes, with thresholds applied to both the PSG and the processed radar data. Time-series features were extracted using centered moving windows with window sizes ranging from 2 to 19. The dataset was then used to evaluate the performance of the contactless sleep analysis system.
- **Reliability and Quality:** The PSG system, which includes numerous sensors, is susceptible to data loss caused by malfunctions or sensor detachment. However, the movement calculation is resilient to this issue since it relies on a subset of sensors. Nonetheless, complete PSG failures or radar system impairments have resulted in data loss, limiting the number of recordings that could be evaluated to 32.
- **Relevance and Suitability:** In contrast to this thesis, the study on early detection of PD aims to include both movement data and vital signs. For this, the PSG recorded a broader range of signals than were strictly necessary for movement validation, resulting in a surplus of data that was not directly utilized in this thesis.

5.3 Evaluation of Movement Event Detection

Accurate movement detection is essential for reliable radar-based sleep stage classification. This thesis aims to develop a method for approximating recorded movements during sleep using radar data.

- **Data Processing:** The movement detection approach employed in this thesis comprises two key steps: signal refinement and movement information extraction. Signal refinement involves applying a MAF, taking the FOD, scaling, and absolute value transformation. This refined data is then utilized to calculate the total energy over a 30-second interval for detecting movement events.

- **Localization and Magnitude:** As a secondary objective of this thesis, the localization and differentiation between large and very large movements using radar data were explored. This side project sought to improve the accuracy and robustness of radar-based sleep stage classification by discriminating between movement magnitudes and their locations.
  The identification of large and very large movements was accomplished by employing two distinct thresholds, one for peripheral and central regions. These thresholds can be seen in Table 4.2.8.
  This approach effectively identified the desired data points. A threshold of 0.2 was established for localizing movement events by comparing the signals from all four radar nodes. This approach aimed to identify significant movements in different regions. However, the implemented localization method failed to produce satisfactory results. For further information see Chapter 6.

- **Performance Measurement:** For the measurement of the resulting approximation, the performance metrics described in Section 5.1 were introduced. These metrics were then calculated to adjust the variables used in the data processing. Furthermore, the effectiveness of the processing pipeline was also showcased in Chapter 4. This calculation was done for all suitable recordings and either averaged or otherwise collected and represented in the following Chapter 6.

5.4 Evaluation of Machine Learning Algorithms

The calculated motion events obtained through the movement event detection pipeline served as the basis for extracting specific motion features. These features were subsequently employed as input for the ML algorithms described in Section 4.4 for sleep stage classification.
This section evaluates the processing pipeline involving feature extraction and the utilization of ML algorithms.

- **Feature Extraction:** In order to get a large data set on which to train the algorithms, 15 motion features were computed from the movement data, as can be seen in Section 4.3. These features were calculated for all 32 recordings and then fed into the ML algorithms for training.
  They included the calculated differentiation of large and very large movements, as well as a variety of metrics such as mean, median, and SD.

- **Machine Learning Algorithms:** Three algorithms were chosen to classify the different sleep stages: SVM, MLP, and XGB.
  All of them were provided with the mentioned features and the corresponding ground truth, consisting of the automatically labeled sleep stages of the PSG. The algorithms were modified to include a new modality and a new dataset type.
  The dataset used in this thesis, required further processing to match the format of the original "Mesadataset" used with the same algorithms. This included changing the labeling of the PSG ground truth, to match the different sleep stage classifications categories Wake/Sleep or Wake/NREM/REM.

- **Performance Metrics:** In accordance to the methodology outlined in the Section 4.4, seven performance metrics were obtained. These metrics provide a comprehensive assessment of the efficacy of the ML algorithms employed.

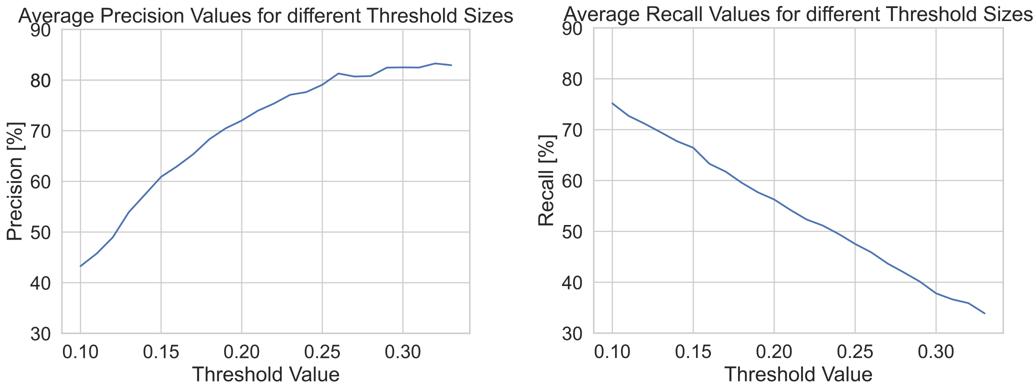
Chapter 6

Results

6.1 Movement Detection

6.1.1 Optimization of Threshold and Window Size

The accompanying visualizations, in Figure 6.1 illustrate the correlation between the diverse performance metrics outlined in Section 5.1. The Precision metric exhibits a gradual improvement in tandem with the rising threshold applied to the signal. Conversely, the Recall results manifest a diminishing performance with the advancement of the Threshold. This phenomenon stems from the aforementioned noise, as well as the small movement suppression effect of the threshold.
The F1-score is a composite of both Precision and Recall, and it attains its peak at the optimal alignment of the two performance metrics. These visualizations in 6.1 were generated utilizing a subset of 10 recordings.



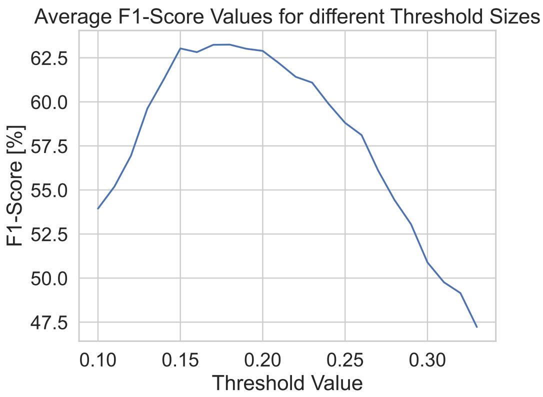
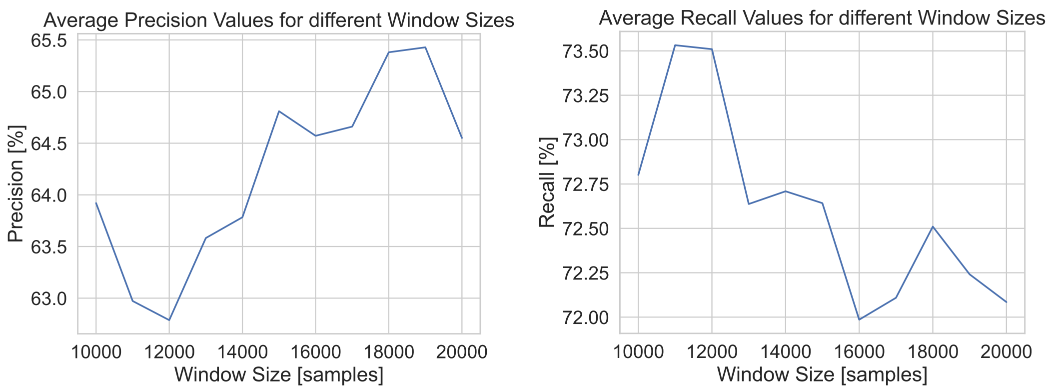


Figure 6.1: These metrics were calculated based on the threshold values.

Variable window sizes have a similar impact on noise filtering and the inadvertent suppression of minor movements as the threshold. Accordingly, the MAF window size exhibits a similar impact on Precision and Recall, albeit with a lower magnitude.
Figure 6.2 illustrates this effect in relation to the MAF window size, where Precision generally exhibits an ascending trend, contrasting with the Recall metric, which declines.
The F1-score follows a similar trajectory to the threshold, reaching a peak at the optimal equilibrium of Precision and Recall. The performance visualization of the various window sizes was generated using a dataset of 24 recordings.



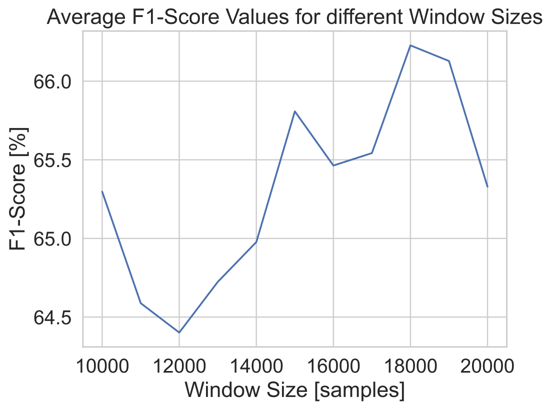


Figure 6.2: These metrics were calculated based on the window size of the MAF.

To identify the optimal thresholds and window sizes for maximizing movement detection, performance metrics were calculated for thresholds ranging from 0.17 to 0.20 and window sizes between 17 000 and 22 000.

Movement detection can be affected by two types of errors: FP (movement is detected when there is no ground truth) and FN (ground truth is present but not detected). To assess these errors, the metrics precision, recall, and F1 were calculated. Precision measures the ratio of true positive (TP) to all detected movements. Raising the threshold and window size effectively filters out noise, but also reduces the detection of genuine movements, consequently improving the precision of movement detection.

Recall, on the other hand, measures the ratio of TP to all movements detected by the ground truth. As the threshold and window size increase, fewer movements are detected, leading to lower recall. To balance precision and recall, the F1-score was also calculated. The F1-score reaches its maximum value when precision and recall are equal. Figure 6.3 demonstrates the trade-off between precision and recall. The Figure also highlights the combination of both threshold and window size that yields the highest F1-score, with the least amount of outliers.

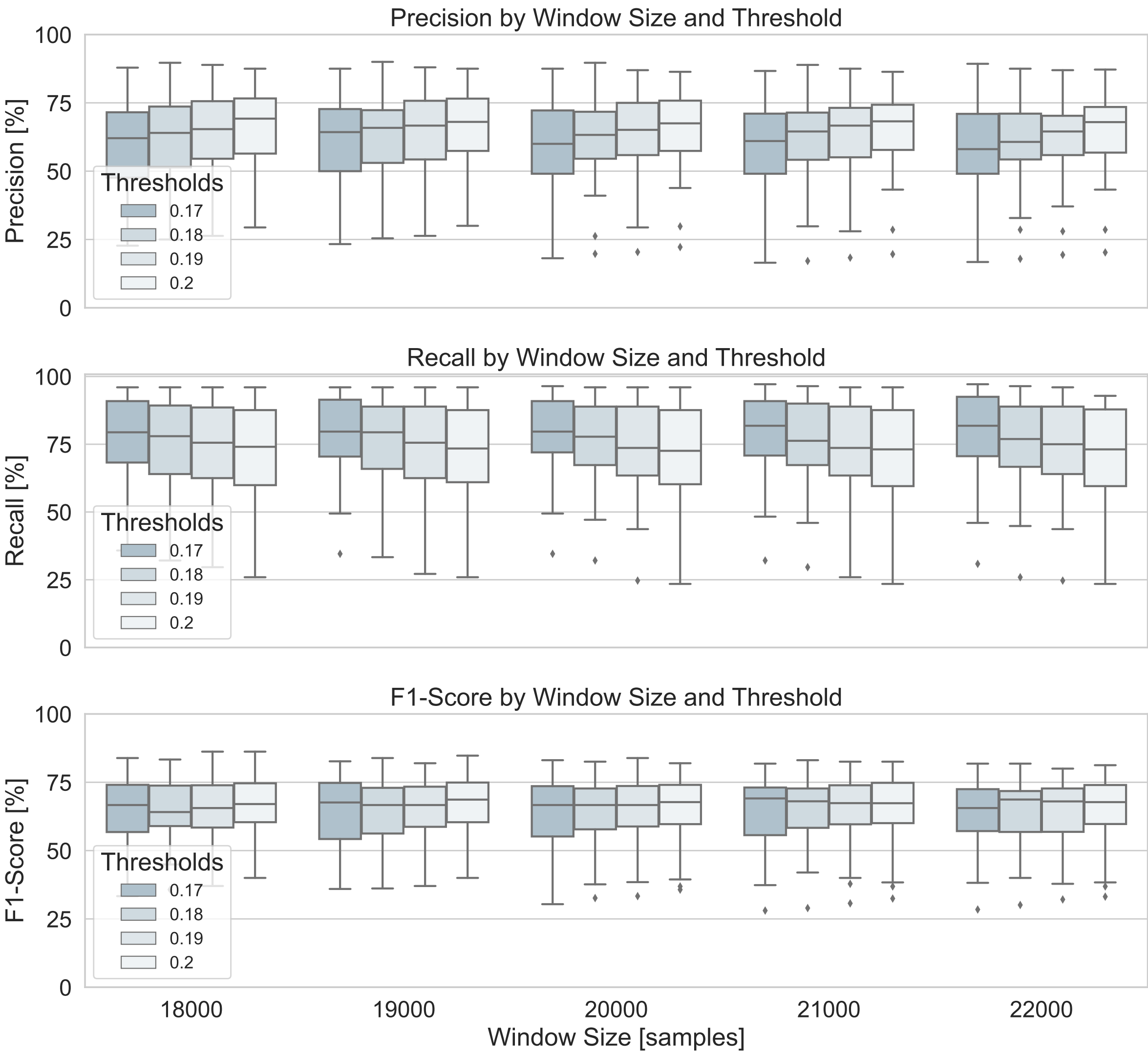


Figure 6.3: The figure shows the distribution of the precision, recall, and F1 metrics of all sleep recordings. It also highlights the detection rates of movement of the processing pipeline for specific thresholds and window sizes.

6.1.2 Calculation of the Performance Metrics

To identify the optimal window size and threshold, all recordings were systematically evaluated. A window size of 19 001 and a threshold of 0.2 were determined to achieve the highest average F1-score with the least number of outliers.
This window size corresponds to an average duration of approximately 8.5 s when synchronized with a sampling rate of 2 kHz. Using these settings and encompassing all eligible recordings, the averaged performance metrics were as shown in Table 6.1.

Precision	Recall	F1
65.56%	73.29%	66.21%

Table 6.1: The overall performance of movement detection across all recordings, as assessed by the mean values of precision, recall, and F1 score.

A significant proportion of FP and FN data points are situated within adjacent time intervals (30 s) to the TP or true negative (TN) indices. The false classification of data points could be attributed to the signal-broadening effect introduced by the MAF algorithm. As the signal is averaged over a sequence of data points, sharp edges tend to be smoothed out, resulting in a less distinct representation of the movement events. This phenomenon is evident in the example depicted in Figure 6.4.
A more comprehensive discussion of the problem leading to this phenomenon was done in Chapter 7.2.2.

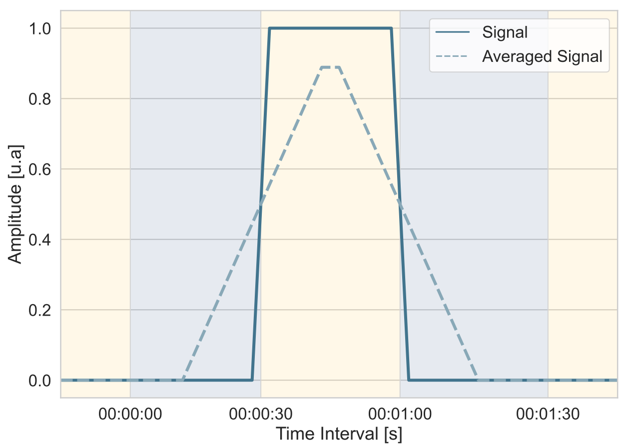


Figure 6.4: This figure demonstrates the signal-broadening effect over three-time intervals, introduced by the MAF algorithm on a simplified rectangular example signal. This phenomenon can lead to movement events detected in one timeinterval extending into neighboring intervals.

When analyzing all available recordings, the mean number of FP or FN in close proximity led to the proportions of all incorrectly identified instances, as presented in Table 6.2.

All False Data Points	False Positive	False Negative
33.18%	19.92%	15.26%

Table 6.2: The mean proportion of inaccurate data points in adjacent time intervals.

Figure 6.5 depicts the percentage of incorrectly detected movement events that are within a close temporal window (30 s) of genuine movements, for each recording. It also provides a comprehensive overview of the distribution of FP and FN data points relative to the total number of falsely identified instances. This serves to approximate the overall temporal proximity of falsely identified movement events to genuine movement events.

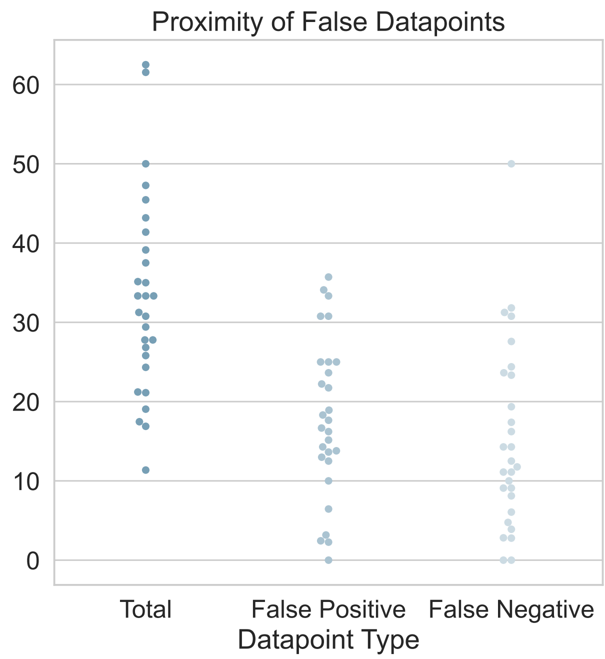


Figure 6.5: The proportion of FP and FN detections across temporally adjacent time windows, with each data point representing an individual dataset.

6.1.3 Localization of Movement

Although not the primary focus of this thesis, the localization of movement was investigated as a secondary objective.
This said, the localization data obtained was deemed imprecise and therefore unsuccessful, primarily due

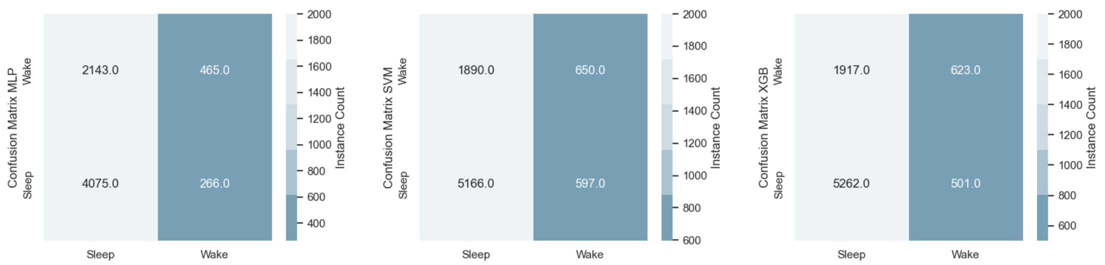
to the inability to consistently track or differentiate between similar movement events. Despite these limitations, the differentiation between large and very large movements produced satisfactory results and was successfully integrated into the feature extraction process.

6.2 Machine Learning based Sleep Stage Classification

The performance metrics of sleep score estimation were evaluated using three ML algorithms: MLP, SVM, and XGB. These algorithms were employed for sleep stage classification in two categories: Wake/Sleep and Wake/NREM/REM.

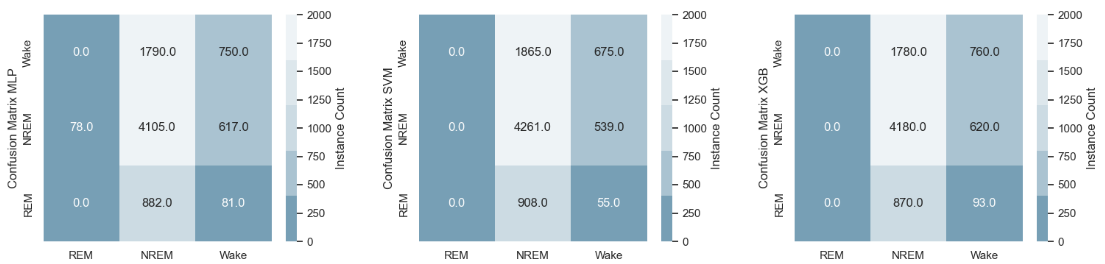
For each category a confusion matrix was generated, to depict the number of instances that were correctly classified (TPs and TNs) and incorrectly classified (FPs and FNs) for each algorithm. These matrices can be used to calculate metrics, such as precision, recall, or the F1-score.

The first matrices in Figure 6.6 depict the binary classification results, indicating the proportion of instances correctly classified as Wake or Sleep events. The latter matrices illustrated in 6.7 show the three-stage classification outcomes, further differentiating between Wake, NREM, and REM.



(a) (b) (c)

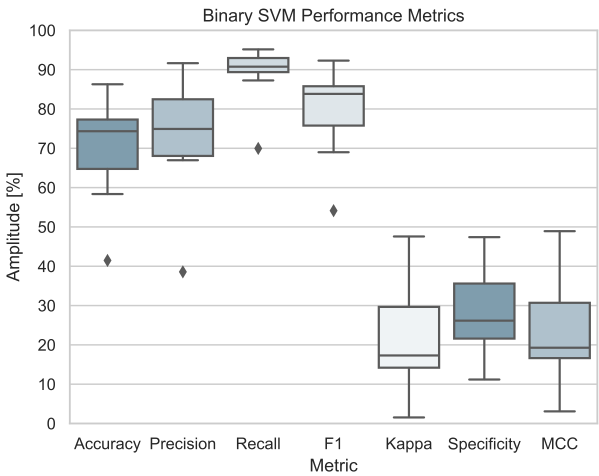
Figure 6.6: The confusion matrices of the binary Sleep/Wake classification.



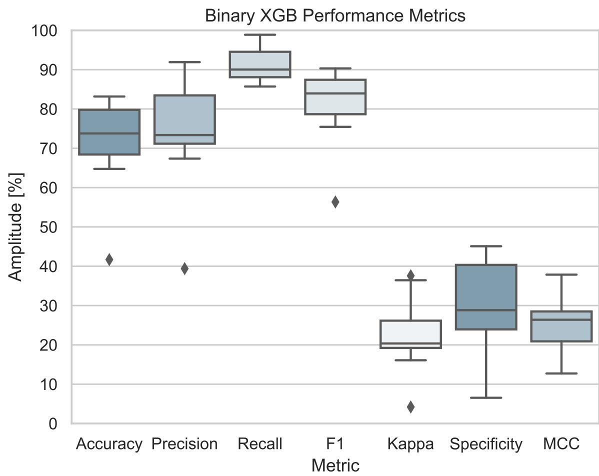
(a) (b) (c)

Figure 6.7: The confusion matrices of the three-stage sleep stage classification.

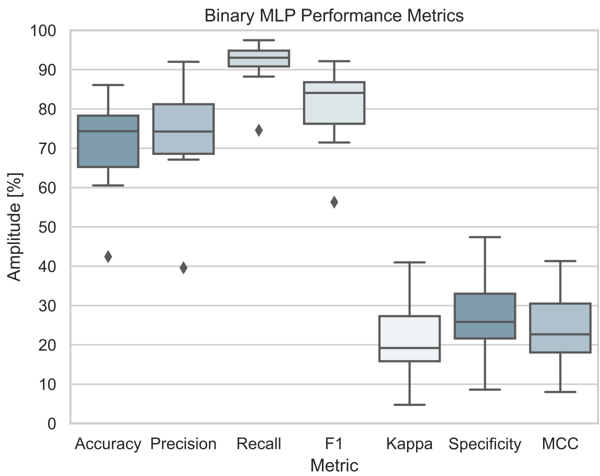
As mentioned in Section 5.1 the results of the ML algorithms were provided in the metrics **Accuracy**, **Precision**, **Recall**, **F1**, **Kappa**, **Specificity**, and **MCC**.
The results of the binary classification and three-stage classification tasks are illustrated in Figures 6.8 and 6.9, respectively.



(a) Binary classification of the SVM algorithm.



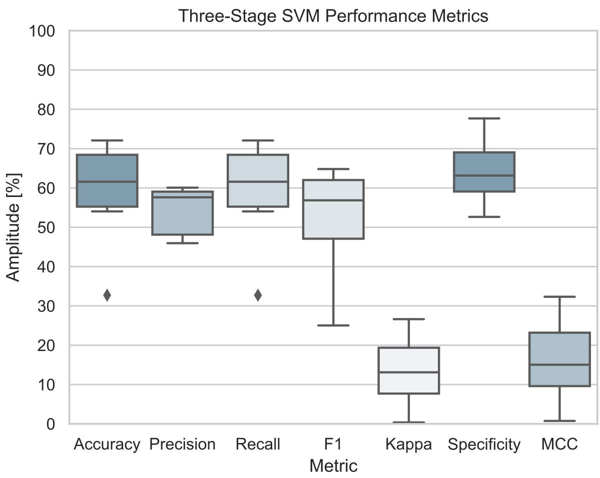
(b) Binary classification of the XGB algorithm.



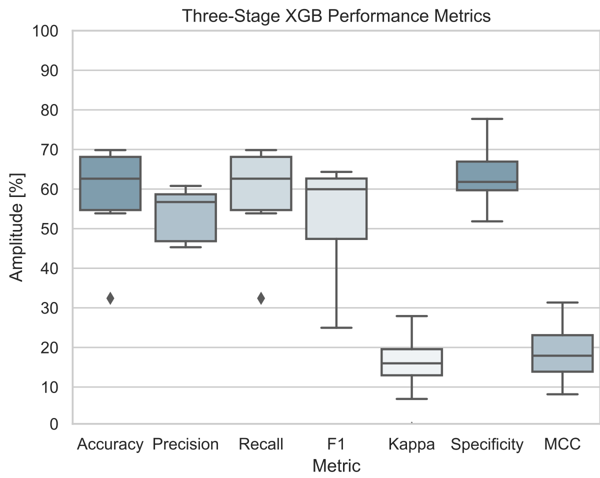
(c) Binary Classification of the MLP algorithm.

Figure 6.8: These boxplots depict the distribution of the performance metrics accuracy, precision, recall, F1, kappa, specificity, and mcc across the three ML algorithms MLP, SVM and XGB for the binary classification of sleep stages.

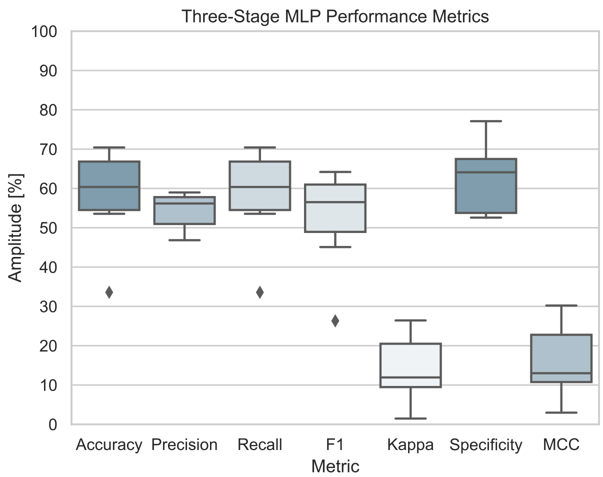
Three ML algorithms, MLP, SVM, and XGB, were evaluated for binary sleep stage classification. The boxplots in Figure 6.8 show that the performance metrics of all three algorithms are relatively high. However, the XGB algorithm consistently outperforms the other two methods. This suggests that the XGB model is, with an F1-score of 81%, more accurate and consistent than the MLP or SVM models. The resulting metrics can be seen in Table 6.3.



(a) Three-stage classification of the SVM algorithm.



(b) Three-stage classification of the XGB algorithm.



(c) Three-stage classification of the MLP algorithm.

Figure 6.9: Performance metrics accuracy, precision, recall, F1, kappa, specificity, and mcc across the three ML algorithms MLP, SVM and XGB for the three-stage classification of sleep stages.

Figure 6.9 presents a comparative analysis of the three-stage sleep classification results obtained using the MLP, SVM, and XGB algorithms. These results are summarized using seven performance metrics, analogous to the binary classification task.

- **SVM:** SVM emerged as the most proficient algorithm in the three-stage sleep classification, consistently outperforming the XGB and MLP algorithms across most metrics.
- **XGB:** The XGB algorithm demonstrated comparable performance in terms of accuracy, precision, recall, and F1-score to the MLP algorithm. However, XGB exhibited a slight inferiority compared to SVM across these metrics.
- **MLP:** Closely following XGB, the MLP algorithm demonstrated a similar pattern of performance across all metrics, albeit slightly lower than SVM.

6.2.1 Algorithm Performance

This chapter summarizes the performance of three ML algorithms (MLP, SVM, and XGB) in classifying sleep data.

Binary Classification

- **XGB:** Emerged as the top performer, achieving an average F1-score of 81%, surpassing MLP (80%) and SVM (80%).
- **MLP and SVM:** Despite not achieving the same value as XGB, with 80% they also demonstrated respectable F1-scores in Wake/Sleep identification.

Three-Stage Classification

- **SVM** consistently outperformed XGB (54%) and MLP (53%) to achieve an average F1-score of 63% in three-stage classification.
- **XGB** exhibited a lower performance than SVM, and was comparable to the results obtained with MLP.
- **MLP** algorithm displayed moderate performance in both binary and three-stage sleep classification. However, its performance did not reach the level of the other two algorithms.

These findings are summed in the Table 6.3 for binary classification and Table 6.4 for three-stage classification.

Algorithm	Accuracy	Precision	Recall	F1	Kappa	Specificity	MCC
MLP	71%	73%	91%	80%	21%	27%	24%
SVM	70%	73%	89%	80%	21%	29%	22%
XGB	71%	74%	92%	81%	23%	29%	26%

Table 6.3: Performance metrics for Wake/Sleep classification.

Algorithm	Accuracy	Precision	Recall	F1	Kappa	Specificity	MCC
MLP	59%	54%	59%	53%	14%	63%	16%
SVM	68%	60%	68%	63%	24%	67%	26%
XGB	60%	54%	60%	54%	15%	64%	18%

Table 6.4: Performance metrics for Wake/NREM/REM classification.

Chapter 7

Discussion

This Chapter discusses the key findings, contributions, and limitations of this thesis. The discussion is organized into sections focusing on movement detection, sleep staging, and a comparison with existing literature.

This work introduces a method for approximating movement events and extracting movement features, these were then employed in ML algorithms for sleep stage classification. The study involved the recording of 37 participants who will later serve as a control group for patients with PD in a broader study.

Despite the occurrence of signal loss and other data impairments, as described in Chapter 4.1.4, only five recordings were excluded from the final evaluation process. This was due to the fact that the PSG system recorded more data channels than were strictly necessary for the movement detection pipeline. As a result, a loss of data in some channels did not necessarily mean that the entire recording had to be discarded. However, this surplus data could still be leveraged to enhance the accuracy of the automated sleep stage classification used as the ground truth for the ML algorithms.

The resulting usable dataset of 32 recordings, consisted of a contactless radar signal as well as a conventional PSG. Additionally, the movement was differentiated in either large or very large movement data if certain thresholds were met, as can be seen in Table 4.4.

7.1 Findings

Movement Detection

The movement detection method developed in this thesis achieved an F1-score of 66%, a recall rate of 73%, and a precision of 65%. These results suggest that the system's ability to detect movements is slightly lower than anticipated.
It also tended to produce both FP and FN values, often in close proximity to TPs. While the resulting performance of movement detection using radar data exhibits only reasonable accuracy, the significant proportion of false data points nearby, as demonstrated in section 6.1.2, indicates that movement can be detected within a tight timeframe. Despite this, it also highlights the potential for enhancement by reducing FP and FNs.
The higher recall relative to precision may suggest that the signal widening introduced by the MAF plays a significant role in the performance metrics. This effect is evident in the visualization 4.2, and described in Section 7.2.2.
The assessment of the movement detection pipeline's performance is further complicated by the absence of reported performance metrics in the studies analyzed, with only sleep stage results being documented.

Sleep Staging

ML algorithms were applied to classify sleep stages using the calculated movement data.
To compute the various performance metrics outlined in Chapter 5.1, the employed ML algorithms generated confusion matrices. These depict the proportion of TPs and TNs relative to FPs and FNs, and are presented in Section 6.2.
In the binary sleep staging task, all employed algorithms exhibited a tendency to overpredict the sleep phase. A trend, that was particularly evident in the NREM phase, as shown by the three-stage classification results. This is exemplified by the confusion matrices for both binary (Figure 6.6) and three-stage (Figure 6.7) categorization. One possible explanation for this overprediction may be the exclusive reliance on movement data in this thesis. This could result in misinterpreting wake phases with minimal movement preceding and following the NREM sleep phase as sleep.
The algorithms achieved an F1-score of 81% for the binary (Wake/Sleep) and 63% for the three-stage classification (Wake/NREM/REM).
Among the considered ML algorithms, XGB performed best for binary classification, while SVM yielded superior performance for the three-stage categorization. Overall, the metrics for the ML-based sleep staging were within the expected range. The only deviation from expected performance was observed in the specificity metric, with values ranging between 27% and 29% for binary classification and between 63% and 67% for three-stage sleep staging.
The F1-scores achieved in this study were marginally lower than those reported in a similar study by Rah-

man et al. [Rah15], which achieved 86.2% accuracy for binary classification and 75.8% accuracy for three-stage sleep staging. This difference may be attributed to the imperfect ground truth sleep phase annotations provided by the PSG.

Movement Localization

Although not the main focus of this thesis, the differentiation of movement events was only partially successful. While the distinction between larger movements was successfully attained, the identification of movement locations was excluded from the extracted features due to imprecise results. This is further described in Chapter 6.1.3.

7.2 Comparison with Existing Literature

Movement Detection

The movement detection pipeline yielded lower performance than anticipated. This could be attributed to a multitude of factors, including the influence of the MAF as described in Section 7.2.2.
Rahman et al. [Rah15] documented an average recall rate of 86% for movement detection utilizing radar data. Nonetheless, the authors did not provide complete information on precision and F1-score, limiting a full assessment of their method's overall effectiveness. Kagawa et al. [Kag16] evaluated their motion detection capabilities by using a body motion index to distinguish between sleep and wake states, achieving a sensitivity of 76% and a specificity of 77%. Conversely, the study failed to provide comprehensive performance metrics for motion detection, hindering a thorough evaluation of the method's overall efficacy. Hong et al. [Hon18] did not share their method for approximating movement with the ground truth data. They solely presented the sleep stage classification results obtained using their ML algorithms.
The absence of comprehensive and comparable performance metrics across the examined literature has limited the evaluation of movement detection from the radar data. Consequently, the focus shifted towards comparing the performance of different ML algorithms to classify the sleep stages.

Sleep Staging

Sleep staging was done with two sets of classification metrics, first, it was differentiated between Wake/Sleep, and secondly between Wake/NREM/REM. These same classification tasks have been addressed in various studies with similar research objectives.

Most previous studies on sleep staging have incorporated vital signs along with movement data, achieving detection rates ranging from 66.4% to 89.6%. Only Rahman et al. [Rah15] reported their findings solely based on movement data, achieving an F1-score of 86.2% for Wake/Sleep and 75.8% for Wake/NREM/REM detection.

The obtained performance metrics of this thesis were slightly lower compared to the results reported by Rahman et al. [Rah15]. However, it remains challenging to definitively attribute this discrepancy to shortcomings in the processing pipeline, inaccuracies in the ground truth provided by the PSG, or a combination of these factors.

The reliance on automatically extracted sleep stage information from the PSG introduces an inherent ambiguity in the evaluation process. Not only that, but given the focus of this study on movement as the sole input for sleep staging, it was anticipated that the overall detection rate would be lower compared to studies incorporating vital signs.

Despite these limitations, the obtained classification results demonstrate the significance of movement in the sleep staging process.

7.2.1 Methodology of Related Work

The approach for the extraction of movement data employed in this thesis resembles that of Hong et al. [Hon18]; with the difference being, that the signal energy was measured over a 6-minute period, which differs from the 30-second interval used in this work.

Rahman et al. [Rah15] adopted a different approach by calculating the RMS of the signal and dividing it into 30-second intervals.

This Thesis integrated elements from both approaches, employing 30-second intervals for movement detection while calculating the mean movement value within each interval. The strategy employed aimed to distinguish between different movement magnitudes by analyzing mean values while effectively addressing the influence of outliers.

In addition, this thesis uniquely incorporated the FOD of the filtered signal to enhance movement event detection. Unlike previous studies that solely relied on a MAF, the proposed preprocessing pipeline aimed to simultaneously address signal distortion and enhance movement detection by combining MAF with a fixed threshold and the FOD.

7.2.2 Effects of the Moving Average Filter

The MAF was utilized to suppress high-frequency noise and to handle signal anomalies.
Although effective in smoothing signals, the MAF can inadvertently blur signal events, particularly sharp changes or short-duration pulses. This occurs because the MAF averages the value of all data points in the window size, resulting in a widening of the motion event, as can be seen in the visualization in Figure 4.2 and Figure 6.4. This can result in FPs, if the motion occurs at the edge of a time interval, as the broadened signal may spill over into the adjacent time segments.
To decrease the number of FP detections, it may be beneficial to use a smaller window size and adjust the threshold accordingly. Alternatively, other filters, such as a Wiener filter, could be considered for improved performance.

Localization

The localization effort of this work was unique when compared to the examined related work. Even though it was only partially successful, it gave valuable insight into the magnitude of the recorded movement. Furthermore, the approach for differentiating between areas of movement is still valid and could be picked up and refined in future work.
The inability to accurately localize movement could arise from various error sources, including wrong thresholding or insufficient discrimination between the maximum amplitude of other significant movements.
Setting an excessively high threshold could effectively eliminate noise but also inadvertently discard numerous smaller movements. Moreover, overly inclusive criteria for classifying large movements might obscure distinct movement events. This issue might stem from the lowered threshold applied to peripheral nodes for classifying movements as large. As a result, movements primarily concentrated in inner radar nodes but displaying low amplitude in outer nodes could be mistakenly identified as large movements.

7.3 Strengths and Limitations

Strengths

The thesis employed a data-driven methodology leveraging a substantial dataset of radar signals, enabling a thorough evaluation of the movement detection and sleep stage classification algorithms. This approach offered several advantages, including objective and unbiased assessment of the processing pipeline, performance evaluation across diverse movement patterns, and the ability to incorporate data irregularities that may be overlooked in smaller datasets.
Furthermore, the use of ML algorithms makes it possible to optimize the classification of sleep stages in this work without the need for frequent code adjustments.
The methodology is evaluated using a rigorous set of metrics, which ensures that the results are reliable and generalizable.

Limitations

Despite the valuable insights obtained, the thesis acknowledges several limitations that warrant further investigation. A notable limitation pertains to the inherent unreliability of the ground truth data. Medical professionals from the university clinic of the FAU, will score the sleep stages, but this process will occur after the completion of the thesis. Consequently, this study utilizes the automated PSG classification as the ground truth for evaluating the performance of the ML algorithms. This approach introduces potential limitations, as automated methods may not be as reliable as human expert annotations. These limitations could reduce classification performance and introduce uncertainty into the evaluation process, making it difficult to determine the precise cause of any errors.

Another challenge was the incomplete synchronization of the radar and PSG signals. The available synchronization methods from empkins-io were not operating correctly at the time of this research. This said this research primarily focused on large body movements, which are generally less susceptible to minor temporal variations compared to smaller ones. Consequently, only the “local_datetime” index was employed for signal synchronization.

While a comprehensive grid search was conducted to optimize the movement detection performance, exploring threshold values ranging from 0.01 to 0.31, the final evaluation focused on a narrower range between 0.17 and 0.20. Within this restricted range, the threshold of 0.20 demonstrated the highest performance. This suggests that the optimal value may lie just beyond the explored range, and further investigation is warranted to identify the optimal threshold value.

This work was done in the context of PD early detection. However, the dataset employed for developing this methodology was exclusively comprised of the healthy control group, preventing a comprehensive evaluation of PD-specific movement detection capabilities. This said the main goal of this thesis was to create a model that could estimate the movement patterns captured by the PSG using radar data. This implies that even though only the control group of the PD cohort was employed in this study, the data were suitable for this purpose. It is worth mentioning that the quality of the recorded sleep had little impact on the processing pipeline implemented in this thesis. The recorded data's most crucial aspect was exhibiting traceable and logically consistent movement features.

Chapter 8

Conclusion and Outlook

This research demonstrates the feasibility of utilizing radar data to discern movement patterns during sleep. It also underscores the potential importance of movement as a feature for sleep stage classification using ML algorithms.

8.1 Conclusion

Movement Detection

To achieve the goal of using movement for sleep stage classification, the detected motion in the radar signal was compared to the ground truth of the PSG.
The processing pipeline achieved reasonable accuracy in detecting movement, although its performance in this regard was slightly lower than expected. It is important to note that a comprehensive comparison to related work, is not possible due to the lack of comparable data.

Sleep Staging

The sleep stage classification performance achieved in this study is comparable to the examined studies [Kag16; Rah15; Hon18]. For binary classification (Wake/Sleep), the highest performing ML algorithm XGB, yielded a precision of 74%, recall of 92%, and F1-score of 81%. In the case of three-stage classification (Wake/NREM/REM), the results were 60% precision, 68% recall, and 63% F1-score, with the SVM algorithm. Rahman et al. [Rah15] reported slightly higher performance results for the sleep stage classification, using only movement data. The full comparison of the results is presented in Section 7.2.

8.2 Outlook

The results of this thesis demonstrate the potential of radar as a promising alternative to PSG for movement detection.
To further enhance the findings of this thesis, future research could investigate alternative noise reduction methods to the application of a MAF. This could potentially reduce FP and FN detection rates, thereby enhancing the overall performance of the movement detection system.
Further investigations should also assess whether employing higher thresholds for noise suppression could potentially enhance the performance of the movement detection pipeline.
The localization accuracy in this thesis could be improved by refining the applied thresholds or by adjusting the radar node positioning to better capture the movements of different body parts.
Furthermore, alternative approaches to motion localization beyond solely considering the amplitude of the movement could be explored. This could help to gain a more comprehensive understanding of the movement's source.
Future research should investigate how sleep staging can be accomplished using movement as a feature, without the potential of misclassifying wakeful periods with minimal movement, as sleep. One potential solution, as demonstrated by recent studies [Rah15; Hon18; Hon19], lies in integrating multiple sleep features, including movement, respiratory rate, and heart rate. This approach can enhance the robustness and overall performance of sleep staging. This work highlights the importance of movement as a feature for sleep stage detection, achieving reasonable detection rates for the examined sleep phases.

Appendix A

Questionnaire

Fragebogen

1 PSQI

During the past month, what time have you usually gone to bed at night?

Please enter the time using the 24-hour system (e.g. "22.00").

During the past month, how long (in minutes) has it usually taken you to fall asleep each night?

minutes

During the past month, what time have you usually gotten up in the morning?

Please enter the time using the 24-hour system (e.g. 08.00).

During the past month, how many hours of actual sleep did you get at night? (This may be different than the number of hours you spent in bed.)

Please enter the hours of sleep.

hours

During the past month, how often have you had trouble sleeping because you cannot get to sleep within 30 minutes?

During the past month, how often have you had trouble sleeping because you wake up in the middle of the night or early morning?

During the past month, how often have you had trouble sleeping because you have to get up to use the bathroom?

During the past month, how often have you had trouble sleeping because you cannot breathe comfortably?

During the past month, how often have you had trouble sleeping because you cough or snore loudly?

During the past month, how often have you had trouble sleeping because you feel too cold?

During the past month, how often have you had trouble sleeping because you feel too hot?

During the past month, how often have you had trouble sleeping because you had bad dreams?

During the past month, how often have you had trouble sleeping because you have pain?

Other reason(s), please describe:

How often during the past month have you had trouble sleeping because of this?

During the past month, how would you rate your sleep quality overall?

During the past month, how often have you taken medicine to help you sleep (prescribed or "over the counter")?

During the past month, how often have you had trouble staying awake while driving, eating meals, or engaging in social activity?

During the past month, how much of a problem has it been for you to keep up enough enthusiasm to get things done?

Do you have a bed partner or room mate?

If you have a room mate or bed partner, ask him/her how often in the past month you have had loud snoring:

If you have a room mate or bed partner, ask him/her how often in the past month you have had long pauses between breaths while asleep :

If you have a room mate or bed partner, ask him/her how often in the past month you have had legs twitching or jerking while you sleep:

If you have a room mate or bed partner, ask him/her how often in the past month you have had episodes of disorientation or confusion during sleep:

Other restlessness while you sleep; please describe:

How often in the past month have you experienced this other restlessness you described?

2 STOP BANG

Do you snore loudly?

Louder than talking or loud enough to be heard through closed doors

- ○ Yes
- ○ No

Do you often feel tired, fatigued, or sleepy during the daytime?

- Yes
- No

Has anyone observed you stop breathing during sleep?

- Yes
- No

Do you have (or are you being treated for) high blood pressure?

- Yes
- No

BMI

- <=35
- >35

Age

- <=50
- >50

Gender

- Female
- Male
- Other

3 RBD Screening Questionnaire

I sometimes have very vivid dreams.

- Yes
- No

My dreams frequently have an aggressive or action-packed content.

- Yes
- No

The dream contents mostly match my nocturnal behaviour.

- Yes
- No

It thereby happened that I (almost) hurt my bed partner or myself.

- Yes
- No

I have or had the following phenomena during my dreams:
speaking, shouting, swearing, laughing loudly

- Yes
- No

I have or had the following phenomena during my dreams:
sudden limb movements, “fights”

- Yes
- No

I have or had the following phenomena during my dreams:
gestures, complex movements, that are useless during sleep, e.g., to wave, to salute, to frighten mosquitoes, falls off the bed

- Yes
- No

I have or had the following phenomena during my dreams:
things that fell down around the bed, e.g., bedside lamp, book, glasses

- Yes
- No

It happens that my movements awake me.

- Yes
- No

After awakening, I mostly remember the content of my dreams well.

- Yes
- No

My sleep is frequently disturbed.

- ◯ Yes
- ◯ No

I have/had a disease of the nervous system (e.g., stroke, head trauma, parkinsonism, RLS, narcolepsy, depression, epilepsy, inflammatory disease of the brain)

- ◯ Yes
- ◯ No

If yes, which?

4 SF-12 Health Survey

In general, would you say your health is:

- ◯ Excellent
- ◯ Very good
- ◯ Good
- ◯ Fair
- ◯ Poor

The following questions are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

Moderate activities such as moving a table, pushing a vacuum cleaner, bowling, or playing golf.

- ◯ YES, limited a lot
- ◯ YES, limited a little
- ◯ NO, not limited at all

Climbing several flights of stairs.

- ◯ YES, limited a lot
- ◯ YES, limited a little
- ◯ NO, not limited at all

During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

Accomplished less than you would like.

- Yes
- No

Were limited in the kind of work or other activities.

- Yes
- No

During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

Accomplished less than you would like.

- Yes
- No

Did work or activities less carefully than usual.

- Yes
- No

During the past 4 weeks, how much did pain interfere with your normal work (including work outside the home and housework)?

- Not at all
- A little bit
- Moderately
- Quite a bit
- Extremely

These questions are about how you have been feeling during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling.

How much of the time during the past 4 weeks...

Have you felt calm & peaceful?

- All of the time
- Most of the time
- A good bit of the time
- Some of the time
- A little of the time
- None of the time

Did you have a lot of energy?

- All of the time
- Most of the time
- A good bit of the time
- Some of the time
- A little of the time
- None of the time

Have you felt down-hearted and blue?

- All of the time
- Most of the time
- A good bit of the time
- Some of the time
- A little of the time
- None of the time

During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting friends, relatives, etc.)?

- All of the time
- Most of the time
- A good bit of the time
- Some of the time
- A little of the time
- None of the time

5 Endseite

List of Figures

List of Tables

Bibliography

[Agn66] H. W. Agnew, Wilse B. Webb, and Robert L. Williams. “THE FIRST NIGHT EFFECT: AN EEG STUDYOF SLEEP”. In: *Psychophysiology* 2.3 (Jan. 1966), pp. 263–266. ISSN: 1469-8986. DOI: 10.1111/j.1469-8986.1966.tb02650.x. URL: http://dx.doi.org/10.1111/j.1469-8986.1966.tb02650.x.

[Alb23] Nils C. Albrecht, Markus Heyer, Marvin Wenzel, Dominik Langer, Hui Lu, and Alexander Koelpin. “Long-Distance Heart Sound Detection using 122 GHz CW Radar with 3D Printed High-Gain Antennas”. In: *2023 IEEE Radio and Wireless Symposium (RWS)*. IEEE, Jan. 2023. DOI: 10.1109/rws55624.2023.10046319. URL: http://dx.doi.org/10.1109/RWS55624.2023.10046319.

[Bor15] Olga Boric-Lubecke, Victor Lubecke, Amy Droitcour, Byung-Kwon Park, and Aditya Singh. *Doppler radar physiological sensing*. en. Ed. by Olga Boric-Lubecke, Victor M Lubecke, Amy D Droitcour, Byung-Kwon Park, and Aditya Singh. Wiley Series in Biomedical Engineering and Multi-Disciplinary Integrated Systems. Hoboken, NJ: Wiley-Blackwell, Dec. 2015.

[Buy89] Daniel J. Buysse, Charles F. Reynolds, Timothy H. Monk, Susan R. Berman, and David J. Kupfer. “The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research”. In: *Psychiatry Research* 28.2 (May 1989), pp. 193–213. ISSN: 0165-1781. DOI: 10.1016/0165-1781(89)90047-4. URL: http://dx.doi.org/10.1016/0165-1781(89)90047-4.

[Che22] Xiao Chen and Xuxiang Ni. “Noncontact Sleeping Heartrate Monitoring Method Using Continuous-Wave Doppler Radar Based on the Difference Quadratic Sum Demodulation and Search Algorithm”. en. In: *Sensors* 22.19 (Oct. 2022), p. 7646. ISSN: 1424-8220. DOI: 10.3390/s22197646. URL: https://www.mdpi.com/1424-8220/22/19/7646 (visited on 05/13/2023).

[Chu08] Frances Chung, Balaji Yegneswaran, Pu Liao, Sharon A. Chung, Santhira Vairavanathan, Sazzadul Islam, Ali Khajehdehi, and Colin M. Shapiro. "STOP Questionnaire". In: *Anesthesiology* 108.5 (May 2008), pp. 812–821. ISSN: 0003-3022. DOI: 10.1097/aln.0b013e31816d83e4. URL: http://dx.doi.org/10.1097/ALN.0b013e31816d83e4.

[Col08] Nancy A. Collop, Rachel E. Salas, Michael Delayo, and Charlene Gamaldo. "Normal Sleep and Circadian Processes". In: *Critical Care Clinics* 24.3 (July 2008), pp. 449–460. ISSN: 0749-0704. DOI: 10.1016/j.ccc.2008.02.002. URL: http://dx.doi.org/10.1016/j.ccc.2008.02.002.

[Com07] Cynthia L. Comella. "Sleep disorders in Parkinson's disease: An overview". en. In: *Movement Disorders* 22.S17 (2007), S367–S373. ISSN: 08853185, 15318257. DOI: 10.1002/mds.21682. URL: https://onlinelibrary.wiley.com/doi/10.1002/mds.21682 (visited on 05/07/2023).

[Coo19] Navin Cooray, Fernando Andreotti, Christine Lo, Mkael Symmonds, Michele T. M. Hu, and Maarten De Vos. "Detection of REM sleep behaviour disorder by automated polysomnography analysis". In: *Clinical Neurophysiology* 130.4 (Apr. 2019), pp. 505–514. ISSN: 1388-2457. DOI: 10.1016/j.clinph.2019.01.011. URL: https://www.sciencedirect.com/science/article/pii/S1388245719300306 (visited on 10/26/2023).

[Daw19] Lisa Dawel, Ann-Kathrin Seifert, Michael Muma, and Abdelhak M. Zoubir. "A Robust Genetic Algorithm for Feature Selection and Parameter Optimization in Radar-Based Gait Analysis". In: *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. Dec. 2019, pp. 674–678. DOI: 10.1109/CAMSAP45676.2019.9022515. URL: https://ieeexplore.ieee.org/abstract/document/9022515 (visited on 10/17/2023).

[Fab21] Marco Fabbri, Alessia Beracci, Monica Martoni, Debora Meneo, Lorenzo Tonetti, and Vincenzo Natale. "Measuring Subjective Sleep Quality: A Review". en. In: *International Journal of Environmental Research and Public Health* 18.3 (Jan. 2021), p. 1082. ISSN: 1660-4601. DOI: 10.3390/ijerph18031082. URL: https://www.mdpi.com/1660-4601/18/3/1082 (visited on 03/12/2023).

[Gai22] Maksym Gaiduk, Juan Jose Perea, Ralf Seepold, Natividad Martinez Madrid, Thomas Penzel, Martin Glos, and Juan Antonio Ortega. "Estimation of Sleep Stages Analyzing Respiratory and Movement Signals". In: *IEEE Journal of Biomedical and Health Informatics* 26.2 (Feb. 2022), pp. 505–514. ISSN: 2168-2208. DOI: 10.1109/jbhi.2021.3099295. URL: http://dx.doi.org/10.1109/JBHI.2021.3099295.

[Gar03] Diego Garcia-Borreguero, Oscar Larrosa, and Mauricio Bravo. “Parkinson’s disease and sleep”. en. In: *Sleep Medicine Reviews* 7.2 (Apr. 2003), pp. 115–129. ISSN: 1087-0792. DOI: 10.1053/smrv.2002.0229. URL: https://www.sciencedirect.com/science/article/pii/S1087079202902297 (visited on 04/23/2023).

[Gir21] Gabrielle Girardeau and Vítor Lopes-dos-Santos. “Brain neural patterns and the memory function of sleep”. In: *Science* 374.6567 (Oct. 2021), pp. 560–564. ISSN: 1095-9203. DOI: 10.1126/science.abi8370. URL: http://dx.doi.org/10.1126/science.abi8370.

[Goe21] R. de Goederen, S. Pu, M. Silos Viu, D. Doan, S. Overeem, W. A. Serdijn, K. F. M. Joosten, X. Long, and J. Dudink. “Radar-based sleep stage classification in children undergoing polysomnography: a pilot-study”. en. In: *Sleep Medicine* 82 (June 2021), pp. 1–8. ISSN: 1389-9457. DOI: 10.1016/j.sleep.2021.03.022. URL: https://www.sciencedirect.com/science/article/pii/S1389945721001933 (visited on 05/24/2023).

[Gra17] Michael A. Grandner. “Sleep, Health, and Society”. eng. In: *Sleep Medicine Clinics* 12.1 (Mar. 2017), pp. 1–22. ISSN: 1556-4088. DOI: 10.1016/j.jsmc.2016.10.012.

[Hon18] Hong Hong, Li Zhang, Chen Gu, Yusheng Li, Guangxin Zhou, and Xiaohua Zhu. “Noncontact Sleep Stage Estimation Using a CW Doppler Radar”. In: *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 8.2 (June 2018). Conference Name: IEEE Journal on Emerging and Selected Topics in Circuits and Systems, pp. 260–270. ISSN: 2156-3365. DOI: 10.1109/JETCAS.2017.2789278.

[Hon19] Hong Hong, Li Zhang, Heng Zhao, Hui Chu, Chen Gu, Michael Brown, Xiaohua Zhu, and Changzhi Li. “Microwave Sensing and Sleep: Noncontact Sleep-Monitoring Technology With Microwave Biomedical Radar”. In: *IEEE Microwave Magazine* 20.8 (Aug. 2019), pp. 18–29. ISSN: 1557-9581. DOI: 10.1109/MMM.2019.2915469.

[Imt21] Syed Anas Imtiaz. “A Systematic Review of Sensing Technologies for Wearable Sleep Staging”. en. In: *Sensors* 21.5 (Jan. 2021), p. 1562. ISSN: 1424-8220. DOI: 10.3390/s21051562. URL: https://www.mdpi.com/1424-8220/21/5/1562 (visited on 05/24/2023).

[Ira06] Alex Iranzo, José Luis Molinuevo, Joan Santamaría, Mónica Serradell, María José Martí, Francesc Valldeoriola, and Eduard Tolosa. “Rapid-eye-movement sleep behaviour disorder as an early marker for a neurodegenerative disorder: a descriptive study”. In: *The Lancet Neurology* 5.7 (July 2006), pp. 572–577. ISSN: 1474-4422. DOI:

10.1016/s1474-4422(06)70476-8. URL: http://dx.doi.org/10.1016/S1474-4422(06)70476-8.

[Jen97] C. Jenkinson, R. Layte, D. Jenkinson, K. Lawrence, S. Petersen, C. Paice, and J. Stradling. "A shorter form health survey: can the SF-12 replicate results from the SF-36 in longitudinal studies?" In: *Journal of Public Health* 19.2 (June 1997), pp. 179–186. ISSN: 1741-3850. DOI: 10.1093/oxfordjournals.pubmed.a024606. URL: http://dx.doi.org/10.1093/oxfordjournals.pubmed.a024606.

[K P19] Milena K. Pavlova and Véronique Latreille. "Sleep Disorders". In: *The American Journal of Medicine* 132.3 (Mar. 2019), pp. 292–299. ISSN: 0002-9343. DOI: 10.1016/j.amjmed.2018.09.021. URL: http://dx.doi.org/10.1016/j.amjmed.2018.09.021.

[Kag16] Masayuki Kagawa, Kazuki Suzumura, and Takemi Matsui. "Sleep stage classification by non-contact vital signs indices using Doppler radar sensors". In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. ISSN: 1558-4615. Aug. 2016, pp. 4913–4916. DOI: 10.1109/EMBC.2016.7591829.

[Kao13] Te-Yu Jason Kao and Jenshan Lin. "Vital sign detection using 60-GHz Doppler radar system". In: *2013 IEEE International Wireless Symposium (IWS)*. IEEE, Apr. 2013. DOI: 10.1109/ieee-iws.2013.6616776. URL: http://dx.doi.org/10.1109/IEEE-IWS.2013.6616776.

[Lau20] Timo Lauteslager, Stylianos Kampakis, Adrian J. Williams, Michal Maslik, and Fares Siddiqui. "Performance Evaluation of the Circadia Contactless Breathing Monitor and Sleep Analysis Algorithm for Sleep Stage Classification". In: (July 2020). DOI: 10.1109/embc44109.2020.9175419. URL: http://dx.doi.org/10.1109/EMBC44109.2020.9175419.

[Lee23] Ji Hyun Lee, Hyunwoo Nam, Dong Hyun Kim, Dae Lim Koo, Jae Won Choi, Seung-No Hong, Eun-Tae Jeon, Sungmook Lim, Gwang soo Jang, and Baek-hyun Kim. "Developing a deep learning model for sleep stage prediction in obstructive sleep apnea cohort using 60 <scp>GHz frequency-modulated continuous-wave</scp> radar". In: *Journal of Sleep Research* 33.1 (Sept. 2023). ISSN: 1365-2869. DOI: 10.1111/jsr.14050. URL: http://dx.doi.org/10.1111/jsr.14050.

[Leo20] Kai Wen Leong, Amanda Griffiths, Anne-Marie Adams, and John Massie. "How to interpret polysomnography". eng. In: *Archives of Disease in Childhood. Education* *and Practice Edition* 105.3 (June 2020), pp. 130–135. ISSN: 1743-0593. DOI: 10.1136/archdischild-2018-316031.

[Li18] Meiyu Li and Jenshan Lin. "Wavelet-Transform-Based Data-Length-Variation Technique for Fast Heart Rate Detection Using 5.8-GHz CW Doppler Radar". In: *IEEE Transactions on Microwave Theory and Techniques* 66.1 (Jan. 2018), pp. 568–576. ISSN: 1557-9670. DOI: 10.1109/tmtt.2017.2730182. URL: http://dx.doi.org/10.1109/TMTT.2017.2730182.

[Liu14] Lanbo Liu and Sixin Liu. "Remote Detection of Human Vital Sign With Stepped-Frequency Continuous Wave Radar". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7.3 (Mar. 2014). Conference Name: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, pp. 775–782. ISSN: 2151-1535. DOI: 10.1109/JSTARS.2014.2306995. URL: https://ieeexplore.ieee.org/abstract/document/6767120 (visited on 10/30/2023).

[Mer12] Giovanni Merlino and Gian Luigi Gigli. "Sleep-related movement disorders". en. In: *Neurological Sciences* 33.3 (June 2012), pp. 491–513. ISSN: 1590-3478. DOI: 10.1007/s10072-011-0905-9. URL: https://doi.org/10.1007/s10072-011-0905-9 (visited on 10/30/2023).

[Rah15] Tauhidur Rahman, Alexander T. Adams, Ruth Vinisha Ravichandran, Mi Zhang, Shwetak N. Patel, Julie A. Kientz, and Tanzeem Choudhury. "DoppleSleep: a contactless unobtrusive sleep sensing system using short-range Doppler radar". en. In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Osaka Japan: ACM, Sept. 2015, pp. 39–50. ISBN: 9781450335744. DOI: 10.1145/2750858.2804280. URL: https://dl.acm.org/doi/10.1145/2750858.2804280 (visited on 04/23/2023).

[Run19] Jessica Vensel Rundo and Ralph Downey. "Polysomnography". In: *Handbook of Clinical Neurology*. Elsevier, 2019, pp. 381–392. DOI: 10.1016/b978-0-444-64032-1.00025-4. URL: http://dx.doi.org/10.1016/B978-0-444-64032-1.00025-4.

[Sad15] Avi Sadeh. "III. SLEEP ASSESSMENT METHODS: SLEEP ASSESSMENT". en. In: *Monographs of the Society for Research in Child Development* 80.1 (Mar. 2015), pp. 33–48. ISSN: 0037976X. DOI: 10.1111/mono.12143. URL: https://onlinelibrary.wiley.com/doi/10.1111/mono.12143 (visited on 04/24/2023).

[Shi18] Pramila P. Shinde and Seema Shah. "A Review of Machine Learning and Deep Learning Applications". In: *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*. IEEE, Aug. 2018. DOI: 10.1109/iccubea.2018.8697857. URL: http://dx.doi.org/10.1109/ICCUBEA.2018.8697857.

[Smi99] Steven W . Smith. *The scientist and engineer's guide to digital signal processing*. en. 1999.

[Sti07] Karin Stiasny-Kolster, Geert Mayer, Sylvia Schäfer, Jens Carsten Möller, Monika Heinzel-Gutenbrunner, and Wolfgang H. Oertel. "The REM sleep behavior disorder screening questionnaire—A new diagnostic instrument". In: *Movement Disorders* 22.16 (Dec. 2007), pp. 2386–2393. ISSN: 1531-8257. DOI: 10.1002/mds.21740. URL: http://dx.doi.org/10.1002/mds.21740.

[Tof20] Ståle Toften, Ståle Pallesen, Maria Hrozanova, Frode Moen, and Janne Grønli. "Validation of sleep stage classification using non-contact radar technology and machine learning (Somnofy®)". eng. In: *Sleep Medicine* 75 (Nov. 2020), pp. 54–61. ISSN: 1878-5506. DOI: 10.1016/j.sleep.2020.02.022.

[Wil14] T. Willemen, D. Van Deun, V. Verhaert, M. Vandekerckhove, V. Exadaktylos, J. Verbraecken, S. Van Huffel, B. Haex, and Jos Vander Sloten. "An Evaluation of Cardiorespiratory and Movement Features With Respect to Sleep-Stage Classification". In: *IEEE Journal of Biomedical and Health Informatics* 18.2 (Mar. 2014), pp. 661–669. ISSN: 2168-2208. DOI: 10.1109/jbhi.2013.2276083. URL: http://dx.doi.org/10.1109/JBHI.2013.2276083.

[Wil17] Christoph Will, Kilin Shi, Sven Schellenberger, Tobias Steigleder, Fabian Michler, Robert Weigel, Christoph Ostgathe, and Alexander Koelpin. "Local Pulse Wave Detection Using Continuous Wave Radar Systems". In: *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology* 1.2 (Dec. 2017), pp. 81–89. ISSN: 2469-7257. DOI: 10.1109/jerm.2017.2766567. URL: http://dx.doi.org/10.1109/JERM.2017.2766567.

[Zha20] Ye Zhang, Rong Ren, Larry D. Sanford, Linghui Yang, Junying Zhou, Lu Tan, Taomei Li, Jihui Zhang, Yun-Kwok Wing, Jie Shi, Lin Lu, and Xiangdong Tang. "Sleep in Parkinson's disease: A systematic review and meta-analysis of polysomnographic findings". en. In: *Sleep Medicine Reviews* 51 (June 2020), p. 101281. ISSN: 10870792. DOI: 10.1016/j.smrv.2020.101281. URL: https://linkinghub.elsevier.com/retrieve/pii/S1087079220300241 (visited on 10/16/2023).

[Zuz20] José Rafael P. Zuzuárregui and Emmanuel H. During. "Sleep Issues in Parkinson's Disease and Their Management". In: *Neurotherapeutics* 17.4 (Oct. 2020), pp. 1480–1494. ISSN: 1878-7479. DOI: 10.1007/s13311-020-00938-y. URL: http://dx.doi.org/10.1007/s13311-020-00938-y.

Appendix B

Acronyms

AI artificial intelligence

BMI body mass index

CDA canonical discriminant analysis

CSV comma-separated values

CW continuous wave

ECG electrocardiography

EDF european data format

EEG electroencephalography

EMG electromyography

EOG electrooculography

EmpKinS Empatho-Kinaesthetic Sensor Technology Sensor Techniques and Data Analysis Methods for Empatho-Kinaesthetic Modeling and Condition monitoring

FAU Friedrich-Alexander-Universität

FNE first-night-effect

FOD first-order derivative

FP false positive

FN false negative

GUI graphical user interface

HRV heart rate variability

I in-phase

IMS industrial, scientific, and medical

KNN k-nearest neighbor classifier

MAF moving average filter

ML machine learning

MLM mean of large movements

MLP multi-layer perceptron

MVLM mean of very large movements

MaD machine learning and data analytics

NREM non-rapid eye movement

PD Parkinson's disease

PLMS periodic limb movement

PPG photoplethysmography

PSG polysomnography

PSQI Pittsburgh sleep quality index

Q quadrature

RBD REM sleep behavior disorder

REM rapid eye movement

RIP respiratory inductance plethysmography

RLS restless-legs-syndrome

RMS root mean square

RTOF roundtrip time of flight

SD standard deviation

STOP-Bang snoring, tiredness, observed apnea, pressure, BMI, age, neck circumference, gender

SVM support vector machines

TP true positive

TN true negative

XGB extreme gradient boosting