



University of Essex
**School of Computer Science and Elec-
tronic Engineering**

CE901-7: MSc PROJECT AND DISSERTATION

Automated Sleep Phase Detection: Unveiling the Night's Secrets with Machine Learning

Priyanshu Banerjee

Supervisor: Dr. Vito de Feo

March 25, 2024
Colchester

Contents

Abstract	6
1 Introduction	7
1.1 Significance of Sleep Phase Differentiation:	7
1.2 Limitations of Traditional Sleep Monitoring Techniques	8
1.3 Integration of Machine Learning into Sleep Research:	8
1.4 Objectives of the Dissertation:	9
1.5 Acknowledgments	9
2 Background Research and Literature Review	11
2.1 The Structure of Sleep: REM and Non-REM Cycles	11
2.2 The Role of Sleep in Health and Disease	13
2.3 Polysomnography (PSG): An In-Depth Look	14
2.3.1 Components	14
2.3.2 Procedure and Setting	18
2.3.3 Data Analysis and Interpretation:	19
2.3.4 Limitations of PSG:	19
2.4 Literature Review:	20
2.4.1 "Do Not Sleep on Traditional Machine Learning: Simple and Inter- pretable Techniques Are Competitive to Deep Learning for Sleep Scoring"	21
2.4.2 Large-Scale Automated Sleep Staging	22
2.4.3 SleepPyCo: Automatic Sleep Scoring with Feature Pyramid and Con- trastive Learning	23
2.4.4 XSleepNet: Multi-View Sequential Model for Automatic Sleep Staging	24

3	Methodology	26
3.1	Data-set Selected For the Dissertation	26
3.2	Feature Extraction	29
3.3	Feature Reduction Using PCA	30
3.3.1	PCA Time Domain	30
3.3.2	PCA Frequency Domain	32
3.4	Machine Learning	34
3.4.1	XGBoost (eXtreme Gradient Boosting)	34
3.4.2	Logistic Regression	34
3.4.3	CatBoost	35
3.4.4	Random Forest	35
3.5	Model Training and Validation Pipeline	36
3.5.1	Overview of the Pipeline	36
3.6	Results	38
3.7	Experiment 1: 4 Channels Feature Extraction + PCA (Table 3.3)	39
3.7.1	Model Performance:	39
3.7.2	Analysis:	39
3.8	Experiment 2: 13 Channels Feature Extraction (Table 3.4)	40
3.8.1	Model Performance:	40
3.8.2	Analysis:	40
3.9	Comparative Analysis	40
4	Conclusions	42

List of Figures

2.1	Sample hypnogram showing one sleep cycle (the first of the night) from NREM through REM [10].	13
2.2	EEG brain wave patterns during sleep [11].	15
2.3	. Electrode placement for (a) conventional EOG and (b) proposed forehead EOG.[12].	16
2.4	EMG electrode positions. [13].	17
2.5	Measurement of respiratory effort in sleep by 3D camera and respiratory inductance plethysmography. [14].	17
2.6	chest wall motion during respiratory cycles and sleep apnea events.[16]. . . .	18

List of Tables

3.1	Dataset Composition	27
3.2	Signal Characteristics in Sleep Study[21]	28
3.3	Model Performance Scores(4 Channels Feature Extraction + PCA)	38
3.4	Model Performance Scores(13 Channels Feature Extraction)	39

Abstract

This dissertation explores the application of machine learning algorithms in the classification of sleep stages using polysomnography (PSG) data. The study leverages advanced machine learning models, including XGBoost, CatBoost, SVM, and Logistic Regression, to analyze sleep data under two experimental conditions: one utilizing a reduced feature set derived from 4-channel data post-PCA, and the other employing a comprehensive 13-channel feature extraction. The methodology encompasses rigorous data preparation, cross-validation, and performance evaluation. The findings reveal that ensemble methods and SVM consistently outperform Logistic Regression, highlighting their potential as robust tools for sleep stage classification. This research not only contributes to the field of biomedical signal processing but also has implications for clinical practices in sleep medicine.

Introduction

Ever found yourself wondering about the mysterious voyage we embark on each night when we surrender to sleep? You're not alone. The world of slumber is a captivating realm, and within it, there's a dance of sleep phases REM, non-REM, and a whole lot more that's as intricate as it is essential to our well-being. But here's the kicker: understanding these sleep phases isn't just a quirky fascination; it's a key to unlocking a treasure trove of insights into our physical and mental health. The thing is, figuring out the specifics of these sleep phases has been a bit like detective work in the dark. Traditional methods involving sleep tests and human interpretations are great but come with their own set of challenges. That's where our superhero, Machine Learning, swoops in. So, imagine this dissertation as a flashlight in the dark, aimed at those elusive sleep phases. We're diving into the sweet spot where sleep science meets machine learning, with a mission: to build smart models that can read the language of sleep. Why? Because once we crack that code, we can open up new possibilities for understanding sleep patterns and, ultimately, improving our snooze game.

1.1 Significance of Sleep Phase Differentiation:

The sleep cycle, characterized by its nuanced stages, plays a pivotal role in various physiological functions. Non-REM sleep, encompassing lighter stages (N1 and N2) and deep slow-wave sleep (N3), is associated with essential processes such as physical restoration, immune system modulation, and hormone regulation. Conversely, REM sleep, distinguished by

heightened brain activity and vivid dreaming, contributes significantly to cognitive functions, particularly memory consolidation.

Disturbances in the natural progression through these sleep phases have been linked to a spectrum of health issues, emphasizing the critical need for accurate identification and classification. Sleep disorders, ranging from insomnia to sleep apnea, often manifest as disruptions in the balance between REM and non-REM sleep. Thus, the ability to precisely delineate these phases holds promise for early diagnosis and targeted interventions.

1.2 Limitations of Traditional Sleep Monitoring Techniques

Despite the wealth of information provided by PSG, its applicability beyond laboratory settings is constrained by logistical challenges. The necessity for specialized equipment and expert supervision, coupled with the subjectivity inherent in manual scoring, hinders scalability and accessibility in real-world scenarios. As such, there is a compelling call for innovative methodologies capable of accommodating the evolving landscape of sleep research.

1.3 Integration of Machine Learning into Sleep Research:

The emergence of machine learning as a transformative force across various disciplines has prompted its exploration in sleep research. Machine learning algorithms, capable of processing extensive datasets and identifying intricate patterns, offer a potential solution to the challenges posed by traditional monitoring methods. With the proliferation of wearable devices equipped with sensors capable of unobtrusive and continuous monitoring, the synergy between machine learning and sleep science is poised to revolutionize the field.

The application of machine learning to sleep phase detection involves the extraction of informative features from diverse physiological signals. By training models on labeled datasets and validating their performance against established benchmarks, the aim is to surpass human expertise in sleep stage scoring and overcome the limitations of traditional methods.

1.4 Objectives of the Dissertation:

Against this backdrop, the primary objectives of this dissertation become evident. Firstly, it seeks to comprehensively review existing literature on sleep phase detection, encompassing both traditional methodologies and recent advances involving machine learning. Secondly, it aims to contribute to the field by developing and evaluating machine learning models for automated sleep phase detection. The dissertation will navigate through the complexities of feature extraction, model training, and validation, with a critical eye on the practical implications of these methodologies in real-world applications.

By engaging in this academic exploration, we aspire to not only deepen our understanding of sleep dynamics but also to pave the way for scalable and accessible sleep monitoring solutions. The subsequent chapters will delve into the intricacies of the methodologies employed, the challenges encountered, and the implications of our findings in the broader context of sleep research.

In the pursuit of unraveling the mysteries of sleep through the lens of machine learning, we embark on a scientific journey that bridges the realms of data-driven methodologies and the nuanced intricacies of human sleep.

1.5 Acknowledgments

The successful completion of this dissertation is the result of collaborative efforts and dedication from multiple individuals who have contributed their expertise and time to this extensive project. We would like to express our sincere gratitude to all those who played a role in making this research endeavor possible.

First and foremost, we extend our appreciation to Syed Akash, whose collective efforts have been instrumental in the various stages of this project. Each team member brought a unique set of skills and perspectives, enriching the overall quality of the research.

We are grateful to our advisors and mentors Dr Vito De Feo, for their guidance, encouragement, and valuable insights throughout the entire research process. Their expertise and unwavering support have been indispensable in shaping the direction of this dissertation.

Additionally, we would like to acknowledge the support received from the University of Essex. The resources, facilities, and academic environment provided by the institution have

been essential in facilitating our research endeavors.

We extend our appreciation to the participants who took part in the data collection process, contributing to the foundation of our study. Their involvement has been crucial in ensuring the relevance and applicability of our findings.

As we reflect on the collaborative nature of this dissertation, we recognize and appreciate the collective effort that has gone into unraveling the complexities of sleep dynamics through the lens of machine learning. This project stands as a testament to the power of teamwork and interdisciplinary collaboration in advancing scientific understanding.

Background Research and Literature Review

The journey into the enigmatic world of sleep science and the innovative realm of machine learning begins with a comprehensive exploration of existing knowledge and scholarly discourse. This chapter, titled "Background Research and Literature Review," serves as the bedrock upon which the subsequent analysis and experimentation of this dissertation are built. It aims to intricately weave together the threads of past research, theoretical frameworks, and technological advancements to provide a coherent backdrop against which our study is positioned.

2.1 The Structure of Sleep: REM and Non-REM Cycles

Sleep, a state of altered consciousness and diminished interaction with the surrounding environment is broadly classified into two distinct types: Rapid Eye Movement (REM) and Non-Rapid Eye Movement (Non-REM) sleep. These stages form a cyclic pattern that repeats several times during a typical night's sleep, each playing a crucial role in various aspects of physical and mental health.

Non-REM Sleep: The Foundation of Sleep Architecture

Non-REM sleep is subdivided into three stages, each marked by progressively deeper levels of sleep:

1. **N1 (Light Sleep):** This initial stage acts as a transition from wakefulness to sleep. During

N1, which typically lasts for 5 to 10 minutes, the brain produces theta waves, which are slower and higher in amplitude than the alpha waves seen in wakeful relaxation. This stage is characterized by the slowing down of both heart rate and breathing, a decrease in body temperature, and a reduction in overall muscle tension. N1 accounts for roughly 5% of a night's sleep.[1]

2. **N2 (Intermediate Sleep):** Lasting approximately 20 minutes, N2 marks the onset of true sleep. It is during this stage that sleep spindles and K-complexes occur, which are thought to play a role in memory consolidation and decreasing the brain's responsiveness to external stimuli. N2 constitutes about 45-55% of total sleep in adults.[2]
3. **N3 (Deep or Slow-Wave Sleep):** N3, also known as delta sleep due to the presence of slow delta waves, is the deepest stage of sleep. It typically lasts for 20-40 minutes, but this duration decreases with age. This stage is crucial for restorative processes, such as tissue repair and growth hormone secretion. N3 is also when the body is least responsive to external stimuli, making it the most difficult stage from which to awaken. [3]

REM Sleep: A Unique Phase of Sleep Following the Non-REM stages, typically after about 90 minutes of sleep, is REM sleep. This stage is characterized by rapid eye movements, increased brain activity, and vivid dreams. REM sleep periods lengthen across the night, with the longest occurring just before awakening.

Physiological Characteristics: During REM, the brain exhibits patterns similar to wakefulness, but the body experiences atonia, a temporary paralysis of the muscles, preventing the enactment of dreams. As explained by Aserinsky and Kleitman, who first discovered REM sleep, this phase is associated with increased brain metabolism, irregular breathing, and heart rate, and is believed to be critical for learning and memory.[4]

Dreaming and Cognitive Processes: The heightened brain activity during REM sleep is closely associated with vivid dreaming. The activation-synthesis hypothesis suggests that dreams result from the brain's attempt to make sense of random neural activity during this phase. REM sleep has been implicated in various cognitive processes, including memory consolidation and emotional regulation.[5]

The Cyclic Nature of Sleep Stages Throughout the night, these sleep stages follow a cyclic

pattern, usually repeating every 90 to 120 minutes. The proportion of REM and Non-REM sleep changes as the cycles progress, with REM sleep occupying a larger portion of the cycle in the second half of the night. This cyclic nature of sleep is thought to support various restorative processes for the brain and body.[6]

Understanding the intricate structure of REM and Non-REM sleep cycles is fundamental to the study of sleep science. It provides insights into various physiological and psychological processes and is crucial in diagnosing and treating sleep disorders. The ongoing research in this field continues to shed light on the complexities of these sleep stages, unraveling the mysteries of our nocturnal life.

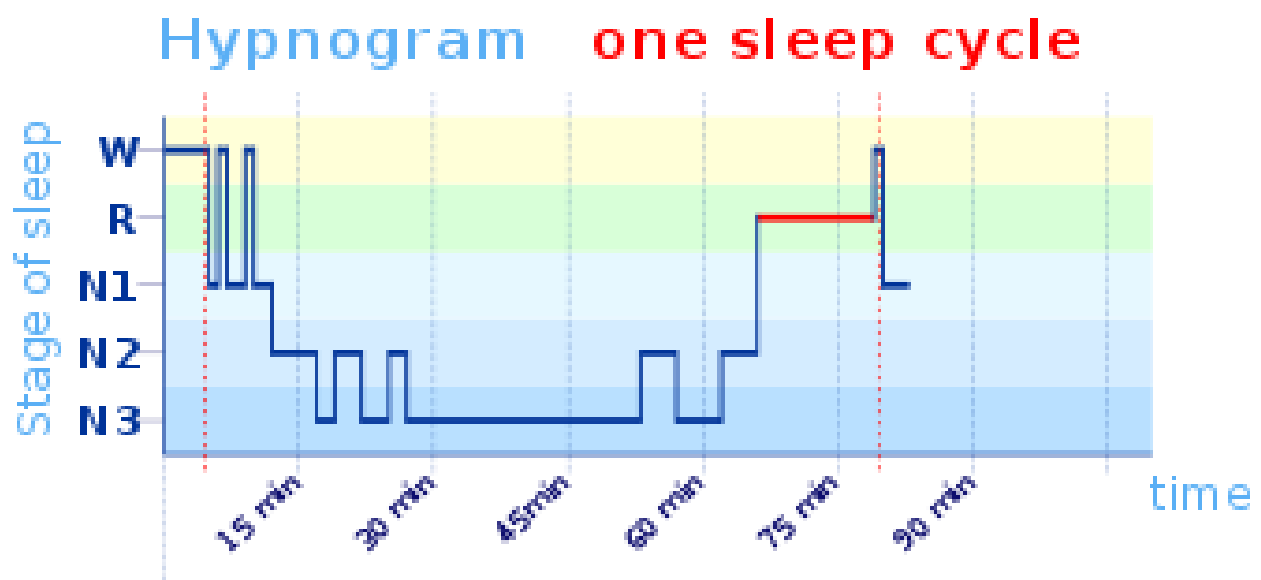


Figure 2.1: Sample hypnogram showing one sleep cycle (the first of the night) from NREM through REM [10].

2.2 The Role of Sleep in Health and Disease

Sleep plays a pivotal role in overall health and well-being, impacting a wide array of physiological and psychological processes. Disruptions in normal sleep patterns are associated with various health conditions.

Physical Health Implications:

1. **Cardiovascular Health:** Poor sleep quality and sleep disorders like sleep apnea are linked to an increased risk of cardiovascular diseases. Disturbances in sleep can lead to hypertension, arrhythmias, and even heart failure.[7]
2. **Metabolic Function:** Sleep deprivation can adversely affect glucose metabolism and appetite regulation, increasing the risk of obesity and type 2 diabetes.

Mental Health Correlations:

1. **Cognitive Functioning:** Sleep is essential for cognitive processes including learning, memory, and problem-solving. Impaired sleep can lead to deficits in these areas, impacting daily functioning and quality of life
2. **Psychiatric Disorders:** There is a strong correlation between sleep disturbances and psychiatric conditions such as depression and anxiety. Disrupted sleep can exacerbate these conditions, creating a vicious cycle of mental health deterioration and sleep problems.

The physiological and neurological underpinnings of sleep stages, and their role in health and disease, underscore the importance of sleep in human life. As this section of the dissertation elucidates, understanding these mechanisms is not just academically fascinating but also crucial in addressing the myriad of health issues associated with sleep disturbances.

2.3 Polysomnography (PSG): An In-Depth Look

Polysomnography (PSG) is a multifaceted diagnostic tool used in sleep medicine to evaluate sleep disorders. This test records several physiological parameters during sleep, each providing crucial insights into the various aspects of sleep architecture and quality. The following is a detailed examination of each component involved in a PSG.

2.3.1 Components

1. **Electroencephalogram (EEG):**

Function: The primary function of EEG in PSG is to monitor and record the electrical activity of the brain. This is crucial for identifying and differentiating between the

various stages of sleep (N1, N2, N3, and REM).

Implementation:Electrodes are strategically placed on the scalp in accordance with the 10-20 system, a standardized method for electrode placement in EEG recordings.

Data Interpretation:The EEG traces display different brain wave patterns characteristic of each sleep stage. For example, theta waves for light sleep (N1), sleep spindles and K-complexes during N2, and delta waves for deep sleep (N3).

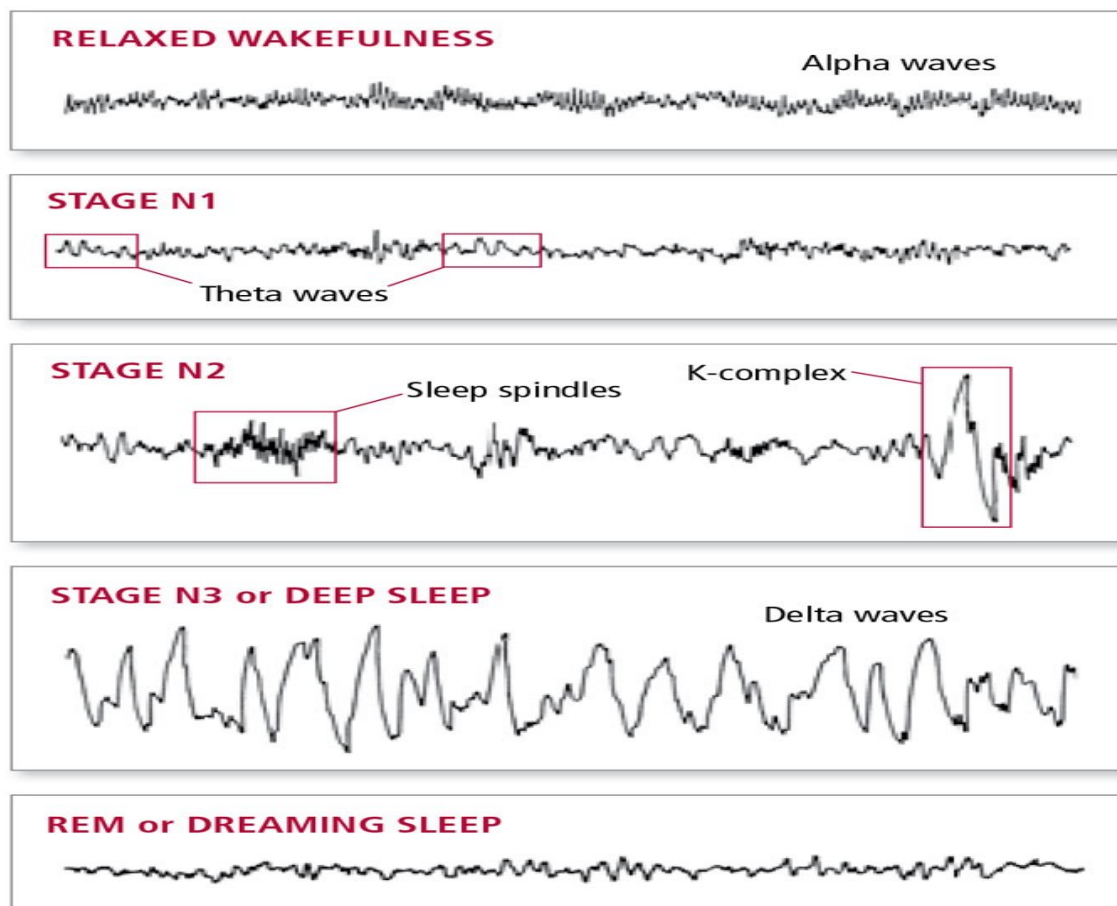


Figure 2.2: EEG brain wave patterns during sleep [11].

2. Electrooculogram (EOG):

Function: EOG records eye movements, which are particularly important for identifying REM sleep where rapid eye movements are a defining feature.

Implementation:Electrodes are placed near the eyes, typically above and below one eye or at the outer corners of both eyes.

Data Interpretation:The recording captures the movements of the eyes, distinguishing between the slow, rolling eye movements of Non-REM sleep and the rapid, jerky

movements of REM sleep.

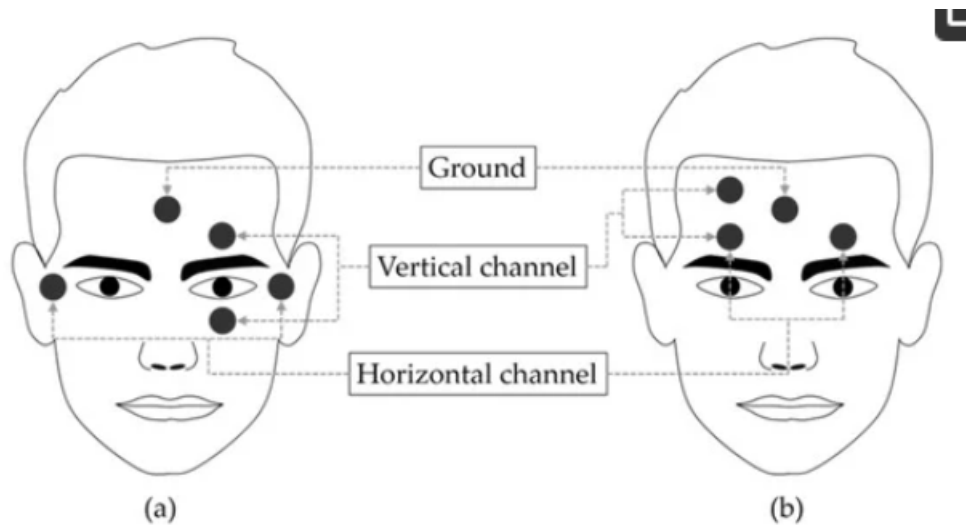


Figure 2.3: . Electrode placement for (a) conventional EOG and (b) proposed forehead EOG.[12].

3. Electromyogram (EMG):

Function: EMG in PSG is used to measure muscle activity, which helps in detecting the reduced muscle tone during REM sleep (REM atonia) and muscle relaxation in other sleep stages.

Implementation: Electrodes are placed on the chin to measure submental muscle tone and, in some cases, on the limbs to monitor limb movements.

Data Interpretation: The EMG trace varies from high muscle tone during wakefulness to reduced tone in Non-REM sleep and almost complete atonia in REM sleep, except for twitches.

4. Respiratory Monitoring:

Function: This aspect of PSG assesses breathing patterns, effort, and airflow, crucial for diagnosing sleep-related breathing disorders.

Implementation: Electrodes are placed on the chin to measure submental muscle tone and, in some cases, on the limbs to monitor limb movements.

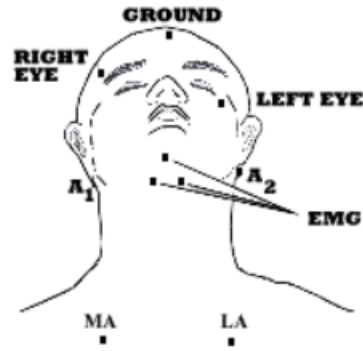


Figure 2.4: EMG electrode positions. [13].

- (a) **Effort:** Respiratory effort is measured using belts placed around the chest and abdomen.

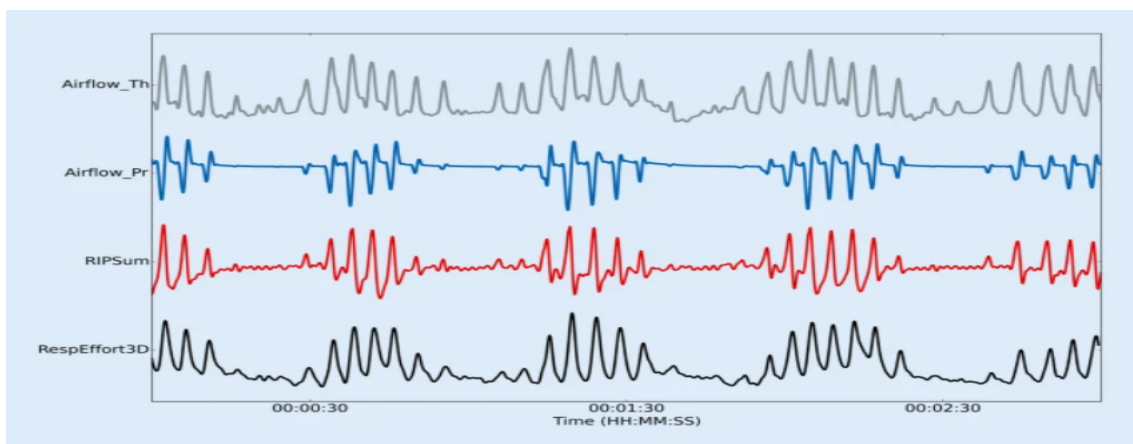


Figure 2.5: Measurement of respiratory effort in sleep by 3D camera and respiratory inductance plethysmography. [14].

- (b) **Airflow:** Airflow is typically measured using thermistors or nasal pressure transducers placed at the nose and/or mouth.
- (c) **Oxygen Saturation:** Pulse oximetry is used to monitor blood oxygen levels.

Data Interpretation: The EMG trace varies from high muscle tone during wakefulness to reduced tone in Non-REM sleep and almost complete atonia in REM sleep, except for twitches.

5. Cardiac Monitoring:

Function: Cardiac monitoring, typically through an electrocardiogram (ECG), is included to observe heart rate and rhythm during sleep.

Implementation: ECG electrodes are placed on the chest to record the electrical activity of the heart.

Data Interpretation: This data can reveal any sleep-related cardiac irregularities, such as arrhythmias, which may be related to sleep apnea or other sleep disorders.

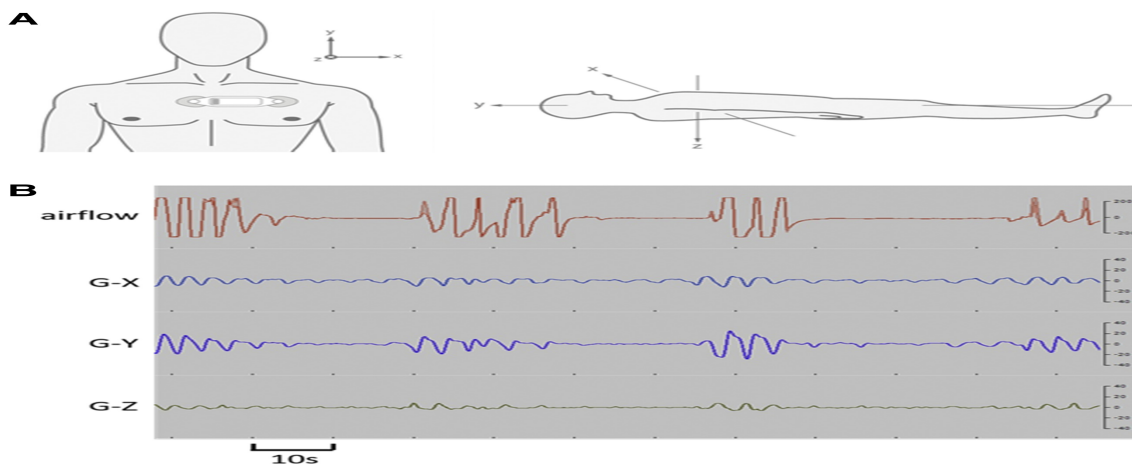


Figure 2.6: chest wall motion during respiratory cycles and sleep apnea events.[16].

The comprehensive nature of PSG, with its various components, offers a detailed view of the physiological processes occurring during sleep. This detailed recording and analysis make PSG an invaluable tool in diagnosing a wide range of sleep disorders, providing essential information for targeted treatments and interventions. As technology advances, we can anticipate more sophisticated and patient-friendly enhancements to this crucial diagnostic method.

2.3.2 Procedure and Setting

PSG is typically conducted in a sleep laboratory, where a patient is monitored overnight. The environment is designed to be comfortable and conducive to sleep. Prior to sleep, a technician attaches the various sensors to the patient's body. The process is non-invasive, although it can be unfamiliar and slightly uncomfortable for some patients.

2.3.3 Data Analysis and Interpretation:

The data collected during a PSG test is extensive and requires expert interpretation. Sleep specialists analyze the various waveforms and physiological signals to identify sleep stages, diagnose sleep disorders, and assess the overall quality of sleep. This analysis is pivotal in formulating a treatment plan for sleep-related conditions.

2.3.4 Limitations of PSG:

While polysomnography (PSG) is considered the gold standard for diagnosing sleep disorders, it is not without its limitations. Understanding these limitations is crucial for researchers and clinicians in sleep medicine to interpret PSG results accurately and explore potential alternative methods.

1. **Environmental and One-Night Snapshot Limitations:** PSG typically requires an overnight stay in a sleep laboratory. This unfamiliar environment can affect the patient's normal sleep patterns, a phenomenon known as the 'first-night effect'. H W Agnew found significant differences in sleep patterns between the first and subsequent nights in a sleep lab, indicating that the first night may not be representative of a person's typical sleep.[17] PSG usually provides data from only one night. Yet, sleep patterns can vary significantly from night to night. This variability can lead to misdiagnosis or an incomplete understanding of the patient's sleep disorder.
2. **Cost and Accessibility Issues:** The cost of a PSG test can be prohibitive for many patients. The expenses associated with the specialized equipment and staff required for PSG make it a costly diagnostic tool. Limited availability of sleep laboratories and trained personnel can make it difficult for patients to access PSG services.
3. **Invasiveness and Comfort Concerns:** The array of electrodes and sensors used in PSG can cause physical discomfort, potentially affecting sleep quality. The invasive nature of the test can lead to anxiety or distress, further impacting sleep.

Despite its limitations, PSG remains a cornerstone in the field of sleep medicine, providing invaluable insights into sleep disorders and physiology. Its comprehensive nature allows for a detailed understanding of sleep patterns, essential for accurate diagnosis and effective

treatment of sleep-related conditions. As technology evolves, there may be advancements in PSG techniques, making it more accessible and less intrusive for patients.

2.4 Literature Review:

In this literature review, we embark on an analytical journey through the complex domain of sleep science, with a specific focus on the delineation of sleep phases-REM and non-REM, and the emerging application of machine learning in this intricate field. The phenomenon of sleep, a fundamental yet multifaceted physiological process, has consistently piqued scientific curiosity and investigation. It offers profound insights into the intricate mechanisms of the human body and its integral relationship with overall health and mental well-being. The early 20th century marked the advent of sleep research, gradually unraveling the layers of this enigmatic state and uncovering its pivotal role in various physiological functions. The identification of distinct sleep phases, each serving specific physiological purposes, has been a cornerstone in sleep research. Non-REM sleep, encompassing lighter stages (N1 and N2) as well as deep slow-wave sleep (N3), is intricately linked with processes such as tissue repair, memory consolidation, and hormonal regulation. Conversely, REM sleep, known for its distinct brain activity patterns and dream occurrences, is essential for cognitive processing and emotional balance. This review aims to dissect the evolution of our understanding of these phases, synthesizing a diverse array of research that underscores their physiological significance. Central to the discourse of sleep science is the challenge of accurately and effectively monitoring sleep patterns. The field, traditionally reliant on polysomnography (PSG), a method both comprehensive and resource-intensive, faces challenges related to accessibility, cost-efficiency, and the contrived nature of sleep laboratory environments. This review intends to undertake a critical analysis of PSG, evaluating its methodologies, applications, and limitations, thereby contextualizing the emergence of more sophisticated and accessible technologies. The advent of machine learning stands as a paradigm shift in sleep research. This technology promises to overcome the constraints of traditional sleep monitoring methods, leveraging complex algorithms to analyze extensive datasets. This review will scrutinize the intersection of machine learning with sleep science, tracing its development from initial application to its present capabilities and future potential. An integral aspect of this technological advancement is the proliferation of wearable devices. These tools,

equipped with continuous, non-intrusive data collection capabilities, have revolutionized sleep monitoring. The integration of machine learning algorithms with data from these devices, offering insights into real-world sleep patterns, has significant implications for both research and clinical practices. This review will dissect how these developments synergize, transforming our approach to sleep analysis. Through a comprehensive examination of a broad spectrum of studies, meta-analyses, and expert opinions, this review aims to critically assess methodologies, findings, and overarching conclusions within the field, while also pinpointing areas that remain underexplored. The objective extends beyond merely summarizing existing knowledge; it seeks to connect historical discoveries with contemporary innovations and prospective advancements, underlining the dynamic and evolving nature of sleep science intertwined with machine learning. Ultimately, this literature review strives to establish a robust foundation for the research questions and hypotheses at the heart of this dissertation. It aims to position this research within the expansive continuum of scientific inquiry into sleep, highlighting its relevance and potential contributions to the field.

2.4.1 "Do Not Sleep on Traditional Machine Learning: Simple and Interpretable Techniques Are Competitive to Deep Learning for Sleep Scoring"

In the realm of sleep science, the advent of machine learning has spurred a myriad of advancements, particularly in sleep stage classification. This paper presents a compelling case for traditional machine learning techniques, emphasizing their competitiveness against deep learning models in sleep scoring. The authors delve into various machine learning algorithms, underscoring the significance of simplicity and interpretability in clinical settings. This perspective challenges the prevalent notion in sleep research that complexity equates to superiority.[18]

The study systematically compares several machine learning models, including simpler linear models and more complex non-linear ones, to deep learning approaches. It evaluates these models based on their accuracy, computational efficiency, and ease of interpretation. The findings reveal that, in many instances, simpler models provide comparable accuracy to their more complex deep-learning counterparts. This revelation is pivotal, as it suggests that simpler models could be more suitable for real-world clinical applications due to their

interpretability and lower computational demands.[18] One of the key strengths of this paper is its comprehensive analysis of the trade-offs between model complexity and performance. The authors thoroughly discuss the implications of using complex models in clinical practice, particularly the challenges in interpreting these models and the potential risks associated with misinterpretation. They advocate for a balanced approach, where the choice of model is tailored to the specific requirements of the task, taking into consideration factors such as data availability, computational resources, and the need for interpretability.[18] The paper also explores the potential of combining traditional and advanced machine learning techniques to create hybrid models that leverage the strengths of both approaches. This innovative idea opens up new avenues for research, suggesting that the future of sleep scoring could lie in the integration of various machine learning methodologies.[18] In conclusion, this paper makes a significant contribution to the field of sleep research by challenging existing paradigms and advocating for a more nuanced approach to the application of machine learning in sleep scoring. Its findings have far-reaching implications, potentially influencing future research directions and clinical practices in sleep medicine.[18]

2.4.2 Large-Scale Automated Sleep Staging

Haoqi Sun's study, "Large-Scale Automated Sleep Staging," represents a significant leap in the application of machine learning to sleep research, specifically addressing the challenge of automating sleep stage classification on a large scale. This paper is particularly notable for its exploration of how machine learning can be applied to vast datasets, a common scenario in contemporary sleep studies, and its implications for both clinical practice and research.[8]

The core of this research lies in its innovative use of advanced machine learning algorithms capable of handling and interpreting extensive and varied sleep data. The authors employ a range of sophisticated techniques, including deep learning models, to analyze sleep stages across a substantial dataset. This approach is critical given the increasing volume of sleep data generated by both clinical studies and wearable technology.[8]

A key contribution of this paper is its demonstration of the scalability and adaptability of machine learning in sleep science. The study shows that with the correct algorithmic approach, it is possible to accurately analyze large volumes of sleep data, a task that would be impractical with traditional manual methods. This scalability is crucial for translating sleep research into practical applications, such as developing more personalized and effective

treatments for sleep disorders.[8]

Sun's methodology includes a meticulous validation process, ensuring that the algorithms are not only effective in large-scale data analysis but also reliable and accurate. The study's findings suggest that machine learning can significantly enhance the efficiency of sleep stage classification, making it a valuable tool for both researchers and clinicians.[8]

The implications of this research are far-reaching. It paves the way for more extensive and detailed sleep studies, potentially leading to better understanding and treatment of complex sleep disorders. Furthermore, the study highlights the potential of machine learning in transforming sleep medicine into a more data-driven and precise field.[8]

In summary, "Large-Scale Automated Sleep Staging" by Haoqi Sun is a pioneering work that expands the horizons of sleep research and machine learning. It provides compelling evidence of the efficacy of machine learning in analyzing large-scale sleep data, marking a significant advancement in the field.[8]

2.4.3 SleepPyCo: Automatic Sleep Scoring with Feature Pyramid and Contrastive Learning

Seongju Lee and colleagues' paper, "SleepPyCo: Automatic Sleep Scoring with Feature Pyramid and Contrastive Learning," marks a significant stride in the field of machine learning applications in sleep research. This study introduces an innovative method for automatic sleep scoring, leveraging the strengths of feature pyramid networks and contrastive learning. The novelty of this approach lies in its ability to enhance the accuracy and efficiency of sleep stage classification, a critical aspect in diagnosing and managing sleep-related disorders.[19]

The research revolves around the development and implementation of the SleepPyCo model, a sophisticated algorithm designed to process and interpret sleep data more effectively. The use of a feature pyramid structure allows the model to capture and integrate information at various scales, an essential aspect given the complex nature of sleep data. Additionally, the incorporation of contrastive learning techniques aids in improving the model's ability to differentiate between various sleep stages, a challenging task in traditional sleep scoring methods.[19]

A notable strength of the SleepPyCo model is its adaptability and precision. The study demonstrates that this model can effectively handle diverse datasets, making it a potentially

valuable tool for large-scale sleep studies. Furthermore, the model's design addresses some of the key limitations of previous machine learning approaches, such as the need for large amounts of labeled data and the challenge of generalizing across different patient populations.[19]

The findings of Lee et al. are promising, indicating that the SleepPyCo model could significantly refine the process of sleep scoring. This advancement could lead to more accurate diagnoses of sleep disorders and, consequently, more effective treatment plans. Moreover, the study opens up new possibilities for further research, particularly in exploring the integration of various machine learning techniques to enhance the analysis of sleep data.[19]

In summary, "SleepPyCo: Automatic Sleep Scoring with Feature Pyramid and Contrastive Learning" contributes an innovative methodology to the field of sleep research. The paper stands as a testament to the potential of advanced machine learning techniques in transforming the way sleep data is analyzed and interpreted.[19]

2.4.4 XSleepNet: Multi-View Sequential Model for Automatic Sleep Staging

Huy Phan's "XSleepNet: Multi-View Sequential Model for Automatic Sleep Staging" presents a groundbreaking approach to sleep staging, introducing a multi-view sequential model that significantly advances the application of machine learning in sleep research. This paper is particularly noteworthy for its exploration of a sequential modeling approach, addressing the dynamic and complex nature of sleep patterns.[9]

The XSleepNet model distinguishes itself through its multi-view framework, which allows for the integration and analysis of various types of sleep-related data. This approach acknowledges the multifaceted nature of sleep and the need for a comprehensive analysis to accurately classify sleep stages. The sequential aspect of the model is designed to capture the temporal dependencies in sleep data, a critical factor often overlooked in traditional sleep staging methods.[9]

A significant contribution of this study is its technical innovation. The XSleepNet model employs advanced machine learning techniques, including deep learning and sequence modeling, to achieve a high level of accuracy in sleep stage classification. The model's

performance is rigorously evaluated, demonstrating its superiority over existing methods in terms of both accuracy and efficiency.[9]

The implications of Phan's research extend beyond the realm of academic study. The XSleepNet model has the potential to revolutionize clinical practice in sleep medicine, offering a more precise and efficient tool for diagnosing sleep disorders. Additionally, the model's design paves the way for future research exploring the integration of diverse data types in sleep analysis, potentially leading to more holistic and patient-centric approaches in sleep medicine.[9]

In conclusion, "XSleepNet: Multi-View Sequential Model for Automatic Sleep Staging" by Huy Phan is a pivotal work that significantly enhances the understanding and application of machine learning in sleep research. The study not only demonstrates the feasibility of advanced machine learning techniques in sleep staging but also sets a new benchmark for future developments in the field.[9]

Methodology

In this chapter, we outline the methodology employed for the research conducted as part of this dissertation. This section is essential as it provides the framework for the study, detailing the approaches and techniques used in the application and evaluation of machine learning models for sleep stage classification. The methodology is structured to ensure replicability, accuracy, and validity of the research findings.

3.1 Data-set Selected For the Dissertation

For this research, we utilized data from the 2018 PhysioNet Challenge, a publicly accessible repository that offers a rich collection of sleep-related physiological signals. This data set is instrumental in our investigation into the application of machine learning models for sleep stage classification. PhysioNet, an online resource for complex physiological signals, has been a critical data source for numerous studies in biomedical research. The specific dataset employed in this study is from the 2018 PhysioNet/Computing in Cardiology Challenge. The challenge focused on automatic sleep stage classification, providing a comprehensive dataset suitable for this purpose.[\[21\]](#)

The dataset includes overnight polysomnography (PSG) recordings from a diverse population, encompassing various age groups and both genders. PSG recordings, considered the gold standard in sleep research, offer a multi-dimensional view of sleep by capturing a range of physiological signals. These include electroencephalogram (EEG), electrooculogram

(EOG), electromyogram (EMG), and others. Each recording in the dataset is accompanied by expert-annotated sleep stage labels, following the American Academy of Sleep Medicine (AASM) guidelines.[21]

Table 3.1: Dataset Composition

Component	Description
EEG Recordings	Multiple channels of EEG data capturing brainwave activity throughout the sleep period.
EOG Recordings	Recordings of eye movements, essential for identifying the REM stage of sleep.
EMG Recordings	Muscle activity data, particularly useful for detecting the reduced muscle tone during REM sleep.
Sleep Stage Labels	Expert-annotated labels for each 30-second epoch of the recordings, based on AASM standards.
Patient Demographics	Information on the age and gender of the subjects, contributing to the diversity of the dataset.

The table 3.1 provided categorizes and describes the key components of the data set used from the 2018 PhysioNet Challenge for sleep stage classification. Each row of the table presents a different aspect of the data, highlighting its relevance and role in the context of sleep study. Here's an explanation of each element in the table:

1. **EEG Recordings:** These are recordings of brainwave activity captured using electroencephalography (EEG). EEG is a fundamental component of polysomnography and is critical for sleep analysis as it reveals the electrical activity of the brain during sleep. Different sleep stages are characterized by distinct brainwave patterns, which are essential for accurate sleep stage classification.
2. **EOG Recordings:**These recordings capture eye movements using electrooculography (EOG). EOG is particularly important for identifying the Rapid Eye Movement (REM) stage of sleep.The presence and pattern of eye movements are key indicators of REM sleep, a stage associated with dreaming and distinctive brain activity.
3. **EMG Recordings::** Electromyogram (EMG) recordings measure muscle activity. In

Table 3.2: Signal Characteristics in Sleep Study[21]

Signal Name	Units	Signal Description
SaO2	%	Oxygen saturation
ABD	μV	Electromyography, a measurement of abdominal movement
CHEST	μV	Electromyography, measure of chest movement
Chin1_Chin2	μV	Electromyography, a measure of chin movement
AIRFLOW	μV	A measure of respiratory airflow
ECG	mV	Electrocardiogram, a measure of cardiac activity
E1-M2	μV	Electrooculography, a measure of left eye activity
O2-M1	μV	Electroencephalography, a measure of posterior activity
C4-M1	μV	Electroencephalography, a measure of central activity
C3-M2	μV	Electroencephalography, a measure of central activity
F3-M2	μV	Electroencephalography, a measure of frontal activity
F4-M1	μV	Electroencephalography, a measure of frontal activity
O1-M2	μV	Electroencephalography, a measure of posterior activity

the context of sleep studies, EMG is used to detect muscle tone and movement. EMG data helps in identifying sleep stages, particularly REM sleep, where muscle tone is significantly reduced (a phenomenon known as REM atonia).

4. **Sleep Stage Labels:** Each 30-second segment (epoch) of the recording is annotated with a sleep stage label by sleep experts. These labels follow the American Academy of Sleep Medicine (AASM) standards, which define different stages of sleep (N1, N2, N3, REM, and Wake). These labels serve as the ground truth for training and evaluating machine learning models in sleep stage classification.
5. **Patient Demographics:** This includes basic demographic information about the subjects, such as age and gender. Demographic diversity in the dataset helps ensure that the developed models are applicable and accurate across different age groups and genders, enhancing the generalizability of the research findings.

3.2 Feature Extraction

In this dissertation, a detailed and meticulous approach to feature extraction from polysomnography (PSG) data is employed, which is pivotal for the effective application of machine learning models in sleep stage classification. The process involves several steps, each carefully executed to ensure the highest data integrity and relevance for subsequent analysis.

The process begins with the loading of PSG data, which is achieved through the `load data` function in the script. This function efficiently handles `.mat` files, extracting the necessary signals from selected channels. Given the complexity of sleep data, which encompasses various physiological signals like EEG, EOG, and EMG, careful attention is paid to maintaining the integrity of the data during this stage. The signals are then transposed to align with the format required for feature extraction.

Following this, the `import labels` function is employed to associate each 30-second epoch of the PSG data with expert-annotated sleep stage labels. This step is crucial as it lays the groundwork for a supervised learning approach, where these labels serve as the target outcomes for the machine learning models.

Given the susceptibility of PSG signals to noise and artifacts, the `design filters` function is utilized to create specific filters for different types of signals. This step is crucial to enhance the signal quality and ensure the reliability of the features extracted.

Subsequently, the `filter signals` function applies these filters to the respective signals. This function also normalizes the data, standardizing the mean and variance. This normalization is critical as it mitigates the impact of signal variability between different recordings, ensuring a level playing field for model training and evaluation. The feature extraction phase is bifurcated into two primary components: statistical feature extraction (`extract stat features`) and spectral feature extraction (`extract spec features`).

1. **Statistical Features:** This segment of the process focuses on extracting statistical parameters such as interquartile range, skewness, kurtosis, and standard deviation from the windowed signals. These features provide a quantitative overview of the distribution and variability of the physiological signals, aspects that are informative of different sleep stages.
2. **Spectral Features:** The spectral feature extraction is particularly intricate, involving

Welch's method to compute power spectral density across various frequency bands. These bands (delta, theta, alpha, sigma, beta) are pivotal in sleep science as they correspond to different sleep stages. An innovative element in this phase is the detection of sleep spindles using a continuous wavelet transform approach, which adds a layer of sophistication to the feature set.

Once the features are extracted, they are compiled into a comprehensive dataset. A pandas DataFrame is created for each participant, which includes all the extracted features along with the corresponding sleep stage labels. This dataset forms the foundation upon which machine learning models are trained and validated.

The final step involves exporting this compiled dataset as a CSV file. This file represents the culmination of the preprocessing and feature extraction phases and serves as the input for the machine learning algorithms that follow. The methodology adopted for feature extraction in this study is both rigorous and thorough, ensuring that the PSG data is transformed into a feature-rich format conducive to machine learning analysis. The combination of statistical and spectral features provides a multifaceted view of the sleep data, capturing various physiological dynamics that are essential for accurate sleep stage classification. This comprehensive approach to feature extraction is expected to enhance the predictive performance of the machine learning models, thereby contributing significantly to the field of sleep research.

3.3 Feature Reduction Using PCA

In addition to feature extraction, this research incorporates Principal Component Analysis (PCA) for feature reduction, specifically targeting both time-domain and frequency-domain features extracted from the polysomnography (PSG) data.

3.3.1 PCA Time Domain

Preparation and Memory Mapping

Memory Mapping: Given the large size of the dataset, a memory-mapped file approach is used for efficient data handling. This method enables the processing of large arrays that do not fit entirely into memory, reducing memory consumption and enhancing computational

efficiency. **Initial Setup:** The script sets up critical parameters, including window size, sampling frequency, and channels to process. It also defines a small value, epsilon, used to check the standard deviation of the channels to ensure signal quality.

Data Reshaping and Normalization

Reshaping and Normalization Functions: The script includes functions to reshape data for individual channels and normalize it. Reshaping involves segmenting the continuous signal data into 30-second windows, aligning with standard sleep stage classification practices. Normalization adjusts the data to a common scale without distorting differences in the ranges of values.

Batch Processing for PCA

Batch Processing: To manage the large volume of data, the script processes the data in batches. This approach allows for handling substantial datasets by dividing them into smaller, more manageable segments. **PCA on Each Batch:** Principal Component Analysis is performed on each batch of data. The PCA model is configured to reduce the feature set to a specified number of principal components, capturing the most significant variance in the data while reducing dimensionality.

Aggregation and Averaging of PCA Coefficients

Aggregating PCA Coefficients: After processing each batch, the PCA coefficients (principal components) are aggregated. These coefficients represent the transformed feature space where the dimensions are the principal components that best explain the variance in the data. **Averaging Across Batches:** The PCA coefficients from all batches are averaged to obtain a generalized set of components. This step ensures that the feature reduction is representative of the entire dataset and not biased towards any specific batch.

Output and Storage

DataFrame Creation: The averaged PCA coefficients are stored in a pandas DataFrame for easy access and manipulation. Each column in the DataFrame corresponds to a channel in the PSG data, now represented in the reduced feature space. **Exporting PCA Coefficients:** The

final step involves exporting the averaged PCA coefficients to a CSV file. This file contains the reduced feature set that will be used for training and validating the machine learning models.

The inclusion of PCA in the methodology serves a dual purpose. Firstly, it significantly reduces the dimensionality of the feature space, making the machine learning models more computationally efficient. Secondly, by focusing on the principal components, the models are likely to capture the most relevant aspects of the data, potentially enhancing their predictive accuracy. This approach is particularly advantageous in the context of sleep stage classification, where the high dimensionality of raw PSG data can pose challenges in model training and performance.

3.3.2 PCA Frequency Domain

The methodology for frequency-domain feature reduction in this research involves a sophisticated process of Principal Component Analysis (PCA) applied to the frequency-domain features extracted from polysomnography (PSG) data.

Preparation and Memory Mapping

Memory Mapping: To efficiently manage the large volume of PSG data, a memory-mapped file system is utilized. This allows for handling data that surpasses the available memory, ensuring efficient data processing and access.

Setting Parameters: The script sets essential parameters such as window size, sampling frequency, and the selected channels for analysis. An epsilon value is defined to check the standard deviation of the signals, ensuring the quality and reliability of the data.

Data Reshaping and Normalization for Frequency Analysis

Reshaping for Frequency Analysis: The script includes a function `reshape_onesubj_freq` to reshape and prepare data specifically for frequency-domain analysis. This involves segmenting the continuous signals into 30-second epochs and performing Welch's power spectral density estimation.

Normalization: Each frequency-domain feature is normalized to ensure consistency

across the dataset. This normalization involves applying a logarithmic transformation to the power spectral density estimates, followed by scaling and centering.

Batch Processing and PCA

Processing PSG Data: The script iteratively processes PSG data from each subject, handling cases where signal quality is compromised (indicated by a standard deviation below the epsilon threshold).

PCA Application: Principal Component Analysis is then performed on the collected frequency-domain data. The PCA model reduces the number of features to a specified number, focusing on retaining the most significant components that capture the core variance in the data.

Output and Storage

Storing PCA Scores and Coefficients: The PCA scores (transformed data) and the PCA coefficients (principal components) are stored in pandas DataFrames. The scores represent the data in the reduced feature space, while the coefficients provide insight into the feature contribution to each principal component.

Exporting Results: The final PCA scores and coefficients are exported as CSV files for further analysis. This step ensures that the reduced feature set is readily available for machine learning model training and validation.

The inclusion of PCA in the frequency-domain feature analysis offers several advantages. Firstly, it significantly condenses the feature space, making it more manageable and computationally efficient for machine learning algorithms. Secondly, by concentrating on principal components, the models can focus on the most informative aspects of the frequency-domain data, potentially improving their predictive accuracy. This approach is especially beneficial in sleep stage classification, where the frequency characteristics of physiological signals play a crucial role. Through this methodology, the research aims to capture the essence of the frequency-domain features while reducing complexity and enhancing model performance.

3.4 Machine Learning

In this dissertation, we employ a suite of advanced machine learning algorithms to classify sleep stages based on the features extracted from polysomnography data. The chosen algorithms include XGBoost, Logistic Regression, CatBoost, and Random Forest. Each of these algorithms is well-regarded in the machine learning community for its performance in various classification tasks. The following sections provide an overview and rationale for selecting each of these algorithms.

3.4.1 XGBoost (eXtreme Gradient Boosting)

XGBoost stands out as a highly efficient and scalable implementation of gradient boosting. It has gained popularity due to its performance in numerous machine learning competitions and applications. XGBoost is known for its speed and performance, particularly in handling large and complex datasets, which makes it well-suited for sleep stage classification where the datasets are often substantial and multidimensional.[20].

One of the key features of XGBoost is its ability to handle sparse data and its robustness to overfitting, partly due to the regularization components in its objective function. It also offers flexibility in defining custom optimization objectives and evaluation criteria, making it adaptable to various data characteristics and research requirements. The algorithm's effectiveness in capturing complex patterns in data makes it a strong candidate for analyzing the intricate relationships in sleep data [20].

3.4.2 Logistic Regression

Logistic Regression, despite its simplicity compared to more complex models, remains a staple in statistical modeling and machine learning. It's particularly valued for its interpretability and efficiency, especially in binary classification tasks. In the context of sleep stage classification, Logistic Regression serves as a benchmark to assess the performance of more complex algorithms.[22]

The algorithm models the probability of a binary outcome based on a linear combination of the input features. Its simplicity allows for a clear understanding of the influence of each feature on the predicted outcome, an aspect crucial for clinical interpretability in sleep

research.[22]

3.4.3 CatBoost

CatBoost is a relatively recent addition to the ensemble of tree-based machine learning algorithms. It has been designed to provide high performance with categorical data, although it performs equally well with numerical data. One of the key strengths of CatBoost is its robust handling of categorical features, automatically determining useful representations for them, which can be particularly beneficial in sleep data that often comprises both categorical and continuous variables.[23]

CatBoost's algorithmic advancements reduce overfitting and improve generalization, making it a valuable tool in our sleep stage classification task. Its efficient implementation and support for GPU acceleration also make it suitable for handling large datasets typically associated with sleep studies.[23]

3.4.4 Random Forest

Random Forest is an ensemble learning method known for its high accuracy, robustness, and ease of use. It operates by constructing a multitude of decision trees during training and outputting the class that is the mode of the classes of the individual trees. Random Forest can handle a large number of input variables without variable deletion, which is essential for our research given the multidimensional nature of PSG data.[24]

A significant advantage of Random Forest in sleep stage classification is its ability to reduce overfitting while maintaining the accuracy of large datasets. It also offers insights into the importance of different features for classification, aiding in understanding the underlying patterns in sleep data.[24]

The selection of these diverse machine learning algorithms aims to provide a comprehensive analysis of the extracted features from PSG data. By comparing the performance of these different models, we can identify the most effective approach for classifying sleep stages, considering both accuracy and interpretability, which are critical in clinical applications.

3.5 Model Training and Validation Pipeline

This section of the dissertation elaborates on the methodical approach adopted for the training and validation of various machine learning algorithms, including XGBoost, Logistic Regression, CatBoost, and Random Forest, in the context of sleep stage classification. This pipeline is meticulously designed to ensure a robust assessment of each model's performance, accounting for the nuances and complexities inherent in sleep data.

3.5.1 Overview of the Pipeline

The pipeline is built on a foundation of rigorous data preparation, cross-validation, and performance evaluation. Its flexibility allows for the interchangeable use of different machine learning algorithms while maintaining a consistent framework for training and assessment.

Data Preparation and Sampling

Participant Selection: The dataset encompasses a diverse range of participants. For each iteration of model training and validation, a random sample of participant IDs is selected. This randomization is crucial to prevent model bias and ensure the generalizability of the findings.

Data Segregation: The complete dataset is divided into subsets based on the chosen participant IDs, creating a focused dataset for each iteration. This step is vital in maintaining data integrity and avoiding information leakage between the training and testing phases.

Train-Test Split and Preprocessing

Splitting Strategy: A 70-30 train-test split is employed, wherein 70% of the data is used for training the model, and the remaining 30% is reserved for testing. This ratio is chosen to provide a substantial training dataset while ensuring enough data is available for an accurate assessment of model performance.

Data Preprocessing: The features (X) and labels (y) are extracted from the dataset. The features undergo a preprocessing step where missing values are handled (filled with zeros), and label values are adjusted appropriately. This preprocessing is crucial for model compatibility and to ensure accurate predictions.

Feature Scaling: The features are scaled using the StandardScaler, normalizing them to have a mean of zero and a standard deviation of one. Scaling is critical, particularly in datasets with features of varying magnitudes, to prevent any single feature from dominating the learning process.

Cross-Validation and Model Training

KFold Cross-Validation: To validate the model's effectiveness and its generalizability, a KFold cross-validation with 5 splits is implemented. This approach ensures that each subset of the data is used both for training and testing, thereby providing a comprehensive evaluation of the model's performance. **Iterative Model Training:** In each fold, the model is trained on the scaled training data. The use of multiple machine learning algorithms within the same pipeline underscores the versatility of the approach. Each algorithm, with its unique strengths and characteristics, is evaluated under the same rigorous conditions.

Performance Metrics

Accuracy: The accuracy metric assesses the overall correctness of the model predictions. It is a fundamental metric but is complemented by more nuanced metrics to provide a fuller picture of model performance.

F1-Score: The F1-Score is particularly useful in imbalanced datasets. It combines precision and recall into a single metric, balancing the trade-off between false positives and false negatives.

Recall and Precision: These metrics provide insights into the model's ability to correctly classify positive cases (recall) and the proportion of true positive predictions (precision). They are critical in medical applications like sleep stage classification, where missing or incorrectly labeling a stage can have significant implications.

Aggregation and Analysis of Results

Mean Metric Calculation: Post each cross-validation fold, the mean of each performance metric is calculated. This mean provides a consolidated view of the model's performance in that particular fold.

Overall Performance Assessment: The pipeline is repeated for multiple iterations, with

each iteration involving a different set of randomly selected participants. The aggregation of results across these iterations offers a robust estimate of the model's performance, mitigating the impact of any anomalies or biases present in a single iteration.

Comparative Analysis: The final step involves a comparative analysis of the performance metrics across different machine learning models. This comparison is instrumental in identifying the most effective model for sleep stage classification, taking into account factors like accuracy, reliability, and computational efficiency.

This comprehensive training and validation pipeline serves as the backbone of the research methodology. Its systematic approach, encompassing data preparation, model training, cross-validation, and performance evaluation, ensures that the findings are reliable and replicable. The pipeline's adaptability to various machine learning algorithms makes it a valuable framework, not only for this research but also as a reference for future studies in sleep stage classification and related areas.

3.6 Results

In this section of my dissertation, I present a detailed analysis of the performance of four machine learning models - XGBoost, CatBoost, SVM, and Logistic Regression - in the context of sleep stage classification. The evaluation is conducted under two distinct experimental conditions: one with a 4-channel feature extraction followed by PCA, and the other utilizing a more extensive 13-channel feature extraction without PCA. The results of these experiments provide crucial insights into the efficacy of these algorithms in handling complex sleep data.

Table 3.3: Model Performance Scores(4 Channels Feature Extraction + PCA)

Model	Accuracy (%)
XGBoost	80.19
CatBoost	80.66
SVM	80.04
Logistic Regression	74.30

Table 3.4: Model Performance Scores(13 Channels Feature Extraction)

Model	Accuracy%	F1-Score	Recall	Precision
XGBoost	81.66	0.8195	0.7951	0.8568
CatBoost	80.59	0.8027	0.7776	0.8413
SVM	81.66	0.8195	0.7951	0.8568
Logistic Regression	49.26	0.3164	0.3104	0.4659

3.7 Experiment 1: 4 Channels Feature Extraction + PCA (Table 3.3)

3.7.1 Model Performance:

XGBoost: Exhibited an accuracy of 80.19%. This high accuracy reflects XGBoost’s robustness in handling dimensionally reduced data and its efficiency in managing the nuances of sleep stages.

CatBoost: Surpassed XGBoost slightly with an accuracy of 80.66%. This marginally higher accuracy could be attributed to CatBoost’s handling of categorical features, which might be playing a significant role even in this reduced feature set.

SVM: Showed a competitive accuracy of 80.04%. SVM’s performance is particularly noteworthy, given its relative simplicity compared to the ensemble methods. This suggests its capability in capturing essential patterns in the sleep data even after dimensionality reduction.

Logistic Regression: Lagged behind with an accuracy of 74.30%. This lower performance could be due to Logistic Regression’s linear nature, which might struggle with the complexity embedded in the sleep data, even in its dimensionally reduced form.

3.7.2 Analysis:

The results from this experiment highlight the strengths of ensemble methods in sleep stage classification, especially in scenarios where feature dimensionality is reduced. The slight edge of CatBoost over XGBoost could be indicative of its specialized handling of the dataset’s inherent characteristics. SVM’s comparable performance suggests its potential as a simpler

yet effective alternative. Logistic Regression's lower accuracy points to its limitations in dealing with reduced feature sets in complex classification tasks like sleep stage classification.

3.8 Experiment 2: 13 Channels Feature Extraction (Table 3.4)

3.8.1 Model Performance:

XGBoost and SVM: Both models achieved an accuracy of 81.66%, with an F1-Score of 81.95%, Recall of 79.51%, and Precision of 85.68%. These results are indicative of their effectiveness in handling multi-dimensional data, capturing intricate patterns in the more extensive feature set derived from the 13-channel data.

CatBoost: Showed a balanced performance with an accuracy of 80.59%, F1-Score of 80.27%, Recall of 77.76%, and Precision of 84.13%. CatBoost's performance, while slightly lower than XGBoost and SVM, demonstrates its competency in managing complex data structures.

Logistic Regression: Significantly underperformed with an accuracy of 49.26%, F1-Score of 31.64%, Recall of 31.04%, and Precision of 46.59%. This substantial drop in performance across all metrics suggests that Logistic Regression is less suited for high-dimensional data, likely due to its inability to model complex, non-linear relationships inherent in such data.

3.8.2 Analysis:

The results from the 13-channel feature extraction experiment underscore the capability of ensemble methods and SVM in handling high-dimensional data, evident from their high scores across all metrics. The consistency of XGBoost and SVM in both accuracy and other metrics indicates their robustness and reliability. CatBoost's balanced performance across metrics makes it a viable option, particularly in scenarios where trade-offs between different metrics are considered. The stark underperformance of Logistic Regression highlights its limitations in contexts involving complex, multi-dimensional data.

3.9 Comparative Analysis

Comparing the results from both experiments, it is evident that the ensemble methods and SVM consistently outperform Logistic Regression, irrespective of the feature extraction

method employed. The use of PCA in the 4-channel data did not significantly alter the relative performance of the models compared to the 13-channel data. This suggests that the additional channels in the 13-channel setup provide valuable information that is not fully captured in the reduced feature set post-PCA.

Conclusions

This dissertation has provided an in-depth exploration of the application of various machine learning models in the classification of sleep stages. Through a methodical approach to data preparation, feature extraction, model training, and validation, the research has sought to determine the most effective techniques for interpreting the complex data derived from polysomnography.

The experimental results indicate that ensemble methods, such as XGBoost and CatBoost, as well as SVM, are particularly effective in sleep stage classification. These models demonstrate robustness and high accuracy in both the reduced feature set post-PCA and the more comprehensive multi-channel feature extraction. Logistic Regression, while offering benefits in terms of simplicity and interpretability, shows limitations in handling the complexity and high dimensionality inherent in sleep data.

One of the key findings of this research is the importance of feature selection and extraction in sleep stage classification. The comparison between the 4-channel reduced feature set and the 13-channel comprehensive feature extraction illustrates that while PCA is effective in reducing dimensionality and noise, the additional information captured in multi-channel data can be crucial for accurate classification.

The research also highlights the critical role of proper data preprocessing and cross-validation in developing reliable and generalizable machine learning models. By employing a systematic approach to data scaling, splitting, and model evaluation, the study ensures the robustness of its findings.

This dissertation contributes to the broader understanding of how machine learning can be applied in the field of biomedical signal processing, particularly in sleep medicine. The insights gained from this study could have significant implications for clinical practices, potentially leading to more accurate and efficient methods for diagnosing sleep disorders.

Moreover, the methodologies and findings of this research provide a foundation for future studies in this area. Future research could explore the integration of additional physiological signals, the application of newer or more complex machine learning models, and the development of real-time sleep stage classification systems.

In conclusion, this research underscores the potential of machine learning in enhancing our understanding of sleep patterns and disorders. The findings of this study contribute valuable knowledge to the field and pave the way for further innovations in sleep stage classification using machine learning techniques.

Bibliography

- [1] R.B. Berry, R. Budhiraja, D.J. Gottlieb, et al. Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine. *J Clin Sleep Med*, 8(5):597–619, 2012. Published 2012 Oct 15. doi:10.5664/jcsm.2172
- [2] M. Steriade. Grouping of brain rhythms in corticothalamic systems. *Neuroscience*, 137(4):1087–1106, 2006. doi:10.1016/j.neuroscience.2005.10.029
- [3] E. Van Cauter, R. Leproult, and L. Plat. Age-related changes in slow wave sleep and REM sleep and relationship with growth hormone and cortisol levels in healthy men. *JAMA*, 284(7):861–868, 2000. doi:10.1001/jama.284.7.861
- [4] E. Aserinsky and N. Kleitman. Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science*, 118(3062):273–274, 1953. doi:10.1126/science.118.3062.273
- [5] J.A. Hobson, E.F. Pace-Schott, and R. Stickgold. Dreaming and the brain: toward a cognitive neuroscience of conscious states. [published correction appears in *Behav Brain Sci* 2001 Jun;24(3):575]. *Behav Brain Sci*, 23(6):793–1121, 2000. doi:10.1017/s0140525x00003976
- [6] J.M. Siegel. Clues to the functions of mammalian sleep. *Nature*, 437(7063):1264–1271, 2005. doi:10.1038/nature04285
- [7] V.K. Somers, D.P. White, R. Amin, et al. Sleep apnea and cardiovascular disease: an American Heart Association/american College Of Cardiology Foundation Scientific Statement from the American Heart Association Council for High Blood Pressure Research Professional Education Committee, Council on Clinical Cardiology, Stroke Coun-

- cil, and Council On Cardiovascular Nursing. In collaboration with the National Heart, Lung, and Blood Institute National Center on Sleep Disorders Research (National Institutes of Health). [published correction appears in *Circulation*. 2009 Mar 31;119(12):e380]. *Circulation*, 118(10):1080–1111, 2008. doi:10.1161/CIRCULATIONAHA.107.189375
- [8] Haoqi Sun, Jian Jia, Balaji Goparaju, Guang-Bin Huang, Olga Sourina, Matt Travis Bianchi, M Brandon Westover. Large-Scale Automated Sleep Staging. Affiliations: Energy Research Institute @ NTU, Interdisciplinary Graduate School, Nanyang Technological University, Singapore; Fraunhofer IDM @ NTU, Nanyang Technological University, Singapore; School of Mathematics, Northwest University, Xi'an, Shaanxi, China; Massachusetts General Hospital, Neurology Department, Boston, MA; School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. *Sleep*, 2017. DOI: [10.1093/sleep/zsx139](https://doi.org/10.1093/sleep/zsx139) PMID: 29029305 PMCID: PMC6251659
- [9] Huy Phan, Oliver Y. ChÃ©n, Minh C. Tran, Philipp Koch, Alfred Mertins, and Maarten De Vos. XSleepNet: Multi-View Sequential Model for Automatic Sleep Staging.
- [10] Wikipedia contributors. "Sleep cycle." *Wikipedia, The Free Encyclopedia*, Wikipedia, The Free Encyclopedia, 25 Nov. 2023. Web. 1 December 2023. URL: https://en.wikipedia.org/wiki/Sleep_cycle
- [11] HelpGuide.org. "Stages of Sleep: REM Sleep, Deep Sleep, and Circadian Rhythm." *HelpGuide*, in collaboration with Harvard Health, Web. 1 December 2023.. URL: <https://www.helpguide.org/harvard/biology-of-sleep-circadian-rhythms-sleep-stages.htm>
- [12] J. Heo, H. Yoon, and K.S. Park. A Novel Wearable Forehead EOG Measurement System for Human Computer Interfaces. *Sensors (Basel, Switzerland)*, 17(7):1485, 2017. <https://doi.org/10.3390/s17071485>
- [13] CSMA Clinic. EMG electrode position. *Comprehensive Sleep Medicine Associates*, Web. [Access date]. URL: <https://csma.clinic/HTML/EMG.htm>
- [14] Carmina Coronel. Measurement of respiratory effort in sleep by 3D camera and respiratory inductance plethysmography. *Sensors (Basel, Switzerland)*, 17(7):1485, 2017. <https://doi.org/10.3390/s17071485>

- [15] Riccardo Pernice, Tatjana Loncar-Turukalo, Gorana Mijatović. Assessment of Cardiorespiratory Interactions during Apneic Events in Sleep via Fuzzy Kernel Measures of Information Dynamics. *Entropy*, 23(6):698, May 2021. DOI: [10.3390/e23060698](https://doi.org/10.3390/e23060698) License: CC BY 4.0
- [16] Ying-Shuo Hsu, Tien-Yu Chen, Dean Wu, Chia-Mo Lin, Jer-Nan Juang, Wen-Te Liu, et al. Screening of obstructive sleep apnea in patients who snore using a patch-type device with electrocardiogram and 3-axis accelerometer. *J Clin Sleep Med*, 2020. DOI: [10.5664/jcsm.8462](https://doi.org/10.5664/jcsm.8462) PMID: 32267228
- [17] H.W. Agnew Jr, W.B. Webb, and R.L. Williams. The first night effect: an EEG study of sleep. *Psychophysiology*, 2(3):263–266, 1966. DOI: [10.1111/j.1469-8986.1966.tb02650.x](https://doi.org/10.1111/j.1469-8986.1966.tb02650.x)
- [18] J. Van Der Donckt, J. Van Der Donckt, E. Deprost, N. Vandenbussche, M. Rademaker, G. Vandewiele, S. Van Hoecke. Do Not Sleep on Linear Models: Simple and Interpretable Techniques Outperform Deep Learning for Sleep Scoring. URL: <https://ar5iv.org/abs/2207.07753>
- [19] Seongju Lee, Yeonguk Yu, Seunghyeok Back, Hogeon Seo, and Kyoobin Lee. SleepPyCo: Automatic Sleep Scoring with Feature Pyramid and Contrastive Learning.
- [20] T. Chen and C. Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.
- [21] PhysioNet. PhysioNet/CinC Challenge 2018: Training/Test Sets. Retrieved [Access Date], from <https://archive.physionet.org/pn3/challenge/2018/>
- [22] D.W. Hosmer Jr, S. Lemeshow, and R.X. Sturdivant. *Applied Logistic Regression*. John Wiley & Sons, 2013.
- [23] L. Prokhorenkova, G. Gusev, A. Vorobev, A.V. Dorogush, and A. Gulin. CatBoost: unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems*, 2018.
- [24] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.