Universidade do Minho

Escola de Engenharia

Departamento de Informática

António José Almeida Rebelo

Sleep stage classification based on cardiorespiratory signals

Dissertation of Master's Degree

Master's Degree in Informatics Engineering

Work done over the orientation of Paulo Novais and Pedro Fonseca

June 2013

The reproduction of this thesis integral only for research purposes by written declaration of the person concerned, that such commits.

Acknowledgments

I would like to thank Pedro Fonseca for giving me the opportunity to come to the Netherlands and to work within Philips Research allowing me to live a new experience and to grow and build up professionally and personally. I am also thankful for the patience, advice and guidance given throughout the year.

To professor Paulo Novais for the support, availability and guidance given along this year.

To my friends and colleagues from Eindhoven and Portugal for the friendship, support and the leisure activities.

To my family for all the support and care given throughout my whole university years.

Finally, to Ana, who helped me in all moments giving me companionship and support.

Title

Sleep stage classification based on cardiorespiratory signals.

Resumo

O sono está ligado a uma quantidade bastante considerável de patologias que têm impacto direto na maioria das atividades diárias tais como a aprendizagem, memorização ou produtividade. Assim, reduzir as consequências pessoais e os custos associados com os distúrbios do sono tornou-se num dos maiores desafios das últimas décadas. Patologias como distúrbios respiratórios do sono, sonolência, síndrome de pernas inquietas ou distúrbios do sono relacionados com o ritmo circadiano são bastante prevalentes, produzindo grandes distúrbios no dia-a-dia dos pacientes. Para o diagnóstico e tratamento deste tipo de patologias, a capacidade de avaliar o padrão de sono do paciente por períodos de tempo mais alargados poderá ser necessária. A necessidade de avaliação de um determinado medicamento ou a monitorização da qualidade do sono do paciente ao longo do tempo são bons exemplos. O teste clínico PSG, atualmente padrão para a avaliação do sono, é um método caro e complexo, disponíveis apenas em hospitais especializados e equipados com um laboratório do sono e profissionais qualificados. Para além de nem sempre estar disponível, PSG é considerado um procedimento muito penoso devido aos diversos elétrodos em contacto com o corpo e cabeça, que causam desconforto e possivelmente um padrão de sono anormal. Para além destes incómodos, os pacientes têm ainda que dormir num laboratório, sendo continuamente observados ao longo da noite. PSG é, portanto, uma técnica cara, geralmente limitada a uma ou duas noites num laboratório do sono. Métodos como *actigraphy*, que utilizam sensores semelhantes a relógios de pulso para medir os movimentos corporais dos pacientes, podem dar informações úteis sobre os padrões de sono dos indivíduos durante períodos de tempo mais alargados sem perturbar significativamente os hábitos normais de sono dos pacientes. No entanto, este método tem várias limitações, uma vez que apenas avalia movimentos corporais, o que é insuficiente para informações relativas à arquitetura do sono dos pacientes.

Para ultrapassar as limitações dos métodos acima descritos, seria relevante a criação de um novo procedimento capaz de complementar os já existentes. Um sistema de monitorização do sono baseado em informação cardiorrespiratória poderá fornecer mais informação sobre a arquitetura do sono, de forma não intrusiva e durante períodos de tempo alargados, no conforto e privacidade da residência dos pacientes. Esta informação poderia

ser utilizada para o rastreio de doenças, acompanhamento e monitorização de tratamentos ou mesmo complementar o PSG para o diagnóstico de algumas doenças do sono.

O sistema apresentado neste trabalho aborda parte desta hipótese, classificando automaticamente várias fases do sono usando apenas informação cardiorrespiratória. Embora os dados utilizados para este estudo, tenham sido adquiridos através do uso de sensores de contacto, no futuro, esta informação poderá ser obtida através da utilização de métodos não intrusivos, que já se encontram disponíveis comercialmente. Esta hipótese é bastante interessante porque consegue fornecer mais informação aos profissionais do sono, sem interferir com o dia-a-dia do paciente.

Abstract

Sleep pathologies have a direct negative impact into most of daily activities such has learning, memorization or productivity. Decreasing the personal burden and the societal cost associated with sleep disturbances has become one of the major challenges in the last decades. Pathologies like sleep disordered breathing, insomnia, restless leg syndrome or circadian rhythm sleep disorders are fairly prevalent, heavily disturbing the life of affected subjects. For the diagnosis and treatment of these disorders, the ability to assess a patient's sleep pattern over longer periods of time may be required. The need of evaluation of a certain medication or the monitoring of the sleep quality of the patient over time can be named as good examples. The polysomonographic (PSG) clinical test, current gold standard for sleep assessment, is an expensive and complex method only available in specialized hospitals equipped with a sleep lab and qualified professionals. Not always available, PSG is considered a very stressful procedure because of the various electrodes attached to the body and head, which cause discomfort and potentially disrupt the usual sleep patterns. Furthermore people need to sleep in an unfamiliar environment while being observed throughout the entire night. PSG is therefore an expensive technique usually limited to one or two nights in a sleep laboratory. Methods like actigraphy, which measure body movements, can give useful insight about the sleeping patterns of the subjects during longer periods of time without significantly disrupting the normal sleeping habits of a person. However this method has several limitations as it only assesses the movements of the patients and therefore provides little insight about the subjects' sleep architecture.

In order to address the shortcomings of the existing techniques, the introduction of a new system, easy and cheap to deploy and use, capable of complementing the existent approaches is relevant. A sleep monitoring system based on cardiorespiratory data may be able to provide bigger insight of the sleep architecture, while having the potential to be unobtrusive and able to monitor sleep during longer periods of time, in the comfort and privacy of the subject's own room. Furthermore it can potentially enable the screening of diseases, follow-up on treatments, or even complementing PSG for diagnosis of some sleep disorders.

The system presented in this work addresses part of this system by automatically classifying multiple sleep stages using cardiorespiratory information. Although the data used for this study was acquired with contact sensors, in the future, this information might be obtained through the use of non-obtrusive methods that are already commercially available. This possibility is interesting as it provides bigger insight of the subject sleeping patterns and architecture for the sleep professional, without interfering with the daily life of the patient.

Contents

List of figures

List of tables

Glossary

AAL	Ambient Assisted Living
AI	Artificial Intelligence
ICT	Information and Communication Technologies
NREM	Non Rapid Eye Movement
REM	Rapid Eye Movement
PSG	Polysomnography
EEG	Electroencephalogram
ECG	Electrocardiogram
EOG	Electrooculogram
EMG	Electromyogram
AASM	American Academy of Sleep Medicine
SWS	Slow Wave Sleep
WASO	Wake After Sleep Onset
OSA	Obstructive Sleep Apnea
SDB	Sleep Disorder Breathing
OSAS	Obstructive Sleep Apnea Syndrome
ASMD	Absolute Standardize Mean Difference
PDF	Probability Density Function
CFS	Correlation-based Feature Selection

1 Introduction

Modern lifestyles introduced a wide variety of factors capable of perturbing the ordinary habits of the population. Stress, overworking and a fast paced lifestyle made sleep deprivation and fatigue a common problem and hence the source of many accidents, depressions, and health related problems. Although sleep is increasingly recognized as important to public health, it is essential to realize that sleep deprivation is very often due to undiagnosed sleep disorders. After a disturbed night of sleep, a person might not feel restored and refreshed and be sleepy during the day, but be totally unaware that is sleep-deprived or has a sleep disorder. On the other hand existing diagnosis are expensive, obtrusive, limited to one or two night, and require a patient to sleep in an unfamiliar environment. The solution to this problem may reside in new technologies in the area of ambient assisted living that may provide new solutions for a better and cheaper care providing.

1.1 Ambient assisted living

The continuous increase of older population in Europe and worldwide, associated with increased costs of healthcare favor the deployment of AAL intelligent systems for a better, healthier and safer life in the preferred living environment. AAL comprises concepts, products and services that interlink and improve new technologies and the social environment, with a focus on older people.

AAL relies greatly on Ambient Intelligence (AmI) to provide seamless and unobtrusive interaction in the human environment, thus radically moving away from more traditional assistive technologies towards universal access.

Ambient Intelligence is a relatively new field of Artificial Intelligence, in which computers interaction is made in a more natural way since it's made using friendly interfaces such as gestures. The underlying goal of Ambient Intelligence is to involve a wide variety of different technologies, hiding their presence from users or soothingly integrate them within the surrounding context as augmented physical objects, rather than technological

gadgets. This methodology makes ICT accessible and usable for the largest possible population and takes into account the requirements of older users.

The AAL domains range from the AAL for home and mobile support, including AAL for health, rehabilitation and care, personal and home safety and security, AAL in the community, addressing social inclusion, entertainment and mobility and lastly AAL at work, addressing the needs of older people in the workplace.

1.2 AAL for health, rehabilitation and care

In this work we will be focusing on AAL for health, rehabilitation and care, specifically home care monitoring systems, which are intelligent technologies capable of monitoring home inhabitants in their daily activities, and thus preventing health and security risks or alerting family members or healthcare providers when specific situations occur. Current efforts in this context address fall detection and prevention, detection of helplessness, as well as the long term vital signals monitoring.

The monitoring of the vital signs during sleep plays an important role in the quality of life of an individual and therefore in the AAL context. The available knowledge has established that sleep serves an important function, as evidence by the rebound of sleep loss and the developmental, functional, and metabolic impairments produced by sleep deprivation. Persons experiencing sleep insufficiency are more likely to suffer from chronic diseases such as hypertension, diabetes, depression, and obesity, as well as from cancer, increased mortality, and reduced quality of life and productivity[1].

The analysis of a survey taken in the United States in 2009, regarding perceived insufficient rest or sleep, and on sleep behavior, determined that among 74,571 adult respondents in 12 states, 35.3% reported less than 7 hours of sleep during a typical 24-hour period, 48.0% reported snoring, 37.9% reported unintentionally falling asleep during the day at least once in the preceding month, and 4.7% reported nodding off or falling asleep while driving at least once in the preceding month [2].

Although sleep is increasingly recognized as important to public health, it is important to realize that sleep deprivation is very often due to undiagnosed sleep disorders. After a typical night's sleep, a person might not feel restored and refreshed and be sleepy during

the day, but be totally unaware that is sleep-deprived or has a sleep disorder. On the other hand existent treatments are expensive, obtrusive, and long-lasting requiring a patient to sleep in a sleep clinic which is a hassle to the patient. These facts favor alternative solutions following the AAL approach, capable of monitoring sleep at home, in an unobtrusive way during extended periods of time.

1.3 Sleep monitoring

Sleep is a natural, periodic and easily reversible state characterized by reduced or absent consciousness and sensory activity as well as inactivity of nearly all voluntary muscles. It's a state of extreme rest observed in most animals. Sleep scientists remain in the delicate position of not knowing why we sleep, however it is accepted that the function of sleep is likely multidimensional and differential depending on the organism's stage of development [3].

Sleep disorders interfere with the normal sleeping pattern of a patient, and sometimes are serious enough to interfere with normal physical, mental and emotional functioning. Inadequate or non-restorative sleep can markedly impair a patient's quality of life [4]. Sleep disordered breathing, insomnia, restless leg syndrome or circadian rhythm disor-ders may need the examination of the patient's sleeping patterns over extended periods of time for a correct diagnosis and treatment. The current gold standard for sleep assessment is the clinical procedure known as polysomnography (PSG). This procedure comprises a comprehensive recording of the bio-physiological changes that occur during sleep, monitoring the brain (EEG), eye movements (EOG), muscle activity or skeletal muscle activation (EMG), heart rhythm (ECG) and breathing functions. Due to the number of functions monitored in a standard PSG, the number of attached electrodes and the overall complexity of the method, PSG is an expensive method, only available in specialized hospitals equipped with a sleep lab and qualified professionals. Moreover, carrying out a PSG test forces the patient to spend the night in a sleeping lab which cause discomfort and potentially an unusual sleeping pattern. The "first night effect" is good example of a phenomenon that may alter the usual sleeping pattern of the patient [5–7].

To overcome some these problems, PSG is sometimes complemented by actigraphy. This technique is a non-invasive method capable of distinguishing rest from activity cycles. Actigraphy sensors are generally watch-shaped devices, which can be worn for several weeks on the wrist of the non-dominant arm. An actigraphy sensor, usually equipped with accelerometers, continually records the gross movements of the body. Actigraphy has been indicated by the American Academy of Sleep Medicine (AASM) as a suitable method to assist in the evaluation of patients with circadian disorders and sleep-wake disturbances, and also to assess response to therapy of circadian disorders and insomnia [8][9]. Although having some advantages over the PSG test, actigraphy is only capable of differentiating between rest – wake, which is insufficient for the diagnosis of certain diseases.

As a form of support for sleep assessment techniques, a sleep diary is usually used, in which patients are asked to report their perceived sleep quality. The patients use subjective measures that may turn out to be hard to interpret. Moreover different people have different opinions about the quality of their sleep when experiencing a night with similar sleep quality. As a consequence data obtained from PSG and data derived from a sleep diary often do not coincide very well. An alternative system with objective measures, capable of supporting long-term and unobtrusive sleep monitoring at home is therefore extremely interesting.

This work has the aim of developing a system for the analysis of cardiorespiratory signals and automatically classify sleeping stages. Through this method, sleep parameters can be objectively analyzed in an unobtrusive manner during long periods of time.

1.4 Scope of the dissertation

Systems currently implementing multiple sleep stage classification based on cardiorespiratory signals are considered to have great potential. This potential comes from the variations in features related to cardiac or respiratory activities, induced by different sleeping stages [10][11]. The use of unobtrusive techniques to retrieve data as well as the use of hardware, which is simple and cheap to deploy and use, instead of specific and expensive reveals interesting and conforms to the standards of AAL. Moreover this technique is promising as it offers objective measures to assess sleep staging.

This investigation took as a starting point the work already developed in the Sleep Monitoring project at Philips Personal Health Solutions department [3][4].

1.5 Objectives

The goal of this research is to provide a method capable of automatically classify multiple sleep stages based on cardiorespiratory data. This method, when integrated in a system capable of monitoring and capturing cardiorespiratory data from a patient, should be able to monitor the patient's sleeping patterns over long periods of time, in the comfort of his house. Although the available data for this study was collected with contact sensors, in the future, these sensors can be substituted with non-obtrusive devices that are already commercially available. This method would be a very good complement to other methods of sleep assessment and help in the diagnose, treatment and monitoring of patients with sleep disturbances.

In this work, the main objective is to further improve the existent technologies and methods for the creation of a multiple sleep stage classifier based on cardiorespiratory data. To address the objectives proposed, the following tasks will be performed:

- Research on sleeping patterns and sleep architecture of a normal healthy adult.
    - The aim is to provide new designs to acquire new information for the classification process.
- Improve data quality and reduce between-subject variability.
    - Research normalization techniques capable of reducing the differences in the features belonging to different subjects.
    - Search for transformations of the features, such as statistical transformations, capable of providing new information for the classification procedure.
- Use an adequate classifier set-up in order to successfully extend the existent work for the multiple class situation
    - Research and apply feature selection algorithms.
    - Define and set up an appropriate classifier, capable of good discrimination between multiple classes.

- Create methods capable of using probabilistic techniques, to further improve classification techniques.

As a way to have an idea of the overall system, as well as the way the proposed objectives and subsystems are going to be organized, Figure 1, shows the architecture of the system that will be constructed.

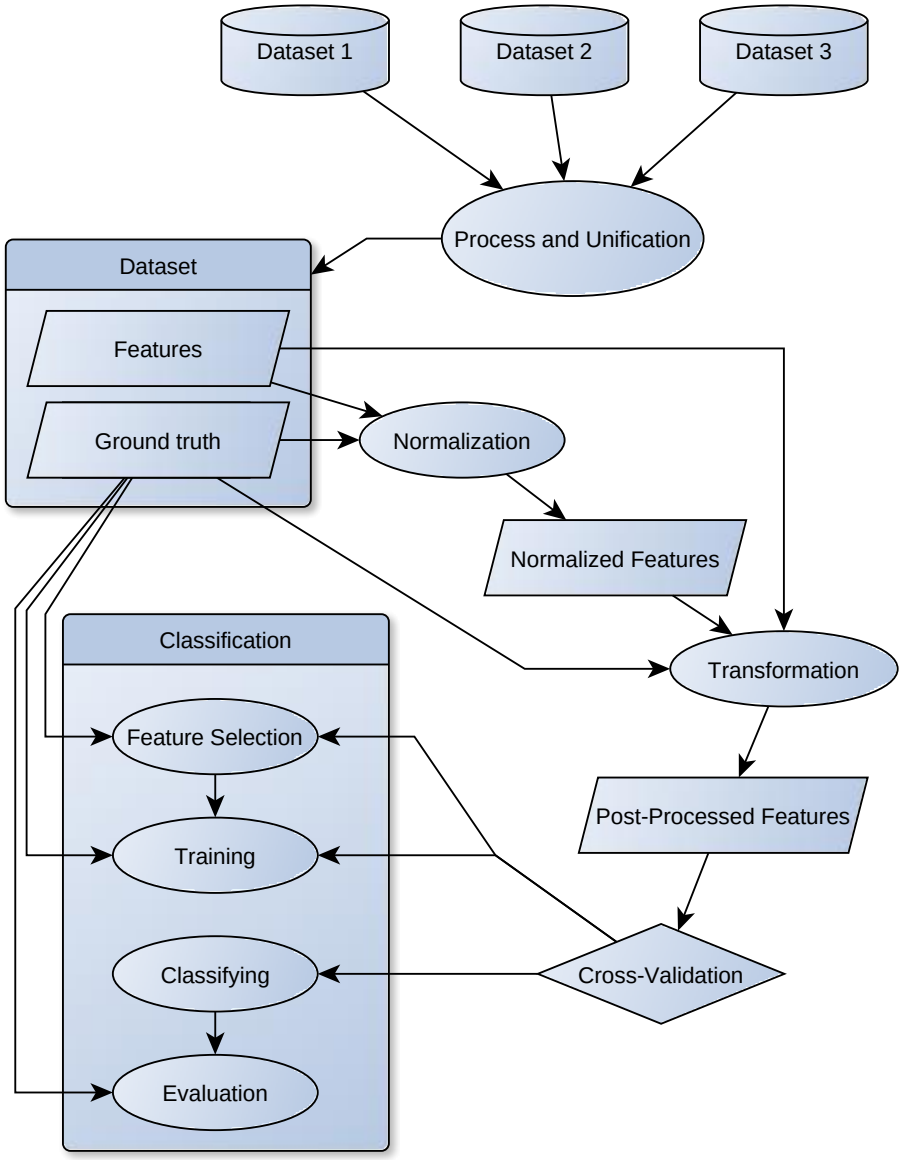


Figure 1.System overview

1.6 Research methodology

This work was developed according to an Action-Research methodology. This method is suitable for solving a problem that seeks to obtain information leading to its resolution through an iterative and recurrent process. This method starts with the identification of the problem so that several hypotheses can be formulated on which development work will be based. Afterwards, the gathered information will be recompiled, structured and analyzed, in order to continuously develop a proposal to solve the identified problem. As an end result, one can make conclusions based on the outcomes obtained during the research.

Using this research model, six complementary steps are defined to achieve the planned objectives:

- Specification of the problem and its characteristics;
- Incremental update and review of the state of the art;
- Idealization of new methods and iterative development of the proposed model;
- Experimentation and implementation of the prototype;
- Results analysis and conclusions;
- Diffusion of knowledge with the scientific community.

1.7 Structure of the document

This document is organized as follows: Section 1 starts with the description the main techniques for sleep assessment focused on the insufficiencies and problems associated with them. After defining the main problem, the scope of the project, followed by the objectives and research methodology are introduced.

Section 2 will begin with a brief introduction to sleep medicine, describing the essential aspects of normal human sleep and some sleep disorders. Some studies related to sleep

assessment are reviewed. Studies that are in the same scope of the research being developed are presented, examined and compared.

In section 3 the datasets used in this work are described. A review and a comparison of the current data with sleep literature is presented as well. The sleep architecture of the subjects of this study will be further analyzed, culminating in the selection of the subjects with a normal healthy sleep architecture.

Section 4 will cover the techniques used to improve the quality of the data and its discriminating power. These techniques will be divided in two distinct procedures, namely normalization and transformation. Normalization aims at reducing between-subject variability while transformation aims at maximizing the number of features with a high discriminative power. The results of the application of this techniques are discussed in this section as well.

Section 5 will describe the classifier, the feature selection algorithm and the probabilistic post-processing step used to improve the classification results. This step aims at capturing the non-stationary temporal characteristics of sleep.

The final results are presented, discussed and compared with literature in section 6.

Finally, section 7 will review the work performed and draw the final conclusions of the present work.

2 Sleep monitoring

2.1 History of sleep medicine

It is difficult to determine or even produce an estimated point in time when interest in sleep first occurred. Insomnia was reported in ancient Egyptian texts and it is thought that opium was used as the first hypnotic medication [5]. Despite early curiosity, the scientific interest in sleep has emerged over the past century and the field of sleep medicine itself has existed for only about five decades. Sleep medicine is devoted to the diagnosis and therapy of sleep disturbances and disorders.

2.1.1 Early theories

In the late nineteenth and early twentieth century, a variety of theories were formed regarding the nature of sleep. A theory gained popularity around the end of the nineteenth century, hypothesizing that toxins were developed during wakefulness and were gradually eliminated during sleep. Legendre and Pieron injected serum from sleep deprived dogs into awake dogs and observed that they became fatigued. The term 'hypnotoxin' was introduced to describe this endogenous sleep factor, which promoted sleep [6].

The development of the electroencephalogram (EEG) in 1929 by the German Psychiatrist, Hans Berger allowed the examination of brain activity during sleep [7]. This measuring technique recorded the electrical activity of the human brain, allowing for a continuously and quantitatively measure of the neural activity of the sleeper without disturbing it. Investigation in the following years established the characteristics of the EEG during sleep. High amplitude, slow waves, and spindles were found to be typical during sleep while wakefulness was characterized by fast and lower amplitude waves and alpha rhythms [8]. Using the EEG it became clear that the brain was not idle and that it actually followed a synchronized pattern of neuronal activity.

Sleep stages were first categorized and described in 1937 by Alfred Lee Loomis and his team, who distinguished the different electroencephalography features of sleep into five levels representing the spectrum from wakefulness to deep sleep [9].

2.1.2 REM sleep

The phenomenon of rapid-eye movement (REM) sleep and its association with dreams was discovered by Eugene Aserinsky, Nathaniel Kleitman and William C. Dement in 1952 at the University of Chicago [10]. The creation of a method which could measure the eye mobility, called electrooculography (EOG), made possible the analysis of the eye mobility across the night. They found that during certain periods in the night, there was a substantial increase in the speed of the movement of the eyes. These periods were therefore called rapid eye movement (REM) sleep [11]. During their investigation, recurring variations were observed, corresponding to cyclical occurrences of REM sleep at intervals of 90 to 100 minutes, with intervals tending to be longer toward the end of night. The total sleep time spent in REM sleep was estimated to be between 20 and 25% of the whole sleep time. The association between REM sleep and dreaming was also established [12].

Meanwhile in 1960, Michel Jouvet demonstrated through electromyographic (EMG) recordings that activity and muscle tone are completely suppressed during REM. Presently it is established that muscle atonia is a fundamental characteristic of REM sleep [13].

Because of these discoveries sleep was therefore reclassified into REM and four non-REM (NREM) stages.

2.1.3 Sleep clinics and sleep disorders

With the discovery of REM sleep, sleep research relying on all-night sleep recordings became standard and was the precursor of sleep medicine and particularly of the clinical test, polysomnography.

The first sleep disorders center was established as a narcolepsy clinic at Stanford University in 1964, evolving into evolved into a full-service sleep disorders clinic by 1970. The sleep center was directed by a sleep specialist, and had the ability to perform polysomnography and multiple sleep latency tests [14].

In 1965, sleep apnea was discovered independently by Gastaut, Tassinari and Duron in France and Jung and Kuhlo in Germany [15][16].

The staging criteria were standardized in 1968 by Allan Rechtschaffen and Anthony Kales in the "R&K sleep scoring manual" [17].

At Stanford University in 1972, respiratory and cardiac measurements became a standard of the all-night diagnostic test. In 1974, Dr. Jerome Holland, a member of the Stanford group named this test ”polysomnography” [18].

Presently people are much more aware of the significance of sleep in their lives as well as the consequences of sleep disorders. However the percentage of undiagnosed sleep disorders is still extremely high.

2.2 Normal human sleep

Healthy sleep is characterized by a consistent and cyclic process in which phases of deep sleep alternate with lighter sleep. Normal human sleep is composed by two major phases of sleep called REM and NREM (Non Rapid Eye movement). These major phases alternate cyclically across the night. In 2007, the former Rechtschaffen and Kales scoring system [28], which was comprised by Wake, Rem and four individual stages of Non-Rem sleep, was revised by the AASM, which resulted in several changes. Arousals and respiratory, cardiac, and movement events were added and stages III and IV were combined into a single stage, stage N3 [28]. Using the AASM definition, NREM sleep is further divided in N1, N2 and N3, increasing depth towards N3 which is sometimes also called slow-wave sleep. When comparing ASSM and Rechtschaffen and Kales scoring, it is acceptable to compare N1 to stage I, N2 with stage II and N3 with stages III and IV.

Presently the current gold standard for sleep assessment is the polysomnography clinical test. As mentioned before, this is a very reliable manual scoring performed by sleep professionals who visually inspect the recordings of brain activity, eye movements and muscle activity in order to determine, for each period of the night, in which sleep stage the subject is. Sleep stages are consequently used to plot a so-called hypnogram, as illustrated in Figure 1. If sleep diseases are suspected, sleep professionals might use additional sensor modalities such as using respiratory flow and respiratory effort to score apnea events, or use video recordings to diagnose REM behavior disorder.

According the AASM guidelines, the night is divided into 30-second epochs which are manually classified as Wake, N1, N2, N3 or REM.

Stage N1 is common in the beginning of the sleeping period, accompanied with slow eye movement and a decline in the tonic activity when compared with the wake state. It is normally a transition phase between wake and sleep. It can also emerge briefly during transitions from sleep to wake or after brief body movements. People are normally unaware of this stage and when aroused they often believe they were still wake. The EEG is characterized by relatively low voltage slow activity in the theta range (4 - 7Hz). N1 is distinguishable from relaxed wake, with closed eyes, since wake during wake higher frequency alpha waves (8 - 13Hz) are generated by the brain [17].

Stage N2, is marked by the appearance of EEG spindles (fast activity in the 7-14 Hz range lasting at least half a second) and K-Complexes, which consist of high-voltage waves with a negative sharp component followed by a positive component. Cardiac and respiratory frequencies are usually slower than during the wake state. Conscious awareness completely vanishes [3].

Stage N3, is the deep sleep stage where slow-wave sleep (SWS) occurs. This stage is scored if delta waves, large amplitude figures with a frequency range of 0.5 - 2Hz, occupy at least 20% of the thirty second epoch in the EEG. These very pronounced waves are much "slower", have a lower frequency, than those characteristic of N1 and N2, explaining why this stage is also called slow-wave sleep. Deep sleep is characterized by lower variations in respiratory and cardiac activity when compared against the other sleep stages. Sleepwalking, sleep-talking or other parasomnias are typically encountered in this stage [29].

REM sleep represents an active form of sleep and is characterized by low amplitude and high frequency cerebral activity, very similar to wake, associated with muscle atonia and rapid eye movements. Both cardiac and respiratory activity are highly variable during REM sleep. Most dreams occur in this phase of sleep. REM sleep can be further divided in tonic and phasic components. Tonic REM sleep is associated with near paralysis of most muscular groups. Only the diaphragm, the cardiac muscle, and some sphincters at

the top and at the bottom of the gastrointestinal tract remain active during REM sleep. Transient swings in blood pressure, heart rate changes, and irregular respiration are associated with tonic REM sleep as well. Phasic REM sleep is characterized by irregular episodes of EMG activity and rapid eye movements [17].

2.2.1 Sleep architecture

In a healthy adult, sleeping on a normal schedule, sleeps follows a certain architecture beginning with NREM, progressing towards N3, before arriving at the first episode of REM sleep approximately 80 to 100 minutes after sleep onset. Afterward NREM sleep alternates with REM sleep with a periodicity of approximately 90 minutes. Across the night the length of the episodes of deep sleep and REM sleep vary. In the first cycles of sleep, N3 episodes are longer than REM sleep episodes. As the nigh progresses, N3 episodes start to be smaller or even absent and REM sleep episodes become longer. In young adults, N1 sleep constitutes about 5-10% of the night. The largest amount of sleep time, 50-60%, is spent in stage N2. Stage N3 constitutes about 10-20% of the total sleeping time while REM sleep 20-25% [19].

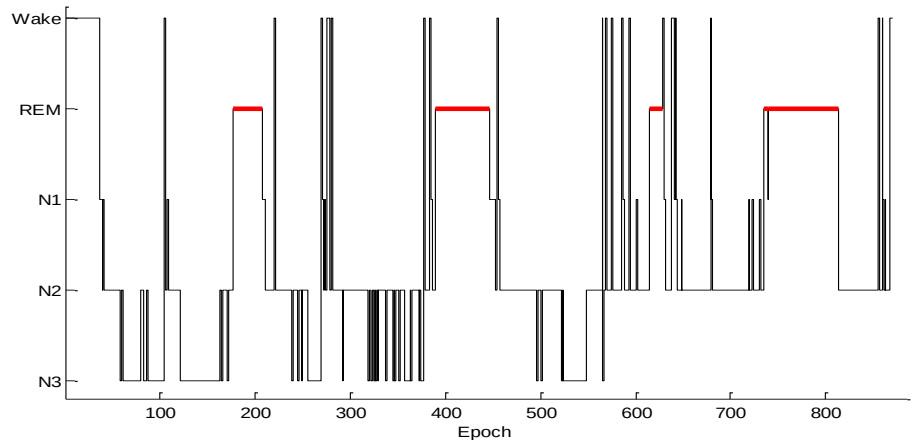


Figure 2. Hypnogram of a healthy adult

Figure 2 illustrates an actual hypnogram of a single night of healthy subject, monitored in a sleep laboratory. Hypnograms were developed as an easy tool to present the sleep architecture during a period of sleep. They allow the easy visualization of the sleep stages, which in turn allow certain parameters which might be indicative of sleep disorders to be

easily identified, for example, high sleep fragmentation or low sleep efficiency, too short REM sleep onset, long sleep onset, low percentage of deep sleep among others. Each epoch correspond to 30 seconds starting from the moment the lights were turned off. In this image it is possible to see the NREM - REM cycles as well as the changes in the duration of these two stages throughout the night.

2.2.2 Changes in sleep with age

In the absence of disorders, the most significant factor affecting total sleep time and sleep stages is age. Newborns have so-called passive and active sleep stages, which are the precursors of NREM and REM sleep respectively. Newborn infants, during the first months of life, sleep 17-18 hours a day, and spend 50% of their sleep time in active sleep. The duration of the passive-active sleep cycle is also shorter in the newborn, when comparing with NREM-REM cycle in adults, at about 50-60 minutes [20]. Also, SWS is maximal in young children and decreases markedly with age. REM sleep and sleep efficiency decrease with age as well.

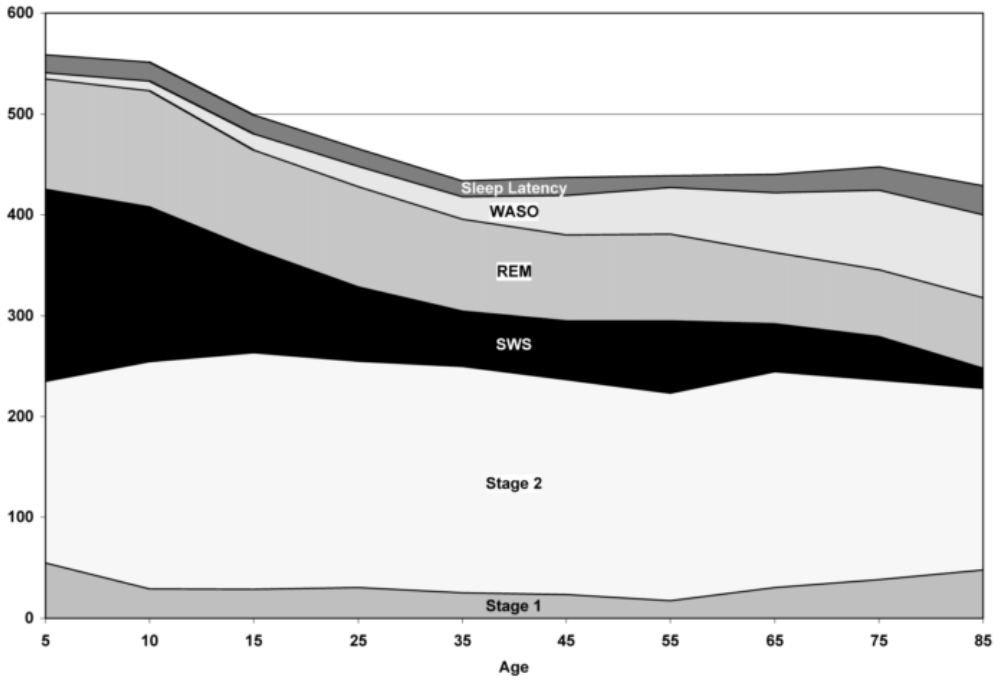


Figure 3. Age related trends for stage 1, stage 2, slow wave sleep (SWS), REM sleep, wake after sleep onset (WASO) and sleep onset latency (minutes) [21]

In Figure 3 it is possible to observe how sleeping time is divided according to the age of the subjects. Sleep onset latency is the length of time that it takes to accomplish the transition from wake to sleep. The changes in sleep onset latency are not very significant, and remain fairly constant as subject's age. WASO is the amount of time spent wake after sleep has been initiated and before final awakening. It is a good metric for measuring the subject's difficulty to stay asleep, a common occurrence in sleep disorders such as insomnia but also a normal occurrence in elderly subjects, since it increases with age.

Sleep efficiency is the ratio of time spent asleep (total sleep time) to the amount of time spent in bed from light out to lights on. Since WASO increases with age, sleep efficiency consequently decreases.

2.3 Sleep disorders

The prevalence of sleep disorders in the general population is considerably high. In fact hundreds of millions of people over the world suffer from sleep disturbances [2]. Sleep disorders can be the cause of impaired academic or occupational performance, accidents at work or while driving and disturbances of mood and social adjustment. Somnolence and the predisposition to fall asleep during the performance of dangerous tasks, is recognized as an important problem in our society. Furthermore, sleep disorders may lead to or aggravate serious medical, neurologic and psychiatric problems. There are some common complaints regarding disturbances in sleep. Patients usually mention problems like insomnia, excessive daytime sleepiness and abnormal movements, behaviors or sensations during sleep or nocturnal awakenings.

Ever since sleep disorders were first accepted, their classification has been of particular interest to clinicians. The Diagnostic Classification of Sleep and Arousal Disorders [22], published in 1979, was the first significant classification and organized sleep disorders in categories which influenced the current classification systems. In order to standardize the sleep disorders and create a systematic approach for their diagnose, the International Classification of Sleep Disorders (ICSD) was created in 1990 and in 2005 it was updated and named ICSD-2 [23]. The sleep disorders were organized into eight distinct categories:

1. Insomnias.
2. Sleep related breathing disorders.
3. Hypersomnias of central origin not due to a circadian rhythm sleep disorder, sleep related breathing disorder or other cause of disturbed nocturnal sleep.
4. Circadian Rhythm Sleep Disorders.
5. Parasomnias.
6. Sleep Related Movement Disorders.
7. Isolated Symptoms, Apparent Normal Variants, and Unresolved Issues.
8. Other Sleep Disorders.

This classification has been widely accepted and used by sleep professionals and has allowed better international communication and cooperation in sleep disorder research.

2.3.1 Insomnia

Insomnia is by far the most common form of sleep disturbances [24]. Usually insomnia is defined as the symptom of difficulty falling asleep or remaining asleep and occasionally as the inability to obtain restorative sleep. Insomnia has a very peculiar characteristic as it is both a symptom and a disorder. It can be secondary to another disorder or an independent disorder. Thus, insomnia may be classified into primary or secondary, with both forms leading to sleep - wake disturbances although the secondary form arises as a function of psychiatric conditions, medical diseases or substance abuse [25]. Insomnia is usually linked with a very active life, high levels of stress, age, shift working, and psychiatric and medical disorders. It is also more prevalent in women. The clinical assessment of insomnia is usually based on a clinical interview with a patient, often supplemented by questionnaires, psychological screening tests, sleep diaries, and interviews with the bed partner. The use of actigraphy in addition to the questionnaires is a common practice before PSG [26]. An efficient assessment and management of insomnia must address psychological and behavioral factors such as poor sleep routines with irregular sleep-wake schedules, hyperarousability, and wrong attitudes and beliefs about sleep. Several treatments are applied in order to address primary insomnia. Examples include sleep re-

striction, stimulus control, cognitive theraphy. Relaxation training and sleep hygiene education are used as well as a combination of those methods, which is referred together as cognitive behavior therapy.

2.3.2 Sleep disordered breathing

These types of sleep disorders are characterized by an irregular and unnatural respiration during sleep. Among this disorders, sleep apnea is by far the most common and well-known. It is associated Sleep apnea can be distinguished between central apnea and obstructive sleep apnea. Central apnea disorders are characterized by diminished or absent respiratory effort intermittently or cyclically as a result of central nervous system dysfunction. Obstructive sleep apnea disorders are formed by an obstruction in the airway resulting in an increased breathing effort and inadequate ventilation. In adults, obstructive sleep apnea is characterized by repetitive episodes of breathing cessation (apneas) or partial upper airway obstruction (hypopneas). The typical consequences are snoring and sleep disruption leading to secondary insomnia or excessive daytime sleepiness [28]. It is also associated with cardiovascular morbidity and hypertension. Although significant prevalent in preschool children and middle age adults, in many countries, up to 90% of the affected subjects remain undiagnosed due to the lack of resources [27–29].

2.4 State-of-the-art in home sleep monitoring

As presented in the previous sections, although sleep disturbances have a significant negative impact on health, Laboratory-based polysomnography (PSG), the gold standard for sleep monitoring, is impractical for long-term and home use. Hence, alternative devices have been developed for home sleep assessment and diagnose of sleep disorders. This section reviews the literature, providing an overview of available projects and commercially available products for sleep monitoring outside the laboratory or the sleep clinic.

2.4.1 Actigraphy

Actigraphy is a non-invasive method that is able to monitor human wake and sleep activity cycles. An actigraph is generally a watch-shaped device, as seen in Figure 4, worn by the patient on the wrist of the non-dominant hand, in order to measure gross motor activity. It continually records the movements it undergoes over extended periods of time,

while allowing the patients to proceed with their normal daily routine and to sleep in their natural environment. Depending on the type of device, the data can be later transferred to a computer or mobile device and further analyzed. The recorded movements, when partitioned by thirty second epochs, lead to the widely known feature called activity counts, which is a set of numbers expressing the total number of movements within each epoch.

Based on the acquired data, scoring algorithms are used to identify sleep and wake states from the activity counts. This information allows the objective assessment of several sleep parameters like time in bed, total sleep time, sleep efficiency or wake after sleep onset. Actigraphy is suitable to provide sleep and wake information that can supplement PSG tests or be used to pre-screen some patients. Furthermore it can track longitudinal sleep information that may be missed by a one-night PSG study.

Actigraphy has been indicated by the AASM as a suitable method to assist in the evaluation of patients with circadian rhythm disorders, such as advanced sleep phase syndrome, delayed sleep phase syndrome or shift work disorder, and also to assess response to therapy of circadian disorders and insomnia [26]. The AASM's Standards of Practice Committee (SPC) also provided recommendations for the use of actigraphy in clinical practice.

Figure 4. Actiwatch Spectrum, Philips Electronics

2.4.2 Unattended portable monitors in the diagnosis of OSA

As a related project regarding sleep monitoring and diagnosis of obstructive sleep apnea, Collop et al. published a review, including an overview [30], and clinical guidelines [31]

on portable monitoring methods specifically for the diagnosis of OSA. As a result of the work performed, the authors concluded that portable monitoring may be used as an alternative to polysomnography (PSG) for the diagnosis of OSA in patients with a high probability of moderate to severe OSA. For patients with significant comorbid medical conditions, portable monitoring is not appropriate for the diagnosis of OSA, as it may significantly degrade the accuracy of the system. This method may be indicated for the diagnosis of OSA in patients for whom in-laboratory PSG is not possible because of immobility, safety, or critical illness. The portable monitoring must record, at least, airflow, respiratory effort, and blood oxygenation. The airflow, effort, and oximetric biosensors should be similar to the ones conventionally used for in-laboratory PSG.

A good example of an unattended portable monitor is the Philips Stardust II. This system records the body position as well as the airflow, gathered by the nasal cannula, in Figure 5, and/or the oral thermistor. The pulse oximetry, gathered by the SpO_2 sensor which corresponds to the grey sensor in the finger in Figure 5. Pulmonary ventilation is measured by respiratory inductance plethysmography which consists of two sinusoid wire coils installed on a flexible band, one placed around the rib cage under the armpits, visible in Figure 5, another placed around the abdomen at the level of the umbilicus (*belly button*).

Figure 5. Stardust II, Philips Electronics

Although these methods allow measurements to be performed at home, which greatly improves the comfort of the patient who is being monitored, the collected data must be

analyzed by a sleep clinician, and the diagnosis of obstructive sleep apnea (OSA) must be performed only in conjunction with a comprehensive sleep evaluation. Furthermore, the application of the sensors must be performed by an experienced sleep technician or the patients must be educated in sensor application.

2.4.3 Ambulatory polysomnography

Sleep in a laboratory may not be representative of the typical sleep of a subject suspected of a sleep disorder. One of the main reasons is related to the unfamiliarity or intimidation with the environment [32]. The labour-saving and cost-saving benefits of home recordings as well as the increased comfort, privacy, and convenience are the main advantages of ambulatory polysomnography.

There are a lot of products capable of performing ambulatory polysomnography. Philips Alice PDx, visible in Figure 6, is a good example, which is a multi-purpose device capable using different sensors depending on the study being performed. Examples of such sensors are the ones used for OSA monitoring as well as others like ECG, EEG and EOG.

Figure 6. Alice PDx, Philips Electronics

2.5 Unobtrusive monitoring in bed

Ambulatory PSG systems greatly reduce the problems related to sleep monitoring in an unfamiliar environment, potentially diminishing effects such as the "first night effect". However, these types of approaches use obtrusive methods with several contact sensors and wires, which decrease the level of comfort of the patient. Furthermore, these systems are still costly to use and not easy to deploy, generally limiting their application to one or two nights. Although some modalities such as airflow or neural activity cannot be readily replaced by unobtrusive counterparts, other modalities such as cardiac and respiratory

activity, and body movements can be monitored using already available commercial systems.

2.5.1 Ballistocardiography

Ballistocardiography is a technique for producing a representation of repetitive motions of the human body, induced by the heartbeat, occurring due to acceleration of blood as it is ejected and carried through the great vessels. Ballistocardiography obtains mass movements of the body, caused by the heart contraction, giving information regarding the overall performance of the circulatory system. Through this technique the mechanical movement of the heart can be captured by unobtrusive methods from the surface of the body.

An example of a device relying on this technology is the EMFIT's bed foil sensor in Figure 7. Through this type of devices it is possible to unobtrusively acquire cardiac information as well as movement and respiration information.

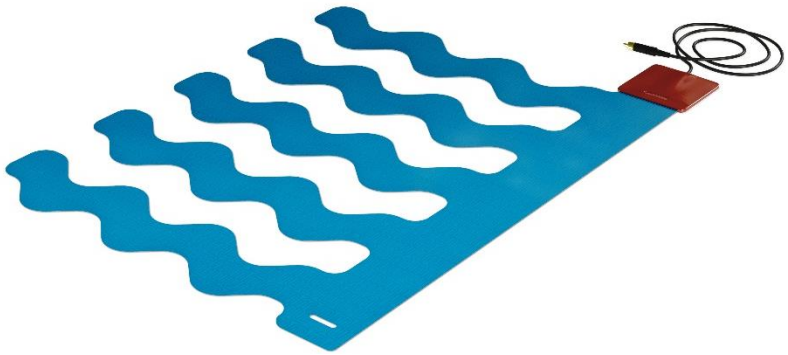


Figure 7. Under-Mattress Bed Sensor, Emfit

2.5.2 Doppler radar

The Doppler radar, uses the Doppler Effect, to acquire information regarding the velocity of the objects at a distance. The Doppler Effect is the change in frequency of a wave, an observer experiences, when moving relatively to its source. The radar beams a microwave signal against the target, waits for the reflection and analysis the frequency of the returned signal, which has been altered due to the target's motion. Doppler radars have a wide

variety of applications including aviation, satellites, meteorology, police speed guns and healthcare.

In the context of sleep monitoring, unobtrusive sleep monitoring devices use the Doppler Effect to measure respiration signals. The emitted microwave signals hit the patient's moving chest wall and are modulated in amplitude and phase. The frequency of the moving chest wall can be calculated through the reflected signal.

An example of a device using this principle for unobtrusive respiration monitoring during sleep is SleepMinder from BiancaMed. This device is capable of unobtrusively monitor respiration during sleep and detecting sleep-disordered breathing events.

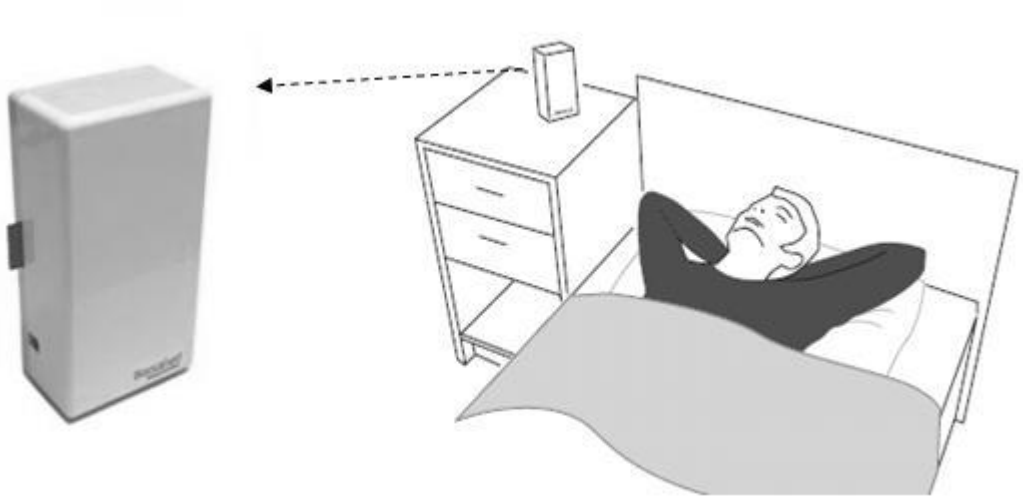


Figure 8. An example set up of SleepMinder, Biamcamed [33]

2.5.3 Video

Different methods have been proposed that rely on video recording for monitoring sleep in an unobtrusive way. For instance Duffy et al.[34] proposed a methodology to access the respiration information of the patient based on an optical chest-wall measurement system. The system used a low powered helium-neon laser to illuminate the chest wall of the patient.

In 2010, Kuo et al.[35] sugested a method to monitor gross body movements and body positions as well as respiration activity of a sleeping patient. In their study they used near-infrared camera with a near-infrared lighting source.

2.6 Sleep stage classification using unobtrusive modalities

The systems described in the previous section have the potential to use non-obtrusive techniques for data collection, which makes then valid options for comfortable, non-obtrusive sleep monitoring over extended periods of time.

In order to compare their performance, the studies presented in this section are described in terms of Cohen's Kappa coefficient. This is a statistical measure of inter-rater agreement for qualitative items. Although there is no consensus about what constitutes a good or a bad performance, Landis and Koch [36], characterized values less than zero as indication of no agreement, 0 to 0.20 as slight agreement, 0.21 to 0.40 as fair agreement, 0.41 to 0.60 as moderate agreement, 0.61 to 0.80 as substantial agreement, and 0.81 to 1 as almost perfect agreement. The metric is further described in section 6.1.

2.6.1 Sleep and wake Discrimination

The study described in this document had used the work of Devot et al. [4] developed within Philips Research as the starting point. It used a Linear Discriminant (LD) classifier trained to classify sleep and wake using actigraphy, cardiac and respiratory features, on a dataset comprised of 35 middle-aged subjects (9 healthy, 27 insomniacs, 16 males and 20 females) and reported an overall Cohen's kappa coefficient of agreement of 0.62 (overall accuracy of 86.7%) , 0.7 for healthy subjects, 0.61 for insomniacs.

This system has been further extended with additional features, such as described by Xi et al. [3]. Using only actigraphy and features derived from respiratory effort extracted from a dataset comprised of nine healthy subjects (eight females) with a mean age of 32 ± 13, it reports a κ of 0.69 (overall accuracy of 95.4%).

2.6.2 Sleep staging using cardiorespiratory signals

The work reported by Redmond et al. [37] is a method for the discrimination of Wake, Non-REM and REM stages, using ECG and respiratory effort signals. In previous work

[38], Redmond et al. executed a similar study, aimed at performing sleep stage classification on subjects being assessed for Obstructive Sleep Apnea. The newly developed study, aimed at examining the effectiveness of the same system when applied to normal healthy subjects.

The dataset used for this work was composed by 31 male subjects with a mean age of 42±7 years.

This study tested and compared the performance of linear and quadratic discriminant classifiers. Temporal varying priori probability information was introduced in the classification results to further increase the classification results. The best performance was obtained by linear discriminant classifier model using temporal varying priori probability. For the 3 class system an agreement of $\kappa = 0.45$ was achieved. When considering only a two class system, Sleep and wake, the agreement would increase to $\kappa = 0.57$.

2.6.3 Sleep staging based on signals acquired through bed sensor

The study performed by Kortelainen et al. [39], had the peculiar characteristic of acquiring the signals using non-obtrusive sensors. It was based on previous work developed in the same scope but with signals acquired through contact sensors[40]. Emfit sensor foils, such as that illustrated in the previous section in Figure 7, were used and placed under the bed mattress, from which the heart-beat interval (HBI) and body movements were obtained.

The dataset used for this work was composed by 18 recordings from shift-work subjects. Nine females, aged between 20 and 54 years, participated in this study with two recordings each.

A time-variant autoregressive model was used for the extraction of the relevant features, and the classification was performed with a hidden Markov model (HMM). The classification results for the 3 classes Wake, NREM and REM achieved κ of 0.44 ±0.19 and an accuracy of 79 ±10% using only three HBI features and one movement feature.

2.6.4 Analysis and Comparison

All the studies presented above reported very promising results, which shows why this area of research is interesting and continues to progress towards a reliable unobtrusive sleep staging system. However, since obtaining cardiorespiratory and movement data in an unobtrusive manner and using this data for automatic sleep stage classification, is not a trivial challenge, there is room for improvements.

In terms of data used, Redmond et al.[37] had the bigger dataset, has he had the bigger number of subjects. Their study had a total of 31 subjects, however all subjects were men of the same age group, suggesting that performance might degrade when using the same system in datasets with female subjects, or with subjects of other age groups. Kortelainen et al.[39] used a total of 18 recordings from healthy subjects, however the subjects were all female and only nine subjects participated in the study (with two nights each). Since the sleep architecture of a subject is usually similar across different nights, classification results when training and testing with data from the same subject, even though they belong to different nights, might be biased. Long et al.[3] used only nine healthy sleeping subjects. The dataset to be used in this research project will consist of more subjects. Furthermore it will be interesting to observe the results obtained with a bigger number of subjects and with a classification process using more features. Moreover it is possible that during the research project more features are created and added or the method or the creation method of the features is modified.

For the maximization of the discrimination power, these studies used various methods like detrended fluctuation analysis or time-dependent auto regressive models. Such methods aim to retrieve information from the features taking the time factor into consideration. Furthermore Redmond et al.[37], suggested that the use of time varying prior probabilities would result in improved results. Time plays therefore an important role in the sleeping architecture and should not be forgotten when developing the classifier.

In terms of the classification process, two of these studies chose a Linear Discriminant and another used a Hidden Markov model. Hidden Markov Models and Bayesian Linear Discriminants are widely known classification algorithms, each with some advantages

and disadvantages. A hidden Markov model (HMM) is a statistical Markov model in which the system being modelled is assumed to be a Markov process with unobserved (hidden) states [41]. Bayesian Classifiers are based on the idea of assigning unknown patterns to the most probable class within a known set of classes. Each pattern is characterized by a feature vector [42].

Although sleep follows a certain sequence of states throughout the night, Hidden Markov Models may not be appropriate for this specific classification process. The complexity of the model and wrong assumptions regarding the process may reduce the classification performance. Assumptions such as the Markov assumption which requires that the next state is dependent only upon the current state, and the stationary process assumption which requires that state transition probabilities are independent of the actual time at which the transitions takes place, may cause the classification results to drop. Besides the probability of a certain sleep stage may not only depend on the previous epoch, but on the sequence of previous epochs. Moreover the probabilities of stage change vary across time, in fact, before the study of Kortelainen et al.[39] and in an attempt to capture the time varying stage probabilities, Redmond et al.[37], proposed a method that uses time-varying *a priori* probabilities, accomplishing positive results.

Linear discriminants are very simple classifiers where probabilities can be added or altered in order to achieve the best representation of the problem at hand [41]. Although time-varying *a priori* probabilities played an important role in the task of capturing the time varying probabilities of sleep stages, this method can be further explored.

Classification results are very similar for the studies performed by Kortelainen et al. [39] and Redmond et al. [37]. The study performed by Long et al.[3] only analyzed the classification for Wake and Sleep which makes the comparison against the systems classifying Wake, Non-REM and REM stages difficult since systems have different problem complexities.

One important detail is the high standard deviation in the classification results of the subjects reported in the literature. The very high standard deviation suggest that these results are being affected by large between-subject variability. Between-subjects variability,

when all data is pooled, greatly decreases the discrimination power comparing against the average discrimination power of the same feature for the each subject. Normalizing the feature values across each subject this between-subject variability can be reduced, possibly improving the classification results.

2.7 Summary

This section began with an introduction to sleep medicine, with information regarding the normal human sleep and sleep disorders. Some studies, methods and devices related to sleep assessment were reviewed.

This section finished with an examination of some techniques, technologies and studies within the same scope of this project. The different studies were compared, taking into consideration the datasets used, the methods applied for the classification and the classification results.

3 Datasets

For the investigation of the automatic classification of sleep stages based on cardiorespiratory data and evaluation of developed algorithms, annotated data is needed. Since the goal of this research work is to use supervised learning to classify data into sleep stages, the annotated sleep stages and corresponding cardiorespiratory measurements are needed.

The annotated sleep stages were annotated by a sleep professional, in 30-second epochs following the AASM or R&K guidelines.

In this work four different datasets are considered, namely the Boston healthy, The Boston insomniacs, Eindhoven and SIESTA.

3.1 Boston Healthy dataset

The Boston Healthy dataset comprises 10 healthy subjects (8 females) with an average age of 31 ± 12 years. Sleep stages were scored by an expert according to the AASM guidelines.

3.2 Boston Insomniacs dataset

The Boston Insomniacs dataset comprises 27 subjects (13 females) diagnosed with insomnia, with an average age of 46 ± 14 years.

Sleep stages were scored by an expert according to the AASM guidelines.

Looking at Figure 9, high amount of epochs score as wake is observable. Sleep efficiency is therefore very low on insomniacs subjects. Moreover the number of epochs scored as deep sleep is much lower than in normal healthy subjects considering the same age of the subjects. As expected, a person who suffers from insomnia, has a predominance in N2 and N1, with difficulty to fall asleep followed by several awakenings which explains the high amount of N1 and wake epochs.

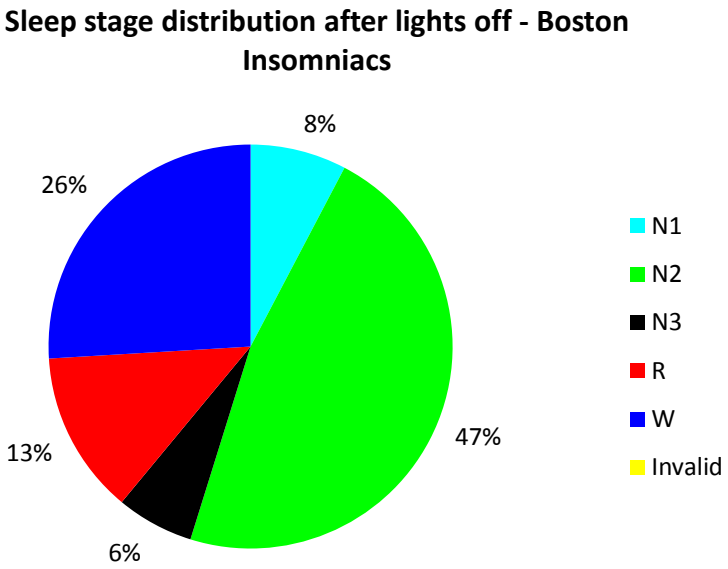


Figure 9. Total epochs spent in each stage from lights out, for Boston Insomniacs dataset

3.3 Eindhoven dataset

The Eindhoven dataset comprises 12 subjects (6 females) with an average age of 29 ± 5 years.

Sleep stages were scored by an expert according to the AASM guidelines.

3.4 SIESTA dataset

SIESTA was a project funded by the European Commission which involved several European partners. The aim of the project was to research the nocturnal human sleep as well as to develop and evaluate new methods of sleep analysis [43].

The SIESTA dataset is the largest dataset in this study. It has a total of 292 subjects, amongst which subjects diagnosed with general anxiety disorder, depressive disorder, sleep apnea syndrome, restless legs or Parkinson.

They have an average age of 52 years ± 17. This dataset is composed by 126 females and 166 males.

Sleep stages were scored by a consensus of at least two experts according to the R&K guidelines.

3.5 Invalid data handling

There are many reasons that can lead to invalid feature values, such as artifacts on the sensors. Nevertheless, the invalid data on a certain feature on some certain epochs is a very delicate problem. If invalid data intervals were not bigger than two consecutive epochs, an interpolation method, such as cubic spline, could be used with some assurance on the trustworthiness of the final data. However, since the invalid data is present in a much further extent, sometimes bigger than 100 consecutive epochs for a certain feature, interpolation of the missing data might not be reliable. Since some transformation techniques use time series statistics, analysing the data with a fixed window of epochs, as will be further explained in chapter 4.4, the deletion of the invalid epochs from the time series of a certain feature, will make the sample rate inconstant which implies that statistical analysis with a window of fixed length can no longer be applied.

Considering this two constraints, epochs with invalid data will be handled in two different ways. Upon the normalization and transformations of the features, the missing data will be interpolated, in order to be able to preserve the time series and a constant sample rate. After the features are normalized and transformed, the epochs previously marked as containing invalid feature values will be ignored and excluded from training and validation. In this work, given the low prevalence of invalid epochs in most subjects, it was decided that the invalid features values would not significantly influence the results. In a real world scenario, this particular problem would have to be addressed separately. Invalid feature values would probably have to be detected and classified into indeterminate, or attempt to use other features to classify those particular epochs.

3.6 Subjects selection

As seen in previous sections, a healthy adult should have sleep efficiency over 90%, and the sleeping time should be distributed between 45% to 55% in light sleep (N1 and N2), 10% to 20% in deep sleep (N3), and 20% to 25% in REM sleep [44]. Sleep efficiency is defined by the sleeping time divided by the total time spent in bed, from the moment lights are turned off with the intention to sleep until lights are turned on again before definitely getting up. Considering the first night effect and the widespread age distribution of the subjects, these parameters are going to be substituted by less strict parameters.

The less strict parameters used for this study are: minimum sleep efficiency 80%, minimum asleep REM ratio 15%, and minimum asleep deep sleep ratio 7%. Due to the removal of a considerable number of epochs in some subjects, a minimum night length of 6 hours (corresponding to 720 epochs) was set. These parameters are only applicable to healthy subjects and therefore will not be used for the Boston Insomniacs dataset.

Knowing healthy subjects have stable sleep architecture, with the order and the timings of the sleep stages being quite similar across different nights, there could be a bias if data from one of the nights of a given subject would be used for training with the other night would be used for validation. Therefore, and considering that the first night effect has a detrimental effect on the sleep architecture, the second night of the SIESTA dataset will not be used.

Using the less strict parameters, the dataset of healthy subjects used in this work comprises a total of 61 subjects (42 female), 54 from the SIESTA dataset, 2 from the Eindhoven dataset and 5 from the Boston Healthy dataset. The mean age of these subjects is 41 ± 17 years.

3.6.1 First night effect

The "first night effect" is a well-known issue regarding PSG recordings on the first night spent by a subject in a sleep laboratory. This effect translates in a number of differences in the sleep architecture between the first night and consecutive nights. Although most differences are observed between the first and the second night, there are studies suggesting the first night effect is present beyond second night [45]. The first night usually contains more wake epochs and less REM epochs. Furthermore there is a delay in the onset of N3 and REM and the sleep is more fragmented [32].

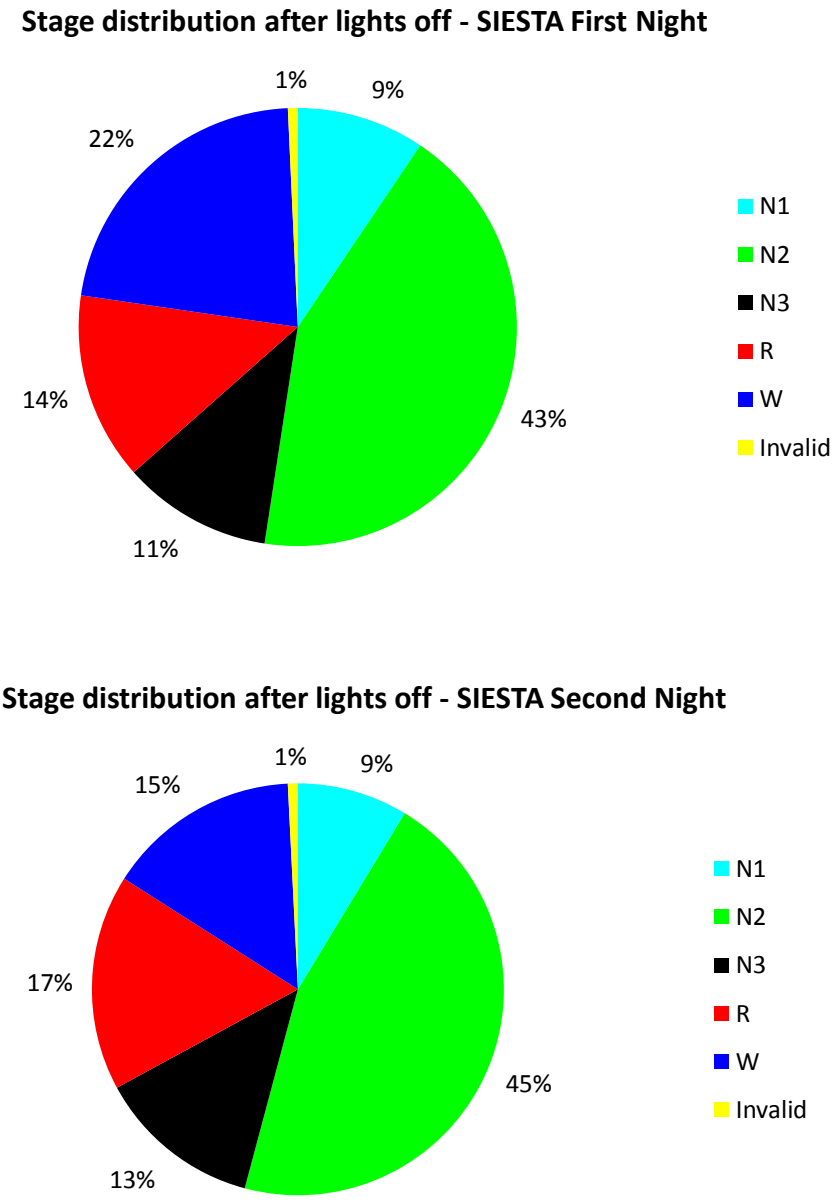


Figure 10. Comparison of sleep stages distribution of first and second night

Looking at Figure 10, it can be seen that there are differences, especially regarding sleep efficiency. The average sleep efficiency in the first night for the SIESTA dataset was 78% which according to literature is very low. However on the second night the same subjects had an average sleep efficiency of 85%. Besides the improvement in sleep efficiency, REM and deep sleep also occupy a larger percentage of the sleep in the second night. This

concurs with the hypothesis that sleep in the second night is less affected by the first night effect.

3.7 Establishing ground truth

The methods used to deal with the epochs with invalid feature values are already described. However, before joining the subjects into a dataset that is going to be used for this study, the ground truth (class labels), have to be converted into a common format.

This problem exists because the subjects belong to different datasets, following different sleep scoring criteria, AASM[46] and R&K [47]. Furthermore, this work will consider only four sleep stages, light sleep, deep sleep, REM and wake.

Considering only three classes, Non-REM, REM and wake, the establishment of the ground truth would be easy since both AASM and R&K consider REM and wake. However when considering four classes, which implies further division of Non-REM sleep, the problem is not so simple.

R&K further divides Non-REM sleep into stages I, II, III, and IV, while AASM divides it into stages N1, N2 and N3. In this work, Non-REM sleep will be divided into light and deep sleep. Therefore in order to convert the ground truth to four sleep stages, for AASM, stages N1 and N2 correspond to light sleep while N3 correspond to deep sleep. For R&K stages I and II correspond to light sleep, while stages III and IV correspond to deep sleep.

3.8 Cardiorespiratory features

In this section, the methods used to record the data will be presented, with examples of features derived from the acquired data.

The available feature set is quite extensive, so as a way to summarize the information contained within each feature, a table with the corresponding feature number, name of the feature and a small description is provided in the Appendix.

The feature numbers are not sequential for several reasons. Some features are not available to all datasets in this study. Other features are being implemented, have errors or, for other reasons, were removed.

3.8.1 ECG data

Electrocardiography is a non-invasive method to monitor the electrical activity of the heart. The electrical activity is detected by electrodes which are attached to the skin. The recordings produced by this methodology are called electrocardiograms. An example of a recording of ECG is visible in Figure 11.

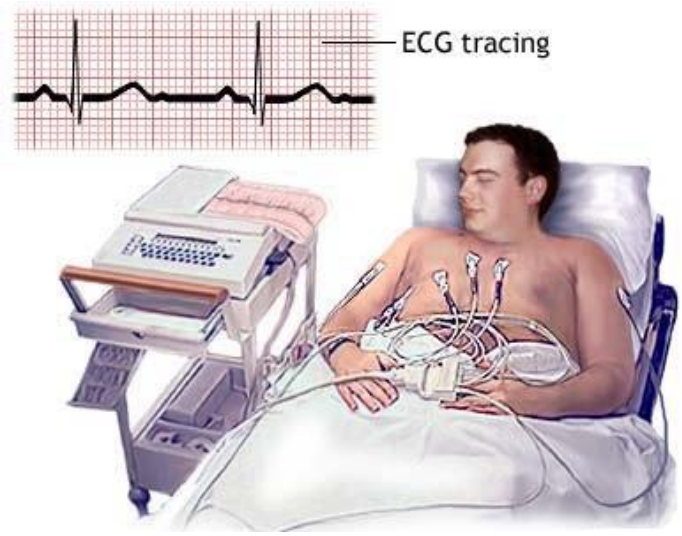


Figure 11. Example of an ECG recording

A typical ECG tracing is composed by a periodic signal that starts with a P-wave, which is followed by the QRS complex and ends with a T-wave. Figure 12, shows the typical tracing and its waves and intervals.

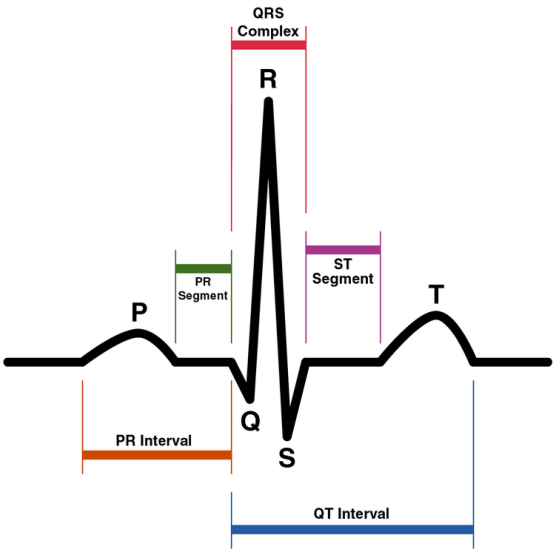


Figure 12. Schematic representation of a typical ECG tracing

Feature 27, which measures the mean RR interval, is a good intuitive example of a feature derived by the ECG data. This feature represents the interval between an R wave and the consecutive R wave. An example showing the feature values of this feature, for a given subject, is visible in Figure 13.

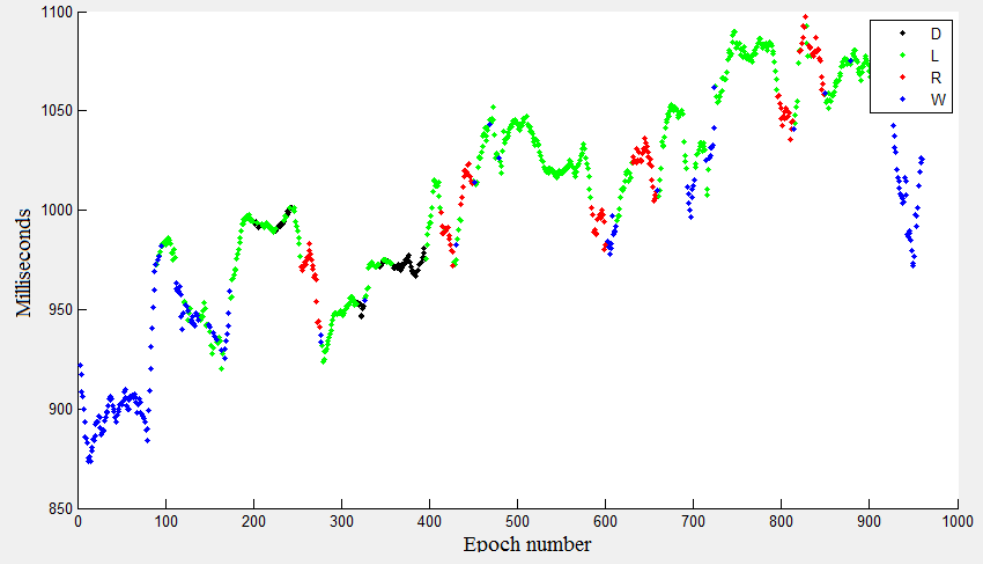


Figure 13. Example of feature 27, mean RR interval, for a given subject

3.8.2 Respiratory inductance plethysmography

Respiratory inductance plethysmography is a non-invasive method to monitor breathing movements, without any connections in the airway opening, by measuring chest and abdominal wall movements. This method uses two elastic belts, in which an inductance sinusoidal wire coil was sew into, which are worn around the ribcage and the abdomen. During inspiration and expiration, the chest and the abdominal area vary, which changes the self-inductance of the coils and the frequency of their oscillation. The detected changes in inductance are proportional to the changes in ribcage and abdominal volume. In Figure 14, and example two elastic respiratory inductance plethysmography belts are visible.

Figure 14. Pro-Tech ezRIP, Philips Electronics

A good example of a respiratory feature acquired through this method, is feature 13 which records the respiration frequency of the subject. An example of the feature values, for a given subject, of respiration frequency can be visible in Figure 15. Note that upon extraction of this feature, the values have been normalized by subtracting the mean.

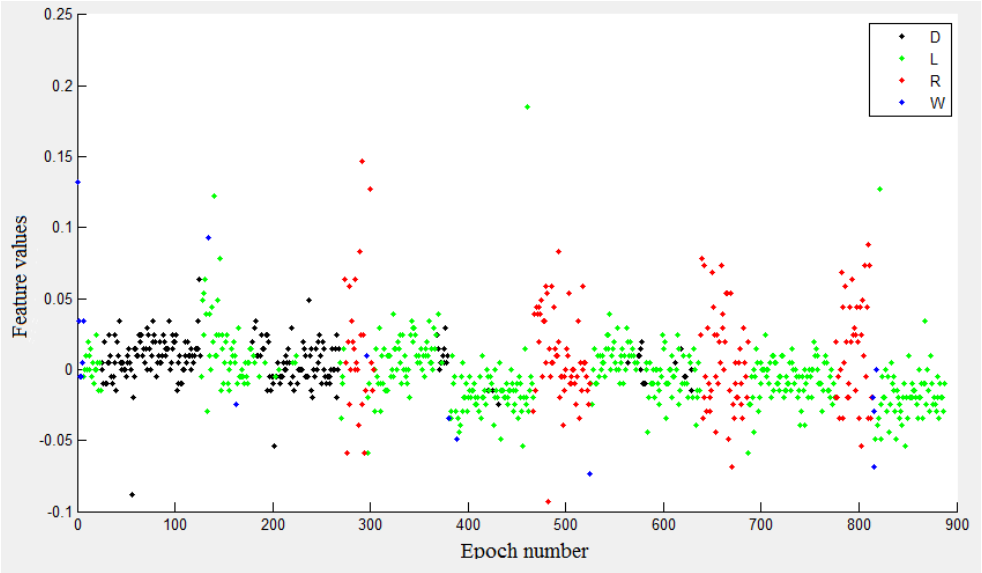


Figure 15. Example of feature 13, respiration frequency, for a given subject

3.9 Summary

This section revised the datasets used for this study. The data from the different subjects were compared against sleep literature which showed that some subjects had an unusual sleep architecture. The sleep architecture of the subjects of this study was further analyzed, which will culminate in the selection of the subjects with a normal healthy sleep architecture. The subjects were filtered according to several relaxed parameters obtained in literature. These subjects formed the final dataset that is going to be used in the next sections. In order to provide an insight of the features used in this work, examples of cardiorespiratory features were presented.

4 Feature Analysis

Feature processing is a very important step in the training/classification process. This is a preliminary phase which transforms the properties of the data which will be more effectively discriminated by the classifier. The stages in a pattern recognition system are in a pipeline fashion meaning that each step depends on the success of the previous phase in order to produce optimal results. In order to improve the discriminatory power of each feature, it will first normalized. This method aims at minimizing the impact of between-subject variability on the expression of the differences between sleep stages on the features. After features are normalized, they will be further processed (or transformed). This method aims at extracting additional information about each feature, for example, about the way it varies in time, or its low- or high-frequency components. In order to determine which normalization and transformation methods actually contribute towards these goals, features will be evaluated individually by their discriminatory power using a class separability measure.

4.1 Class separability measures

In this work class separability measures are used to measure the discrimination power that a feature can provide for a certain individual class or multiple classes. As a way to understand what is a discriminative feature Figure 16 is present. This figure shows feature 169, for a given subject, which is a feature related to respiratory amplitude, which is able to provide good discrimination power for deep sleep.

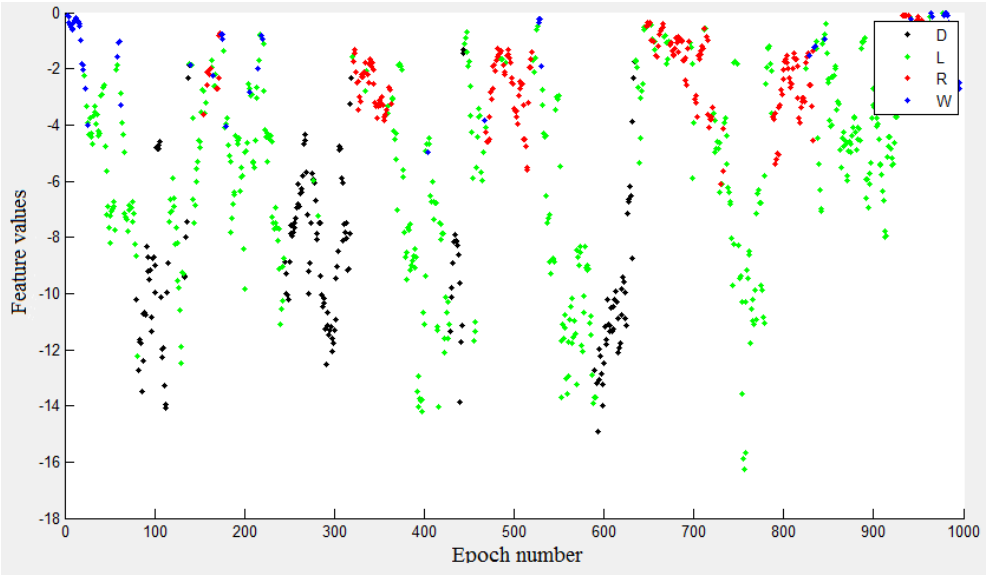


Figure 16. Example of feature 169 for a given subject, a feature with good discrimination power for deep sleep and REM.

As an opposite example, feature 13, visible in Figure 17, representing respiration frequency, for the same subject, is a feature which is unable to provide good discrimination power to any class. Although originally having low discriminatory power, this feature has significant statistical information that can be used to obtain new features with good discriminatory power.

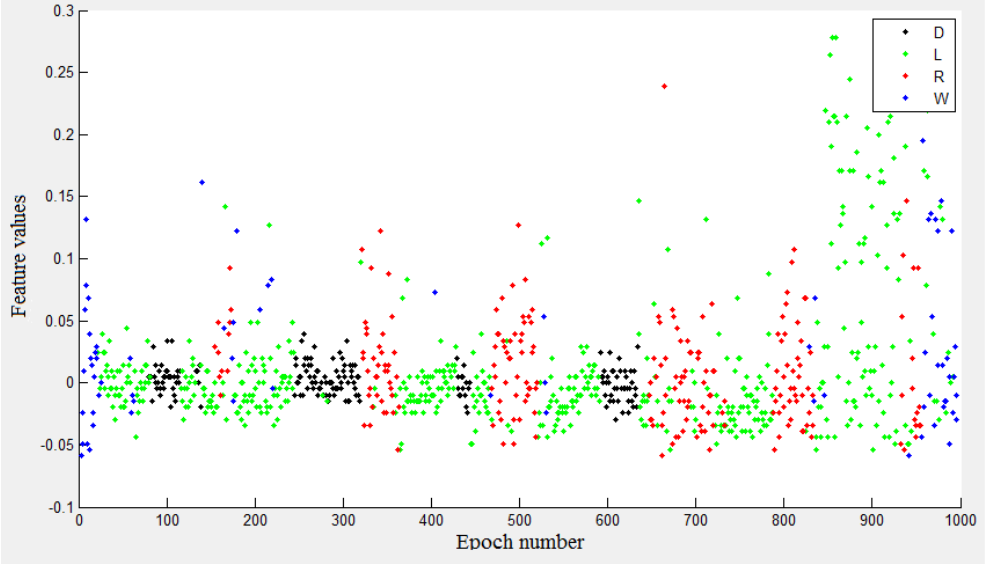


Figure 17. Example of feature 13 for a given subject, a feature with low discrimination power for all classes

4.1.1 Absolute standardized mean distance

One of the simplest and fastest methods for evaluating the class separability of a given feature is the absolute standardized mean difference (ASMD). Furthermore, as will be discussed in section 5, the linear discriminants used for classification depend on the Mahalanobis distance metric, which is a multivariate version of standardized mean distance. As such, if this metric is used to assess normalization/transformation methods, it is guaranteed to yield improvements when using the linear discriminant. For a given feature, with N features values x, the ASMD between class 1 and 2 with mean μ_1 and μ_2 and standard deviation of $\sigma_{1,2}$ the ASMD is given by

$$ASMD = \left|\frac{\mu_1 - \mu_2}{\sigma_{1,2}}\right| \quad (1)$$

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (1.1)$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \quad (1.2)$$

Although very fast the ASMD distance is clearly only applicable in the binary case, i.e., when only two classes are considered. Because the problem addressed in this study concerns multiple classes (corresponding to multiple sleep stages), the ASMD distance of a

certain class will always refer to the absolute standardized mean distance between the mean of the analyzed class and the mean of the aggregated remaining classes. This method was chosen because, as seen in section 5, a classifier will be used per each class and trained as one class versus the rest.

4.1.2 One way analysis of variance F-statistic

One way analysis of variance (or one way ANOVA) tests the null hypothesis that the samples in two or more groups are taken from populations with the same mean values. The F-test is the ratio of two scaled sums of squares which represent different sources of variability. The statistic tends to be greater when the null hypothesis is not true. The one-way ANOVA F-test, for feature values *x*, *N* epochs, and *k* classes, with n_i representing the number of instances from a certain class i, x_i the vector of their respective values, $\overline{x_i}$ the mean of the values and x_{ij} a value of the vector, is given by

$$F = \frac{between - group\ variability}{whithin - group\ variability} \tag{2}$$

$$\bar{X}_{GM} = \frac{\sum x}{N} \tag{2.1}$$

$$between - group\ variability = \frac{\sum_i^k n_i\ (\overline{x_i} - \bar{X}_{GM})^2}{k-1} \tag{2.2}$$

$$whithin - group\ variability = \frac{\sum_i^k \sum (x_{ij} - \overline{x_i})^2}{N-k} \tag{2.3}$$

The between-group variability is described as the sum of squares between groups, and is sometime referred to as the explained variance. The within-group variability expresses the variability within the groups and is sometimes referred to as the unexplained variance.

The F-statistic will be large if the between-group variability is large relative to the within-group variability, which is unlikely to happen if the population means of the groups all have the same value.

4.2 Unprocessed features

Every analysis of data starts with a preliminary part of unifying data from different data sets, cleaning it and preparing it for further processing. That part of the process was addressed in chapter 3. Before features are normalized and further processed, they were first

analysed in terms of their discriminatory power. This step will serve as a starting point for further analysis and as a comparison for the evaluation of the different techniques which will be applied. When the discrimination power of the original features is compared with the processed features, derived from the original ones, it will be clear that the processed features possess bigger discriminatory power and a lower intra subject variation.

In Figure 18 the ASMD is measured for all features for deep, light and REM sleep as well as wake. Note that the ASMD refers to the pooled data after aggregating the feature data from all subjects. From this figure it's possible to visualize that each class has its own good set of discriminative features. For instance feature 67 is good for the discrimination of deep sleep, but not for the other classes.

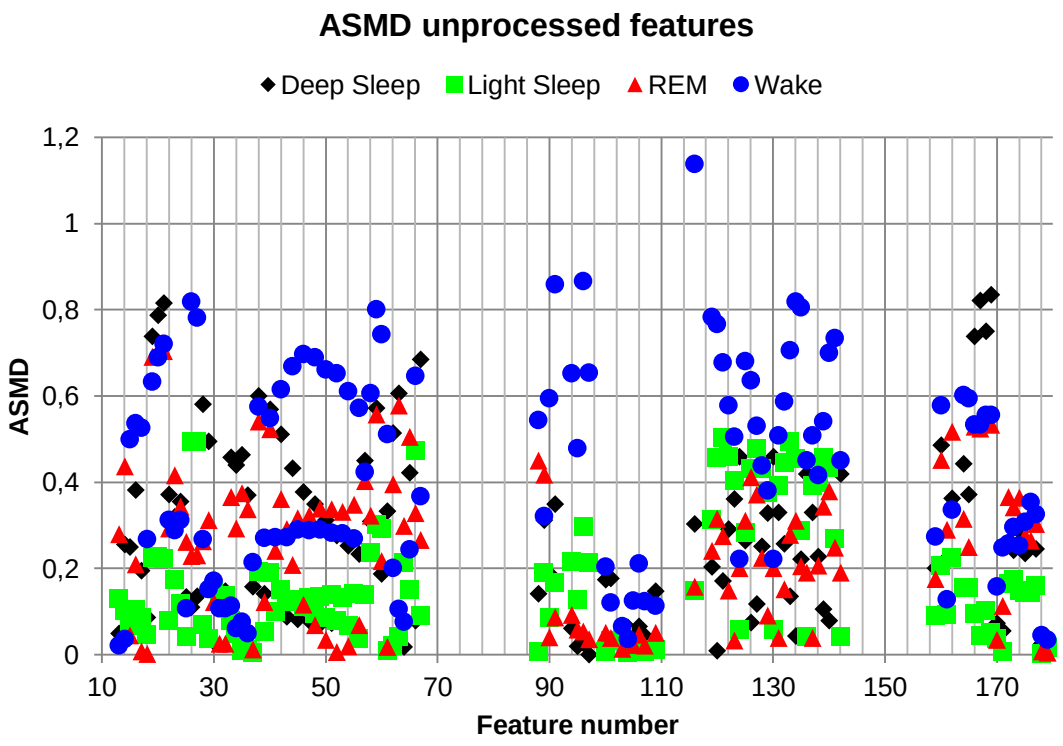


Figure 18. ASMD of all unprocessed features for four classes

To have a better understanding about the relationship between each individual class and the number of features discriminatory power for that particular class, Figure 19 is presented below. This new figure consists of four individual histograms, each illustrating the absolute count of features with given ASMD for each class versus the remaining classes.

In order to compute the histograms, the ASMD was first divided into equally spaced bins. For each bin, the number of features with an ASMD between the boundaries of that bin was counted.

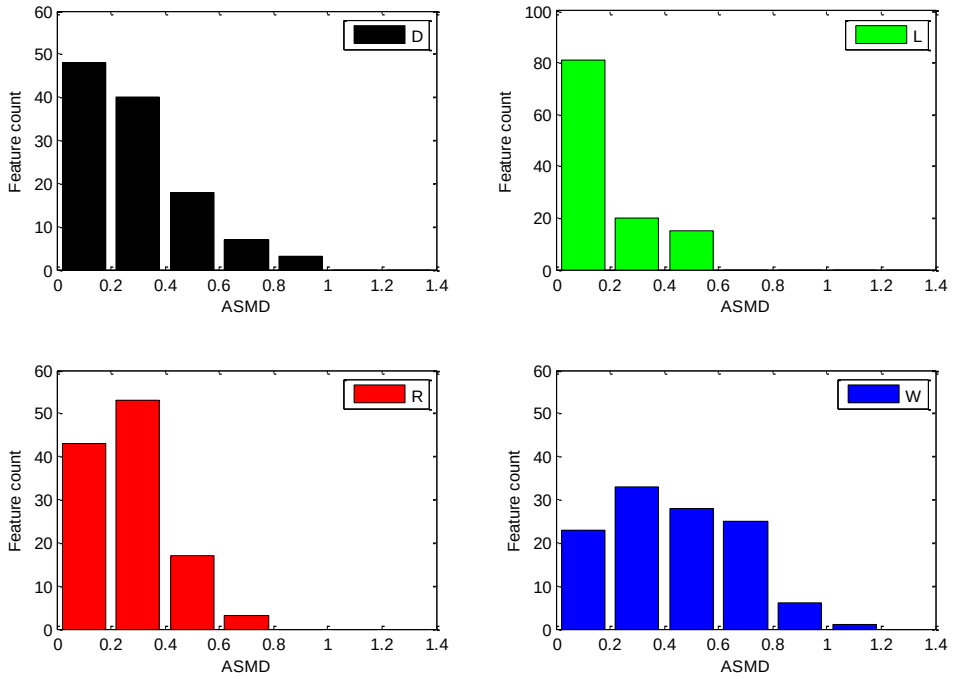


Figure 19. Histograms relating the ASMD with the number of features per class

From these figures, it can be concluded that the class with most discriminatory (unprocessed) features is the wake class. This class has several features with a high ASMD and therefore several feature capable of providing good discriminative power. Following the wake class, deep sleep and REM stages have some features with reasonable ASMD. On the other hand, light sleep seems to have most features below an ASMD of 0.4. Looking at this information, one can say that the wake class is the easiest to discriminate, followed by REM and deep sleep. Light sleep will be harder to discriminate with these features.

4.3 Feature normalization

Feature normalization aims at reducing the between-subject variability arising from differences in the sensors used during PSG recordings (since they were performed in different sleep laboratories), and differences in the physiological expression of sleep stages between subjects.

4.3.1 Motivation

Due to fact that the range of possible values for any measurable characteristic, of a human being analyzed or treated in the course of a particular study can vary significantly, feature normalization is needed in order to compensate for the differences between the subjects. The fact that the data was obtained from several different sources makes feature normalization even more important, as the equipment and techniques used to obtain the data can be different. Therefore the variations between subjects might partially be explained by technical reasons. Different sensors which may have different scales or frequencies for data acquisition, makes the analysis of the pooled data look inconsistent although the data is correct. For instance an electrode on one sleeping center might be recording milivolts while other sleeping centers might have the same electrode recording data at volts. The data is therefore incomparable and requires normalization.

Besides these technical aspects, the between-subject variations can be explained by the biological differences between different persons. These variations are well-known in the biological field as each person has different reactions for the same stimulus. For instance the heart rate of two healthy subjects can be substantially different during similar activities.

As a result from this effect, the pooled ASMD of the features, the ASMD of the concatenated features from all the subjects, will be greatly reduced compared with the mean ASMD for each feature, the mean of the ASMD of a feature for all the subjects. Furthermore this effect produces high disparity on the classification performance between different subjects, as perceived from the high standard deviation in the classification results of similar studies [37], [39].

Figure 20 and Figure 21 illustrate the problem at hand, where the two plots depict the differences between the mean ASMD for all subjects, and the pooled ASMD per feature for deep sleep and REM. Analyzing these figures, it is clear that there are several features, such as 130 for deep sleep and 88 for REM, with good discriminatory power, as expressed by their ASMD. However, when those features are pooled, the discriminatory power clearly decreases. For instance for deep sleep, in feature 130, the mean ASMD is 0.85 while the pooled ASMD is 0.45. For REM for feature 88, the mean ASMD is 0.75 while the pooled ASMD is 0.45.

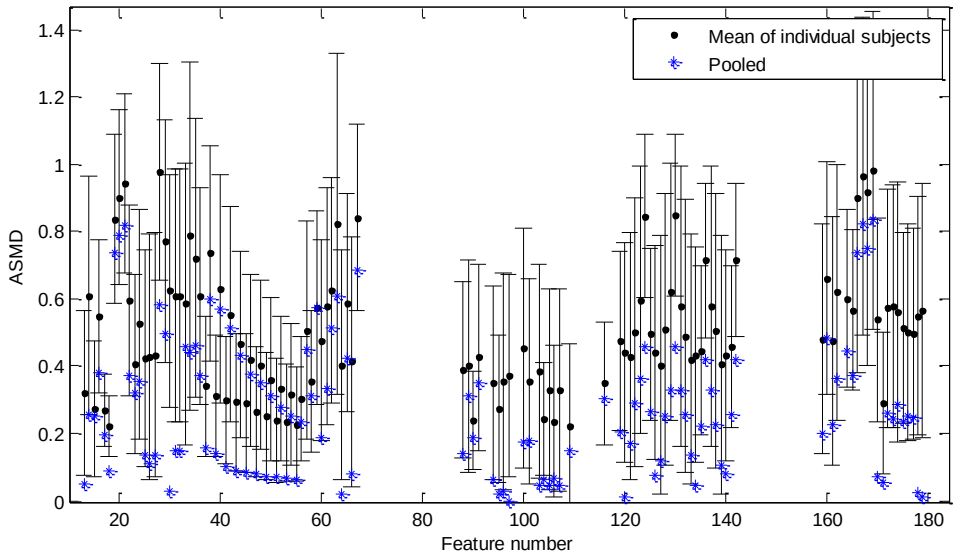


Figure 20. Mean and pooled ASMD for deep sleep versus the remaining classes, for each feature.

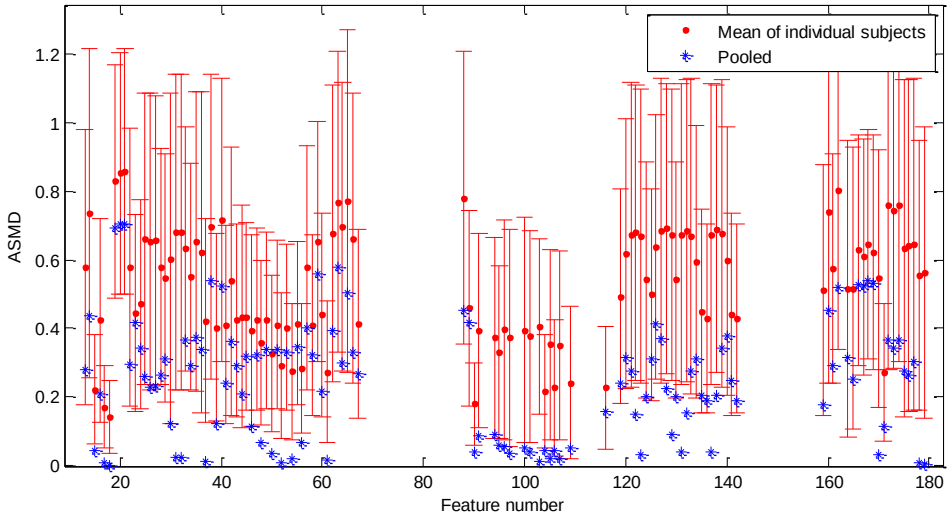


Figure 21. Mean and pooled ASMD for REM versus the remaining classes, for each feature.

By normalizing these features, the between-subject variations should be minimized, which in turn should improve the pooled discriminatory power. Ideally, the pooled ASMD should increase up to the same level as the mean ASMD.

4.3.2 Normalization Techniques

The normalization techniques can be divided in three groups. Clipping operations, which aim at reducing the effects of the outliers, scaling techniques which aim at standardizing the amplitude of the features for all subjects and quantile normalization, a method capable of transforming the features values into a given distribution. Below a detailed description of the methods is available.

4.3.2.1 Winsor

Winsoring or Winsorization of data is a procedure which reduces the effects of possible outliers by limiting the extremes of the feature values. It is named after Charles P. Winsor [48][49]. The effect of this transformation is similar to clipping in signal processing. Values lower or bigger than a given percentile boundary, for instance 5 and 95, are clipped to the values of the 5th and 95th percentile. This technique has no consideration for the mean or scale of the data from the different subjects and is usually only used as a pre-processing step before further normalization techniques.

4.3.2.2 Amplitude

Amplitude normalization is, as the name suggests, a scaling operation. The original feature values are scaled and transposed by defining a new reference maximum and minimum value. Considering a value X and reference values for new maximum and minimum,

$$norm_{amp} = new\_min + \frac{X - min(X)}{(max(X) - min(X)) * (new\_max - new\_min)} \quad (3)$$

4.3.2.3 ZScore

This is a widely known technique for feature normalization. This simple technique removes the mean from the data and normalizes its standard deviation to one [50]. For sample data with mean $\overline{X}$ and standard deviation σ the z-score of a value x is:

$$z = \frac{x - \overline{X}}{\sigma} \quad (4)$$

4.3.2.4 Percentile

This technique removes the median and clips values to an upper and a lower threshold. Since it uses percentiles values (for example, the 25th and 75th), this technique is robust against outliers in the data. A new lower and an upper value are selected as the target values for the percentiles. As an example, assume that the values -1, 0 and 1 are the desired minimum, median and maximum and that the 25th, 50th and 75th percentiles will be used. For a given set of values, the percentiles are computed. The percentile 50th is first subtracted in order to match the target value 0. The values above percentile 50th are then scaled according to the values of percentiles 50th and 75th, so that the new 75th percentile is 1. The same procedure is applied for the values below the 50th percentile, therefore they are scaled according to the values of the 25th and 50th percentiles such that the new 25th percentile is -1. Representing the percentile function as Prc, for a feature values x, the final value of the data point p in that set is:

$$Prc(\mathrm{p}) = \begin{cases} \dfrac{\mathrm{p} - Prc_{50}(x)}{(Prc_{75}(x) - Prc_{50}(x)) \times (new\_max - new\_mid)} + new\_mid, \\ \mathrm{p} \geq Prc_{50}(x) \\ \dfrac{\mathrm{p} - Prc_{50}(X)}{(Prc_{50}(x) - Prc_{25}(x)) \times (new\_mid - new\_min)} + new\_min, \\ \mathrm{p} < Prc_{50}(x) \end{cases} \quad (5)$$

Because some features do not follow a normal distribution, an upper or lower limit can exist with several points having the same value, which caused the values of 25th and 50th or 50th and 75th percentiles to be the same. A good example of this problem is features like activity counts, where most epochs have the value of 0. For this special case, a new approach was implemented, in which the maximum and minimum values of a feature are not considered when calculating the percentiles.

4.3.2.5 Quantile

Quantile is a normalization technique that aims at making two different distributions identical in terms of their statistical properties [51]. In this work, three reference distributions were used to normalize each feature of each subject: the normal distribution, the exponential distribution and the uniform distribution. To normalize the features values, the first step is to create a reference distribution with the same length as the original distribution. The next step is to create two auxiliary vectors with the reference and original distributions sorted according to the feature value. After sorting, the highest element in the original distribution, will take the highest value of the reference distribution. The same procedure is applied to the second highest values and so on. Following this procedure the original feature values will be converted so that the feature will follow the given reference distribution. Figure 22 shows an example of the feature values of feature 169, for a given subject. Below, in Figure 23, it is possible to see the same feature for the same subject after the application of quantile with normal distribution as the reference.

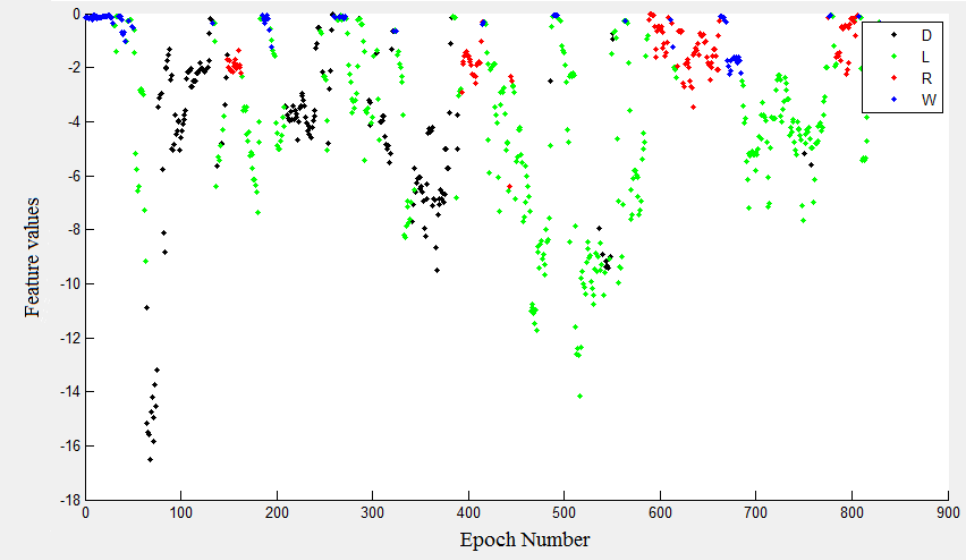


Figure 22. An example of feature values of feature 169 before quantile normalization

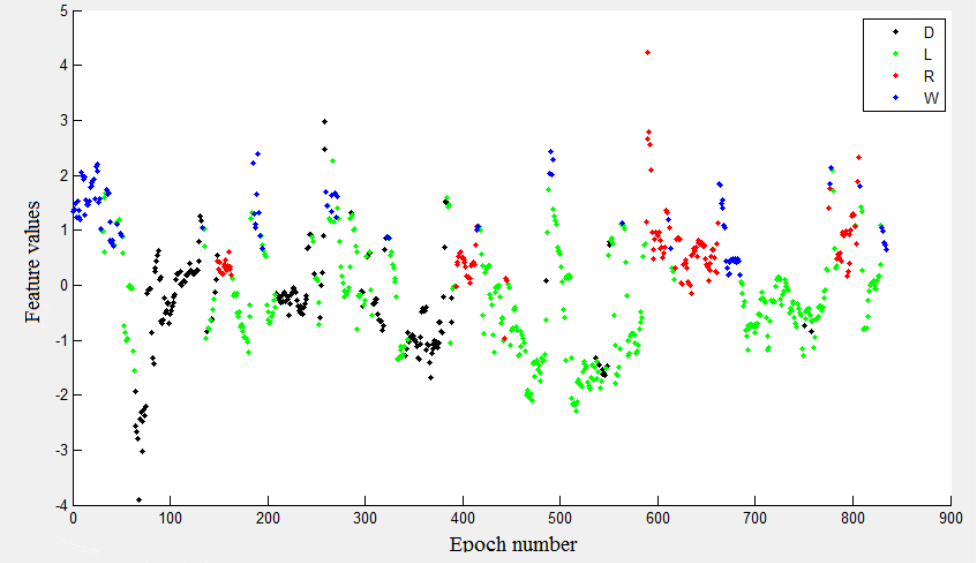


Figure 23. An example of feature values of feature 169 after quantile normalization with normal distribution as reference

4.3.3 Combining techniques

In order to choose which of the different normalization techniques improves the discriminatory power of each feature, a systematic search, described in Figure 24, for the combination of methods that produces the best discrimination is performed. The normalization methods are grouped in Scaling and the Quantile methods. The ZScore, Amplitude, and Percentile techniques change the mean and scale the amplitude of the features, and so

they are grouped within the scaling methods. The Quantile methods change the original distribution of the feature to a predefined distribution. Each normalization technique in a group is treated as an independent node.

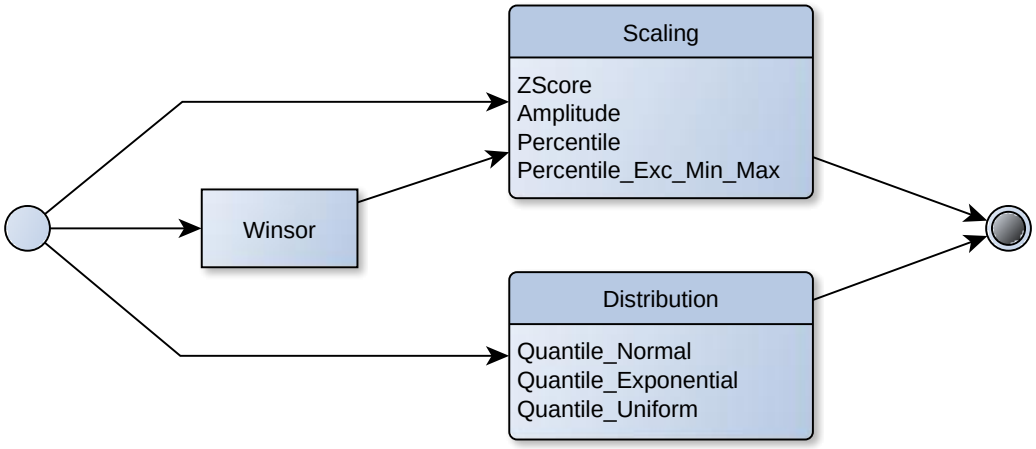


Figure 24. Normalization techniques search graph

The Winsor technique might be used in order to correct the influences of the outliers in the scaling methods. Since the quantile methods are already capable of dealing with the presence of outliers in the data, the Winsor method is not needed for these type of normalizations. With the organization of the normalization methods in groups and in a search graph, the search for the best normalization can be performed using a depth-first search algorithm. This algorithm begins from the starting node and evaluates as far as possible before going back to explore the remaining nodes. It is appropriate for this scenario since the number and depth of nodes in the graph is small.

In order to evaluate the techniques in each node of the search graph, the one way ANOVA F-statistic metric is used. This metric is preferred over the ASMD since it allows the measurement of the discriminatory power of a feature across multiple classes. Since in the normalization procedure the main goal is to improve the pooled feature discriminatory power regardless of the classes, this metric is more appropriate for this task.

4.3.4 Normalization results

After finding the best normalization or set of normalizations that maximize the discriminatory power of each feature, the pooled ASMD of each feature is noticeably larger.

Looking at Figure 19 and Figure 26, there is a clear increase in the number of features with ASMD above 0.6. There is a considerable amount of features capable of providing a ASMD bigger than 0.6 for wake. For deep sleep and REM, the number of features is still good. Light sleep on the other hand, when looking only at the ASMD, did not seemed to benefit from the normalization process. This is due to the inexistence of features capable of providing a good discriminatory power for light sleep.

Using the information present in the same plot, it is possible to see the ASMD values per feature after the normalization. When comparing Figure 18 with its clear that some ASMD values increased considerably. For example, for deep sleep, feature 67 increased from 0.68 to 1 and feature 124 from 0.46 to 0.92. For wake, feature 18 increased from 0.27 to 0.83 and for REM, feature 65 increased from 0.5 to 0.82.

After normalizing the features, there are still some features that are not able to provide any discriminatory power for some classes. This is normal as each feature was designed to provide discrimination to some certain classes and the normalization aims only at reducing the between-subject variability.

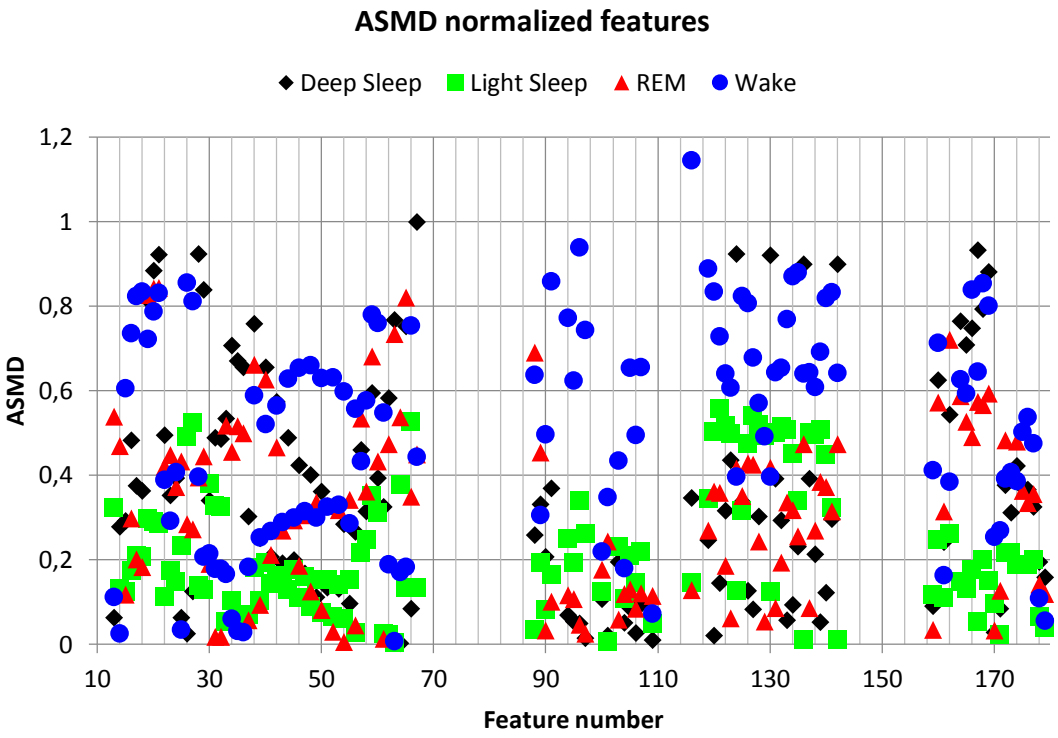


Figure 25. Pooled ASMD of all normalized features for four classes

Figure 26 illustrates the histograms with the number of features per ASMD bin for each. "O" stands for the feature counts in the original feature set, and "N" for the normalized feature set. The color scheme indicating each class is maintained, black corresponding to deep sleep ("D"), green to light sleep ("L"), red to REM ("R") and blue to wake ("W").

Analyzing the figure, it is clear that the number of features with ASMD lower than 0.2 substantially decreased in comparison with the original feature set.

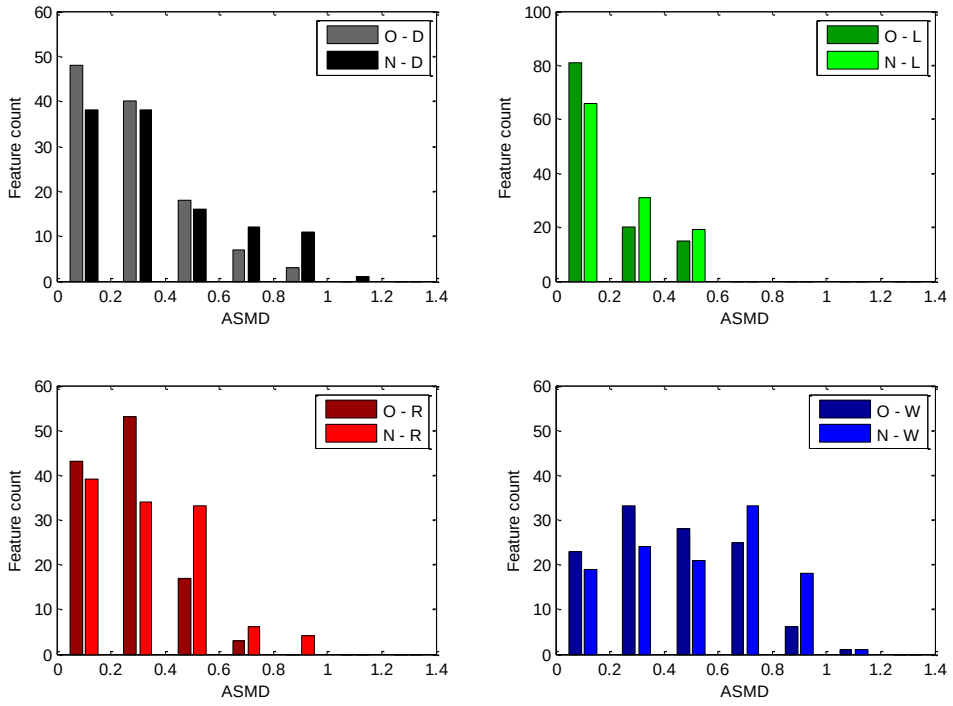


Figure 26. Comparison between Original (O) and Normalized (N) feature set with histograms indicating the number of feature for different ASMDs

More importantly, however, is that the number of features with higher discriminatory power increased for all classes, even for light sleep although not very pronouncedly. For deep sleep and wake, the number of features with ASMD above 0.8 more than doubled. For REM, the number of features above 0.6 also increased with some of the features having an ASMD over 0.8.

4.4 Feature transformation

Feature normalization achieved very good results, decreasing the subject variability which improved the pooled quality of the features and consequently improved the features discriminatory power for all classes. Nevertheless normalization was only concerned in reducing between subject variations. In this section feature transformation will be explored. In this work, feature transformation or post processing will refer to a combination of methods that aim at transforming a feature into a different one in order to maximize the discriminatory power.

4.4.1 Motivation

Linear discriminants are very simple classifiers capable of good generalization. However complex relations and statistical information on the data might be disregarded for the classification task, since the classifier is unable to use this information. Applying some transformation methods to the original features, might extract this information in a recognizable way for the classifier, which in turn will use these newly created features for classification purposes. Furthermore if this new information is able to increase the number of features with a high AMSD for some classes, it's very likely that this step will be noticeable upon examination of the next chapters concerning the classification results.

Linear discriminants are very simple classifiers capable of good generalization. However complex relations and statistical information on the data might be disregarded for the classification task, since the classifier is unable to use this information. Applying some transformation methods to the original features, might extract this information in a recognizable way for the classifier, which in turn will use these newly created features for classification purposes. Furthermore if this new information is able to increase the number of features with a high AMSD for some classes, it is very likely that this step will be noticeable upon examination of the next chapters concerning the classification results.

One of the most noticeable incentives for the transformation of the features was the known differences in physiological variations within each sleep stage. In fact variations in cardiorespiratory signals are correlated with different stages of sleep[52], [53]. These variations are observable in healthy subjects although similar differences between variations in stages are encountered for unhealthy subjects as well[54–56]. These variations are usually bigger in Wake, N1 and REM and tend to be lower in deeper sleep namely in N3.

4.4.2 Techniques

In this work the applied transformations will explore statistical properties of the datasets as well as apply some filters to the original signal.

4.4.2.1 Statistical techniques

The statistical techniques consist in methods to analyze the variations of the features. In literature there are several reference describing how the variability in the cardiorespiratory features varies with the stages of sleep. The purpose of these methods is to evaluate the possibility of finding relevant information in the current normalized features, regarding the variability described in the literature. The concept is to investigate the variations over time of the feature values. This is performed by evaluating the variability within a sliding windows of time. The two methods available for this purpose are the Std (Standard deviation) and the MAD (Mean absolute deviation), computed with the formulas presented below.

$$std = \left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{\frac{1}{2}} \quad (6)$$

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}| \quad (7)$$

Where n is the size of the window and $\bar{x}$ the mean of the values in the window. For each epoch, the Std or the mad is computed for the specific window. In this work two window options were studied. The backwards option, where the values used for the calculation are after the epoch whose value is being calculated and the central option in which the window of values is centered in the epoch which value is being computed.

For each epoch, a value of variability is computed based on one of the above methods. The window changes accordingly with the epoch being calculated. The values are based on the normalized original feature value. As an example, the feature values of feature 13, from a given subject are visible in Figure 27. In Figure 28 the std method, with a backward window of 23 epochs, was applied to the feature values.

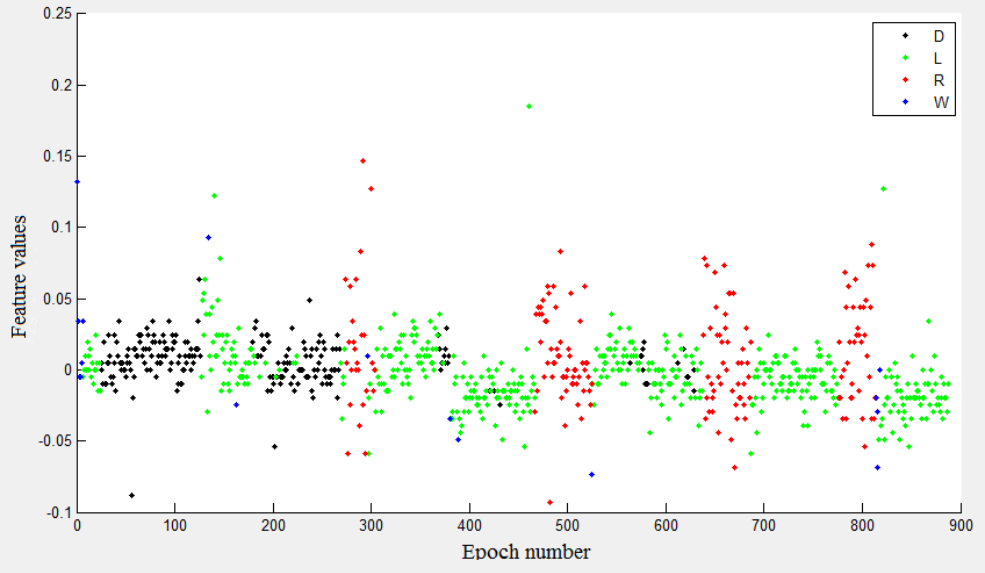


Figure 27. Example feature values for feature 13 for a given subject

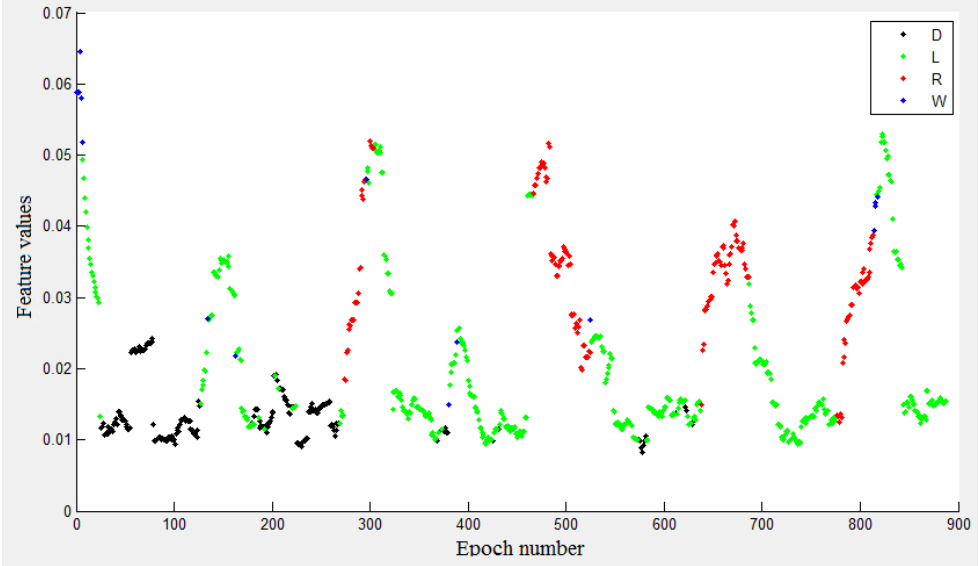


Figure 28. Example feature values for feature 13, for the same subject, after the application of the std method

In order to find best window sizes for the analysis of the variability in the features, an analysis of the way the discriminatory power varies with the variation of the window sizes was performed. In order to understand the effects of low frequencies variations over the feature data, a median filter was investigated as well.

The methodology used was to apply a median filter with different window sizes, in order to remove any trends in the data, perform the Std method, again with several window sizes, and lastly execute a Logarithmic transformation.

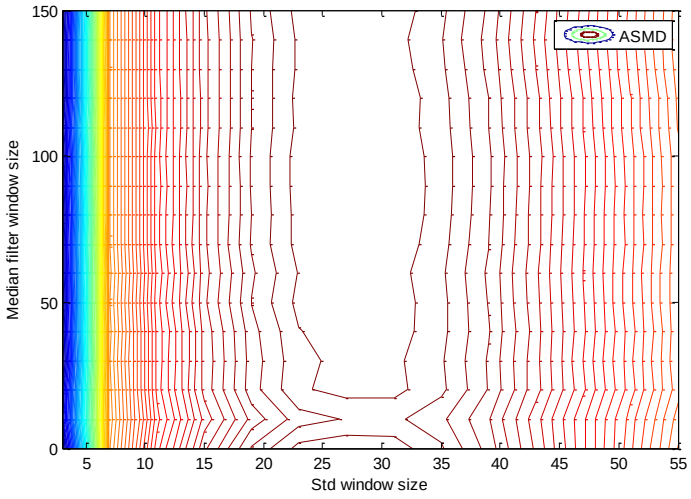


Figure 29. Contour lines of the ASMD of feature 13 for Deep sleep, varying the size of the median filter window and the std window

In Figure 29, it is visible that the variation in the discriminatory power for deep sleep of feature 13, a respiration based feature, varies greatly with the size of the window in which the variation is being computed. The blue section represent low values of ASMD, while darker red represents the biggest values of ASMD. In terms of changes in the low frequency domain, the discriminatory power did not suffered great alteration which makes the application of the median feature rather irrelevant no matter the chosen size of the window. However, as witnessed by the figure presented beneath, with features based on the ECG, like feature 27, ecg_rr_mean, the application of the median filter with a small window size produces very good results.

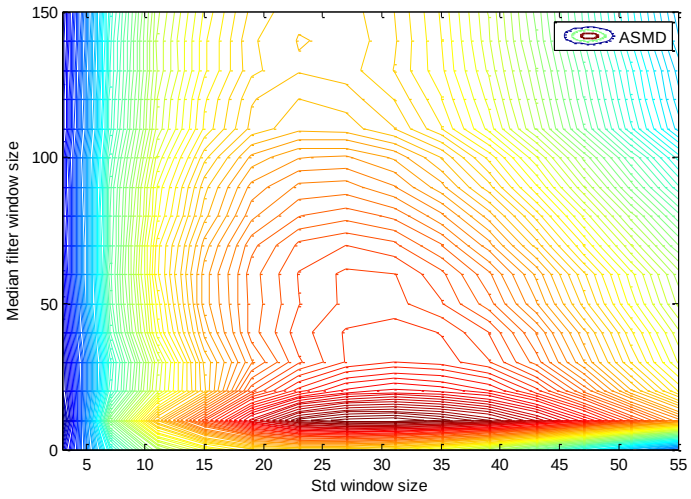


Figure 30. Contour lines of the ASMD of feature 27 for Deep sleep, varying the size of the median filter window and the std window

Undoubtedly, Figure 30, the need for the median filter is obvious. Furthermore, although not visible in this figures, changing the feature or the class would create new optimal setting. With this problem in mind, the need for specific transformations for different classes and features is understandable. Regardless this conclusion there is some information left in these pictures. If a closer look is given, it's quite clear that the size of the window in which the variation is being studied has relatively flexible maximum values. For instance in Figure 29, to achieve the maximum ASMD, a variation window for the std must be between 23 and 33. This is important for the search process as it's possible to set a small collection of predefined values to search, which makes the search much more efficient, without losing much discrimination power.

With the progress of the work the median filter was eventually substituted with a high pass filter that was able to provide similar and in some cases better results.

4.4.2.2 Smoothing techniques

The smoothing techniques compromise a high pass filter and a low pass filter.

In this work the low pass filter is successfully applied to very noise signals as this method is able to remove it.

The high pass filter is used to remove the variation in the low frequency domain, which in turn can make the variance analysis more efficient as it will be displayed in the next sections.

4.4.3 Combining techniques

As a way to combine this different technique into a successful approach, there is a need to define how to combine this different techniques.

Starting from the normalized features, using the methodology described in the previous section, it is possible to perform a statistical analysis of variation or a smoothing operation. If the smoothing operation is performed the transformation process can stop, or a statistical analysis of variation can be applied.

After the application of a statistical method there are three possible scenarios. The transformation can be complete, or the feature values can be transformed further. In the last case, a logarithm may be applied to the values of the statistical analysis of variance as a way to make the distribution of the data more similar to normal, method used in other analysis of variance studies [57][58][54]. From the statistical or log transformation, the data might need to be normalized again. In fact since the statistical methods assess the variation of the features, different subjects might have different variations, which produce between subject variability, and therefore justifies the need for normalization. The normalization is performed in the same manner as it's described in Figure 31.

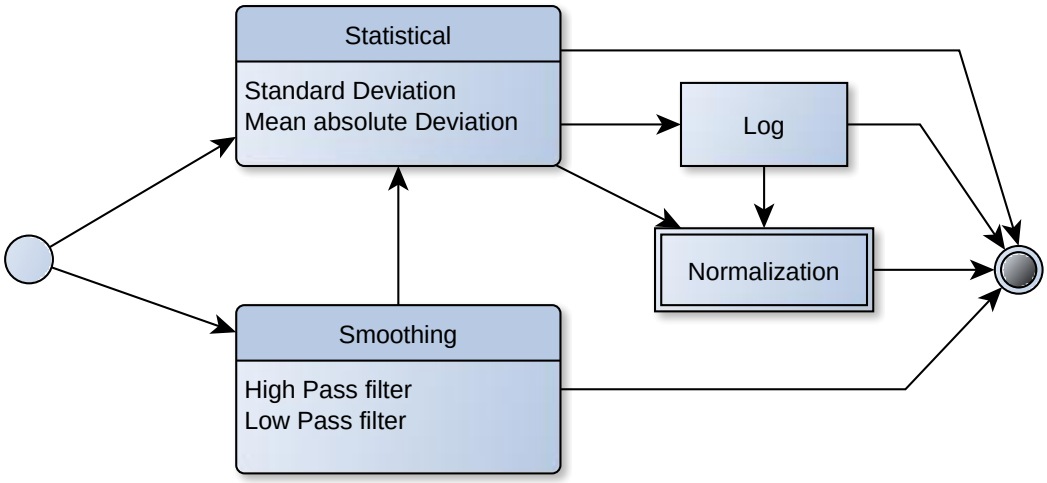


Figure 31. Transformations search graph

4.4.4 Optimizing best transformations for each class

As the transformations have different optimal parameters for each class, the transformation techniques are going to be computed per class in order to maximize de ASMD per feature per class. In the end of the process the final dataset increases linearly in relation to the number of classes as each feature will be added with the best normalization for a determined class.

To find the best combination for each feature for each class the graph represented in the figure above will be used as the search graph for the breath first search algorithm. In order to evaluate each node of the graph the ASMD metric was used. As not all transformations are useful for all features, and a lot of different combinations are present, note that the statistical methods have several different configurations to accommodate the different optimal parameters with different features and classes, a greedy search was implemented. This search has a look ahead of two nodes and only searches for the best configuration of the statistical methods.

With this method the search for the best transformations for each class and for each feature was hastened significantly. At the end of the process, for each class, an xml containing the best transformations, with the designated configurations, was created in order to store the newly attained results.

4.4.5 Transformations results

After the best transformation was found for each class, for each feature, the next step is to apply the transformations on normalized features and add the newly created features to the dataset. As each feature is going to have a transformed version for each class, the inherent result of this method is the growth in the total number of features. The number of features will increase linearly with the number of classes, four in this study, which will result in a dataset with five times more features than the original one.

The large number of features is a problem that will be dealt with at classification time by the feature selection algorithm. In this section the aim is to create features with high discriminatory power in order to improve the results of the next steps of the classification process.

Figure 32 compares the number of features according to different ASMD values for the post processed dataset with the normalized features and the transformed features, the normalized dataset and the original dataset.

Through observation of this figure, the increase in the number of features with high discriminatory power is very substantial for deep sleep, REM and wake. For light sleep the available transformations were unable to create features capable of good discrimination between this class and the others.

For deep sleep, the number of features with an ASMD higher than one increased from very few to nearly fifty features. Furthermore there are now features with an ASMD over 1.2. For REM the increase of features with high discriminatory power was not so expressive, however with the transformation step, the features with an ASMD over 0.6 are now abundant, and there are about twenty features with an ASMD bigger than 0.8. For wake the increase in features with high discriminatory power is very considerable. Without the transformation step, there is only one feature with an AMSD over 1. After transformation, there are around fifty features with an ASMD higher than one.

The post-processed dataset, the dataset with normalized features with the added features from the transformation step, is going to be the dataset used for classification. The epochs

which had invalid feature values, which for normalization and transformation purposes were interpolated, are going to be removed.

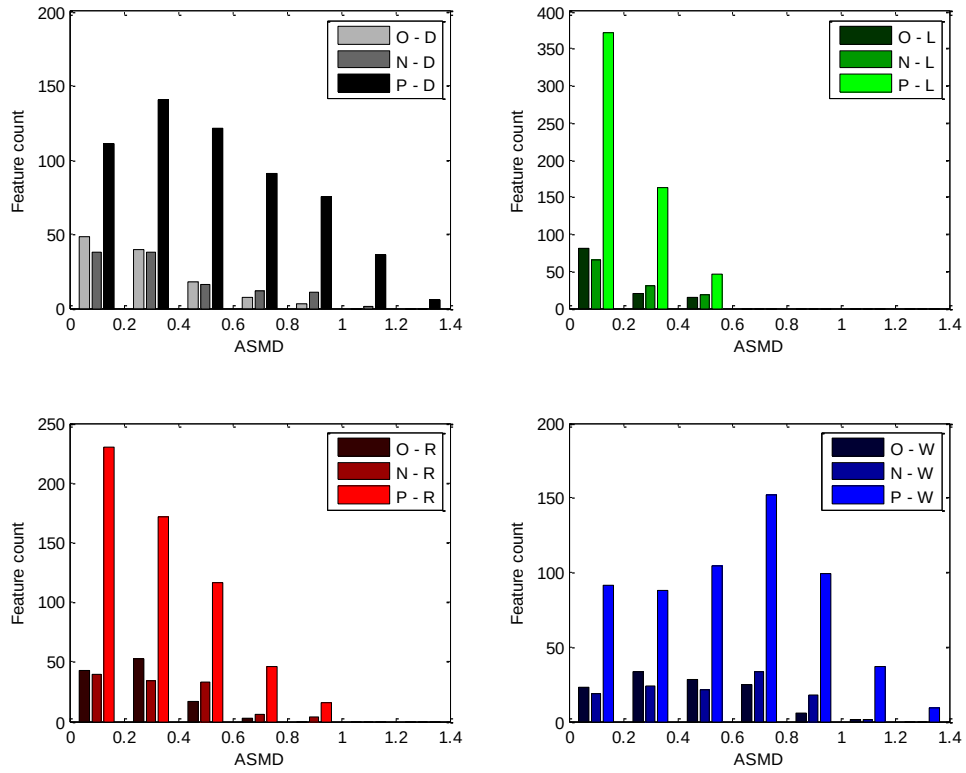


Figure 32. Comparison between Original (O), Normalized (N) and Post-processed (P) dataset with histograms relating the ASMD with the number of features per class

4.5 Normalizations and transformations on subjects with insomnia

The best normalization and transformation techniques were found using the current dataset and the ground truth associated to it. This fact might introduce some bias in the classification results. However this bias should be small as this process was performed only using the ASMD as a performance measure. In fact the purpose was to maximize the discriminatory power of the features. The increases in the results of classification process are a consequence of the increase of the information of the features.

In order to evaluate if the methods are able to generalize on a different dataset with subjects with different sleep architectures, a dataset with subjects diagnosed with insomnia was used.

The normalizations and transformations found on the previous datasets were applied to the dataset of insomnia subjects. Figure 33 shows the results of normalization and transformation.

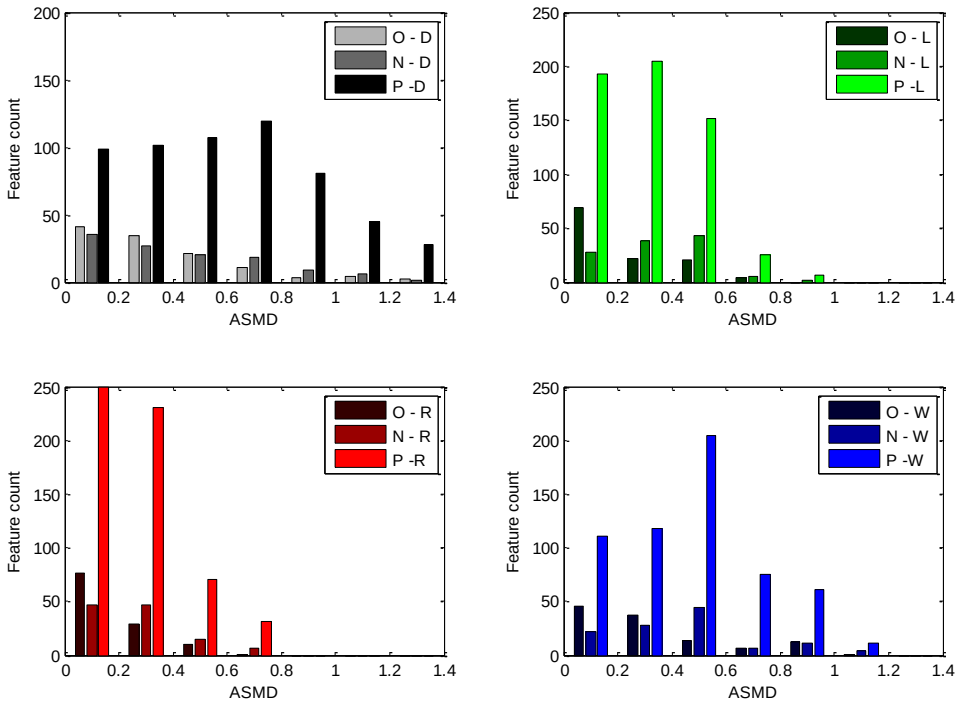


Figure 33. Comparison between Original (O), Normalized (N) and Post-processed (P) dataset with histograms relating the ASMD with the number of features per class for subjects with insomnia

Through the observation of the figure, it is noticeable that normalization once again proved very fruitful, with all the classes increasing the number of features with higher discriminative power. The transformation step also added very discriminative features to the dataset. Nevertheless very significant changes are visible regarding the discriminatory power of the features per class for the healthy subjects compared with the subjects with insomnia. These differences are likely related to the different sleep architecture and class distribution of the datasets. For instance, the insomnia subjects have substantially less deep sleep which may explain why the number of features with higher discriminative power deep sleep and light sleep increased. Furthermore the overall instability of sleep of

subjects with insomnia and the bigger percentage of wake epochs might explain the decrease in discriminatory power of the features for REM and Wake. Nevertheless the normalization and transformation of the data definitely improved the number of higher discriminative features for all classes.

The increase in the number of features with high discriminatory power suggests that the normalization and transformation techniques are useful and appropriate for subjects with a different sleep architecture than a healthy subject.

4.6 Conclusions

The normalization and transformation techniques increased the number of features with high discrimination power and therefore increased the data quality. When compared against the original data, the increase of the number of features with high discrimination power for all classes is very significant. Normalization techniques reduced between subject variability while transformation maximized the number of features with a high discrimination power. The results of the application of this techniques were discussed in this section with a comparison with the results for non-healthy subjects.

5 Classifier

In this section, the process of classification and validation of the data will be reviewed.

The chapter starts with a section regarding Bayesian decision theory followed by the introduction of the Bayesian linear discriminant, which is the classifier used in this work. Linear discriminant is a very fast and simple classifier, capable of good generalization. Unaffected by class imbalance this classifier is suited for this specific problem. Furthermore, other studies regarding multiple sleep stage classification have used it as well [37].

Feature selection, and time varying prior probabilities calculation and integration on the formula output are described in this section as well. In the end a system overview of the classification procedure is presented, with an explanation on the procedure for the validation of the results.

5.1 Bayesian decision theory

The basic idea in Bayes decision theory is to minimize the overall risk, of making a wrong decision, by always choosing the action that minimizes the conditional risk $R(\alpha|x)$. In other words, in a classification problem, the state that maximizes the posterior probability $P(\omega_i|x)$, should always be chosen so that the probability of error is minimized.

For example, given two classes ω_1 and ω_2, and an unknown pattern represented by a feature vector x, the conditional or *a posteriori* probabilities $P(\omega_i|x)$ represent the probability that the unknown pattern belongs to the respective class ω_i given that feature vector x. If $P(\omega_1|x)$ and $P(\omega_2|x)$ are determined, the decision with the lowest classification risk is:

$$If\ P(\omega_1|x) > P(\omega_2|x)\ \ then\ x \epsilon\ \omega_1\ else\ x \epsilon\ \omega_2 \qquad (8)$$

Bayes' formula allows such probabilities to be computed from the prior probabilities $P(\omega_i)$ and the conditional densities $P(\mathrm{x}|\omega_i)$. Using Bayes' rule, the conditional probabilities can be computed as:

$$P(\omega_i|x) = \frac{p(x|\omega_i) \times P(\omega_i)}{p(x)} \tag{8.1}$$

$$p(x) = \sum_{i=1}^{c} p(x|\omega_i) \times P(\omega_i) \tag{8.2}$$

The posterior probabilities $P(\omega_i|x)$ can be computed if the class-conditional probability density functions (pdf), of the distributions of the feature vectors, for both classes, are known. The term $p(x)$ is equation 8.1 is the same for all classes so it is possible to write:

$$If\ p(\mathrm{x}|\omega_1) \times P(\omega_1) > p(\mathrm{x}|\omega_2) \times P(\omega_2)$$
$$then\ \mathrm{x}\ \epsilon\ \omega_1\ else\ \mathrm{x}\ \epsilon\ \omega_2 \tag{8.3}$$

This classification test is in fact equivalent to minimizing the classification error probability. Considering equal prevalence for both classes, $p(\mathrm{x}|\omega_1)$ and $p(\mathrm{x}|\omega_2)$ with R_1 as the feature space region where the decision favors ω_1 and R_2 as the region where the decision favors ω_2, the Figure 34 can be created.

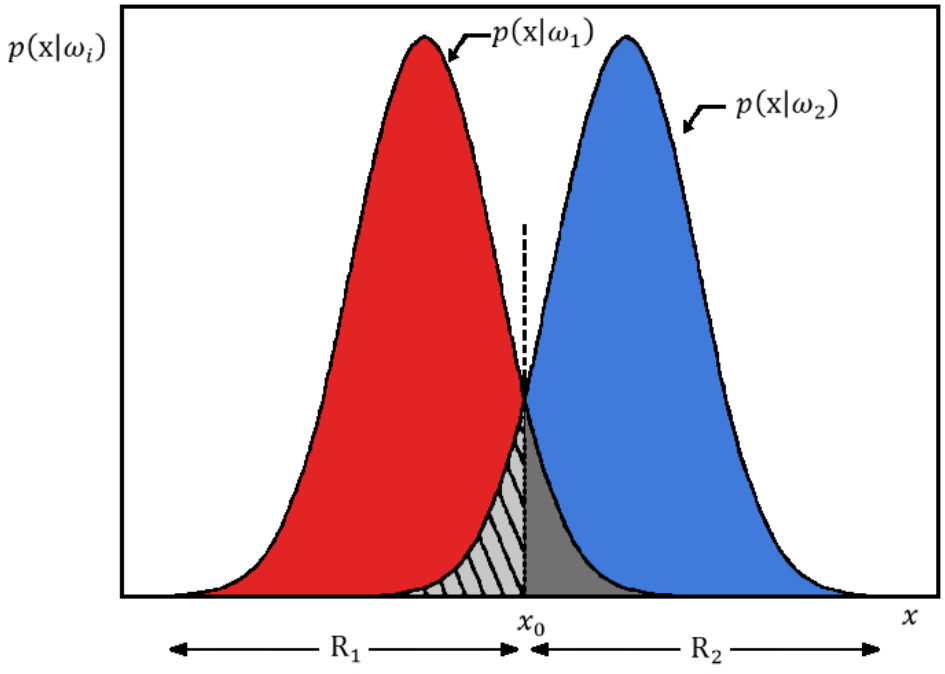


Figure 34. Example of two regions R1 and R2 formed by the bayesian classifier on probability density functions with equal prevalences for both classes

An error occurs whenever $x \in R_1$ although it belongs to ω_2 or when $x \in R_2$ although it belongs to ω_1. The risk is minimized if R_1 is the region of space in which $p(x|\omega_1) > p(x|\omega_2)$ and R_2 is the region where the reverse is true. Since the class prevalences are equal, the decision threshold is at half distance from the means of the classes. The number of misclassifications using the rules above is proportional to the area of the shaded areas and is equal for both classes. In the case of unequal class prevalences, the decision threshold is displaced towards the class with smaller prevalence, therefore decreasing the number of errors of the class with higher prevalence.

5.2 Bayesian linear discriminant

Assuming the normal distribution, the likelihood for class ω_i is given by the pdf in equation 9 where μ_i and Σ_i are the theoretical or true mean vector and covariance matrix for class ω_i respectively.

.

$$p(x|\omega_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|}\exp\left(-\frac{1}{2}(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i)\right) \quad (9)$$

In order to simplify the computation of *a posteriori* probabilities, a discriminant function $g_i(x) = P(\omega_i|x)$ is used. Using the monotonic logarithm $\ln\big(P(\omega_i) \times p(x|\omega_i)\big)$, in order to eliminate the exponential term in $p(x|\omega_i)$ the discriminant function can be further simplified,

$$g_i(x) = P(\omega_i|x) = p(x|\omega_i) \times P(\omega_i) \quad (9.1)$$

$$h_i(x) = \ln\big(g_i(x)\big) = \ln\big(P(\omega_i) \times p(x|\omega_i)\big) \quad (9.2)$$

$$h_i(x) = -\frac{1}{2}(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i) \quad (9.3)$$

The results of this function is often called the discriminant score for the i^{th} class. The first term on the right-hand side of the $h_i(\mathrm{x})$ is also called the Mahalanobis distance between x and μ_i. This discriminant function separates the regions of the feature space with a quadratic boundary. For that reason it is called a quadratic discriminant. When the covariance matrices can be assumed to be identical, the discriminant function can be simplified further, resulting in a linear discriminant function,

$$h_i(x) = -\frac{1}{2}(x-\mu_i)'\Sigma^{-1}(x-\mu_i) + \ln P(\omega_i) \qquad (9.4)$$

In the case of the quadratic discriminant there is a need to compute the covariance matrices for all classes, whereas in the case of the linear discriminant, it is assumed that all classes share the same covariance matrix. In both cases, the *a priori* probabilities $P(\omega_i)$, which is usually a constant, can be provided by the user or deduced from the data. Quadratic and linear discriminant functions can be expected to work well if the class conditional densities are approximately normal and good estimates can be obtained for the population parameters defining the distributions, namely the class mean vectors and covariance matrices. Since quadratic discriminants generally require larger sample sizes than linear discriminant and seem to be more sensible to violations of the basic assumptions [41], linear discriminants functions are used. Furthermore, similar studies achieved better results using linear discriminants than using quadratic discriminants [37]. Robustness to the violations of the assumptions are desirable since not all features follow a normal distribution.

For a two-class problem, assuming the same covariance matrix for both classes, the linear discriminant $d(x) = h_1(x) - h_2(x)$ is easily computed as

$$d(\mathrm{x}) = \mathrm{w}'x - w_0 \qquad (9.5)$$

$$\mathrm{w} = \Sigma^{-1}(\mu_1 - \mu_2) \qquad (9.6)$$

$$w_0 = -\frac{1}{2}\left(\mu_1'\Sigma^{-1}\mu_1 - \mu_2'\Sigma^{-1}\mu_2\right) + \ln\frac{P(\omega_1)}{P(\omega_2)} \qquad (9.7)$$

5.3 From binary classification to multiple class

Multi-class classification can roughly be divided into two groups. The first group consists of classifier algorithms that can naturally handle multi-class cases. Examples of this types of algorithms are nearest neighbors [59] or regression and decision trees such as C4.5[60] or CART (Classification And Regression Tree) [61]. The second group comprises methods that reduce the multi-class classification problem into binary cases. These methods, depending on the approach used, can be further divided into one-versus-the-rest [62], [63], pairwise comparison [64–66], direct graph traversal [67], error-correcting output coding [68], [69] and multi-class objective functions [70]. The first three methods are the most applicable for the linear discriminant case, but considering that studies indicate that the performance of the three methods are very similar and no one method is statistically better the others [71] , one-versus-the-rest , the simplest method, was chosen. Moreover linear discriminants were used successfully in previous studies multiple class sleep stage classification [37].

Since different features were suitable for the discrimination different classes, choosing an overall set of features that would provide good discrimination power for all the classes would prove to be a hard task. On the other hand there were some classes that were known, from previous work, to discriminate well in a one-versus-the-rest setup. Using this information, a linear discriminant classifier was created for each existent class. Each classifier was trained in a one-versus-the-rest setup, with the target class corresponding to the positive class and the rest of the classes merged in the negative class. Although the vector x_i for each discriminant is comprised of different features, a direct comparison of the discriminants scores can be performed as the inverse covariance matrix of the linear discriminant function standardizes the features and therefore the distances and scores. This is very important as it allows direct comparison between different linear discriminants scores, which is very useful to establish the final classification.

5.4 Feature selection

Feature selection is method to evaluate and select the most appropriate features, for a certain classification task, in order to achieve better classification performance.

As visible in Figure 35 there are two main methodologies for feature selection: a filter and a wrapper approach. Wrappers methods use the classifier to evaluate the performance of different sets of features according to some performance or error metric. This might be a very time-demanding task, especially when the number of features is large. Filter models rely on the characteristics of the training data to select features such as relevance, discriminatory power, redundancy and correlation between features.

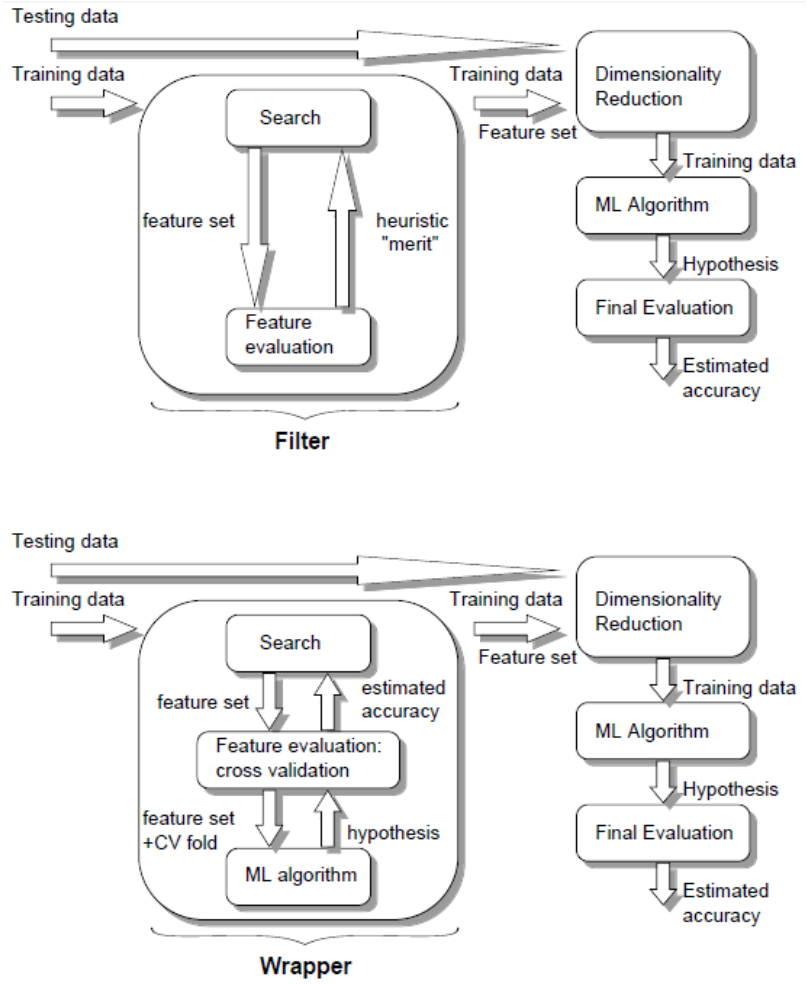


Figure 35. Filter and wrapper feature selectors [72]

In this work, since the number of features is considerably large (580 features), the wrapper approach takes an intolerable amount of time to proceed with the selection of the best

combination of features. One important aspect to be considered when selecting features is the correlations between the different features. According to empirical evidence from feature selection literature, features with low discriminatory power and redundant information should be minimized [73–75]. A feature is considered redundant if one or more of the other selected features are highly correlated with it. After transforming the features, as explained in Section4, it is expectable that there are highly correlated features, which do not add information and therefore should be removed from the final set of selected features. So the filter algorithm should take in consideration the relevance of the features individually as well as their redundancy.

Feature selection was performed with the Correlation Feature Selection (CFS) algorithm described by M. Hall [72].This method is a fast and simple method to perform feature selection, which takes into consideration the discriminatory power of the features individually as well as the correlation between the features.

This iterative method selects the features that are correlated with or which are predictive of the positive and negative classes, avoiding features which are highly correlated features selected in previous iterations. For each iteration, CFS computes a heuristic measure of "merit" of a feature subset from pairwise feature correlations and a formula adapted from test theory

$$M_S = \frac{k\overline{r_{fc}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}} \tag{10}$$

where M_S is the heuristic "merit" of a feature subset S containing k features, $\overline{r_{fc}}$ is the mean feature-class correlation, and $\overline{r_{ff}}$ is the average feature-feature intercorrelation.

The numerator of this equation provides an indication of how predictive of the class a set of features are, whereas the denominator indicates how much redundancy there is among the features.

A greedy sequential forward selection search is used to traverse the space of feature subsets with the subset with the highest merit found during the search being reported.

The stopping criteria of the method is attained once the addition of any unselected feature can no longer improve the merit of the subset of selected features.

Although CFS discretizes all continuous features in the training data, in this work, since discretization of a big amount of features is either a very slow process or a process where there is a big amount of information lost, discretization was not performed. Instead, in order to evaluate the correlation between the ground truth and features, the absolute value of Pearson linear correlation coefficient was used. This coefficient is able to compute the correlation between continuous variables and therefore the discretization step of the CFS algorithm can be skipped without significantly degrading the results.

The fact that features in this study are continuous had a great impact on the decision of the feature selection algorithm. Popular algorithms for feature selection based on mutual information, such as MRMR [76], are hard to compute and time inefficient as it is often difficult to compute the integral in the continuous space based on a limited number of samples.

5.5 Temporal stage changes

As discussed in Section 2.2.1, sleep follows a certain repeatable pattern on most healthy subjects. It has been shown that the probabilities of each sleep stage occurring vary from the moment lights are turned off, until the subject wakes up in the morning. Redmond et al. [38], in an attempt to track the non-stationary behavior of sleep, suggested that the use of time dependent *a priori probabilities* $P(\omega_i|t)$ can improve the classification performance. This information can be used in order to improve the results of classification. In this work, the non-stationary behavior of sleep is further explored in order to add meaningful information regarding sleep architecture to the classification results.

In this section a new method is proposed, which extends the sleep stage probability information of the time dependent *a priori* probabilities proposed by Redmond et al. [38]. The method use the data, to capture the changes in sleep stages in each consecutive epoch for each subject. Saving information regarding temporal changes in sleep stages allows the posterior calculation of temporal prior probabilities as well as the time varying probabilities of transitions between stages. Time varying probabilities of transitions between sleep

stages can help to improve classification results as this information can use the already classified epochs to provide statistics regarding sleep stage changes in the next neighbor epochs.

For instance it is unlikely that a subject who, in a certain epoch in the beginning of the night was classified as being in REM sleep, to be in deep sleep after one or two epochs. Since time dependent *a priori* probabilities only have time into consideration, and deep sleep is likely to occur in the beginning of the night, the *a priori* probability of deep sleep would be high. With the addition of the time varying probabilities of transitions between stages, previously classified epochs will be taken into consideration and so, for this example, deep sleep would have small *a priori* probability.

As a way to model this information and store the stage changes along the length of the night, a three dimensional matrix is created which is called temporal stage changes. The temporal stage changes, has information relative to the epoch of the night, original sleep stage and the next sleep stage. In other words, this matrix is composed by several squared sub-matrix, each having information on the number of sleep stage changes during a certain time of the night. The numbers inside the cells of these sub-matrixes are ordinary counters that express the total amount of stage changes from one stage, in its corresponding row, to some other stage, in its corresponding column. An example of such matrix can be seen in Figure 36.

In order to process this matrix, an iterative process is followed in which each subject is processed individually. The temporal stage changes is complete when all subjects have been processed.

The first step for the insertion of stage changes for a certain subject, is to compute the specific subject's constant. This constant is calculated based on the night length of the subject, last subject's epoch number (SLEN), and the predefined nigh time (PNL) of 8 hours, which corresponds to 960 epochs. The aim of this constant is to normalize the different night lengths of the subjects, in order to obtain a more accurate description of

the likelihoods of the different sleep stages across the night. The formula to compute the subject's constant (SC) is given by:

$$SC = \frac{PNL}{SLEN} \tag{11}$$

As having 960 matrixes of stage changes, one sub-matrix per epoch, would be too specific and redundant, the number of epochs per sub-matrix (NESM) was set to 4 epochs, which makes this structure, by default, a set of 240 sub-matrixes which will be referred to as the total number of sub-matrixes (TNSM).

Because the night length of the subjects is rarely the same as the predefined night constant, a formula to scale the subject's night length and correctly input the stage changes of an epoch into the correct sub-matrix number (SBN) is needed.

The sub-matrix number of an epoch e of a given subject is given by:

$$\mathrm{SMN} = 1 + \mathrm{floor}\left(\frac{\mathrm{e} \times \mathrm{SC} - 1}{\mathrm{NESM}}\right) \tag{12}$$

The function floor(X) is a function that rounds X to the nearest integer less than or equal to X.

To achieve better results, generalizing the data obtained and smoothing the final probability values, each sleep stage change is inserted with 90 epoch (45 minutes which is approximately half of the REM Non-REM cycle) tolerance (T). This step is needed in order to compensate for the lack of subjects in this study. This tolerance will create a smoothening effect much like a low pass filter. Although this tolerance is applicable to most of the night, in the beginning of the nigh as well as in the end, this tolerance is lower in order to avoid losing information. To determine the initial sub-matrix number (ISNM) and the final sub-matrix number (FSMN) the following formulas are applied:

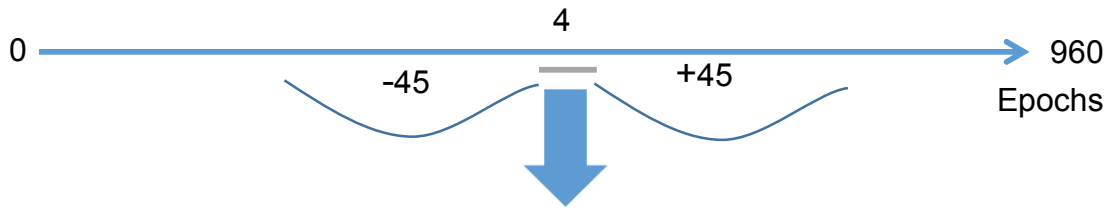
$$\mathrm{ISMN} = 1 + \mathrm{floor}\left(\frac{SMN}{2}\right) \tag{13}$$

$$\text{FSMN} = \text{TNSM} - f\text{loor}\left(\frac{\text{SMN}}{2}\right) \quad (14)$$

This calculation allows the epochs near the beginning and the end to be less affected by the sleep stage changes of the epochs referring to the middle of the night. On the other hand for epochs in the middle of the night, the tolerance provided by these formulas is too high. To correct this, a tolerance limit (TL) is established for tolerance of initial and final sub-matrix number, which is given by:

$$\text{TL} = \frac{\text{T}}{2 \times \text{NESM}} \quad (15)$$

In Equation 15, the constant two, divides the tolerance between the upper limit and the lower limit.



Next / Original	N12	N3	REM	Wake
N12	1091.2	60.2	16.4	18.6
N3	52.3	351.6	0	1.3
REM	12.6	0	384.9	5.6
Wake	33	0	0	36.2

Figure 36. Sub Matrix of Temporal Stage Changes being processed

After calculation of the initial and final sub matrix, in which the sleep stage change should be recorded to, the cell of the matrixes corresponding to the sleep stage change should be incremented by the subject's constant. This step is to ensure that sleep stage changes from

subjects with a bigger night length do not get more significance than sleep stage changes from subjects with a lower night length.

5.6 Temporal *a Priori* probabilities calculation

The temporal *a priori* probabilities are calculated from the information in the previously computed temporal stage changes. In order to perform the calculation of the *a priori* probabilities for each class, the epoch number and the number of total epochs of the subject's nights are needed. At first, the subject's constant is calculated based on Equation 11, and then the sub matrix number is found using Equation 12. With the correct sub matrix number, it is possible to sum the values of the column's corresponding class and divide by the sum of all the values of the matrix. When looking at all the epochs of the night, assuming predefined nigh time, the time varying prior probabilities for all the classes will take the values visible in Figure 37.

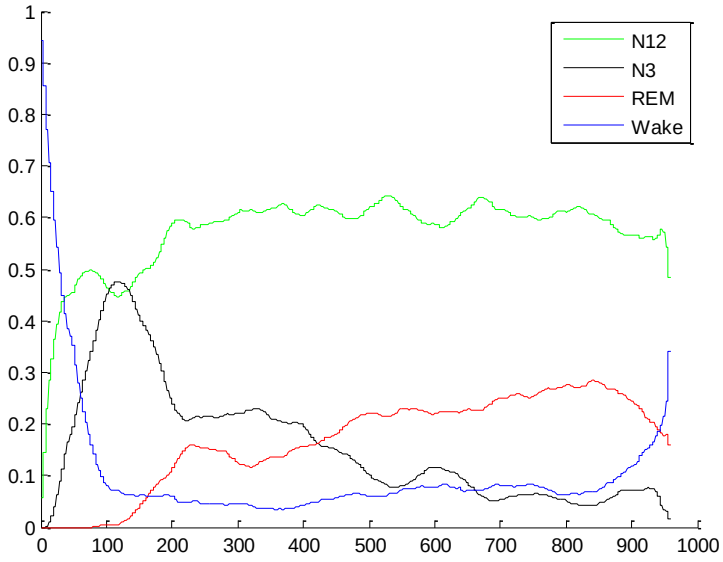


Figure 37. Temporal a Priori Probabilities for all classes

5.7 Stage change probability and Markov Chain

As the epochs start to be assigned a final classification, it is possible to check which class has a higher discriminant value, and therefore select it as the most likely class. Assuming that this class is correct, the *a priori* probabilities of the next classes can be updated based on the time varying probabilities of transitions between stages. This probability can easily

be computed using the correspondent row of the correspondent sub matrix number of the epoch. Using the information on that specific row, it is possible to find the probabilities of the next class based on the currently selected class. The problem with this method is that it assumes that classified epoch was assigned the correct class, which might not be the case, which by turn introduces wrong information on the probabilities.

In the specific case of having a discriminant with a considerable higher output than the rest (when the higher discriminant output minus the second best discriminant output is higher than a certain threshold) the probabilities of the epoch being incorrectly classified are low. In this case, which is going to be referred to as a confidence point, the probabilities of a class being correctly classified are increased. When an epoch with these characteristics is found, and the next epochs are not confidence points anymore, a first-order Markov Chain is applied to supplement the time varying probabilities. When another confidence point is found, the Markov Chain method stops, restarting when a confidence point is followed by a non-confidence point.

The transition matrix is of the Markov process is obtained from the sub matrix from the temporal stage changes, according to the epoch and the last epoch of the subject's night. As the transitions grow, so does the epoch and so, the transition matrix can be updated according to the sub matrix that the analyzed epoch corresponds to. This way the Markov chain is applied to small segments of the night, which means the transition probabilities do not change significantly, and therefore the Markov stationary process assumptions, which states that the transition probabilities are independent from the time in which the transition takes place, is respected.

5.8 Combination with Linear Discrimination formula and output

The introduction of the probabilities in the discriminant formula is quite simple. In fact discriminant functions have the *a priori* probability variable $P(w_i)$ in the formula which is usually used as a constant and for instance can be used to compensate for problems like class imbalance.

$$h(i) = \frac{1}{2}(\mathrm{x} - \mu_\mathrm{i})' \sum^{-1} (\mathrm{x} - \mu_\mathrm{i}) + \ln P\,(w_i) \tag{16}$$

In this work as shown above, this variable will be computed based on the time of the epoch and, if present, with the information of the nearest confidence point. To achieve the final value of $P(w_i)$, the temporal prior probability value for class *i* is multiplied by the probability given by the Markov chain assuming the starting point to be the last confidence point.

5.9 Post processing and establishing final classification

The classification process is done gradually, subject by subject, starting with the first epoch until the last one, with each epoch being classified by the class which linear discriminant provides with a higher score.

After the calculation of temporal varying prior probability for all epochs, it is necessary to know how much weight the prior probabilities should have in the final classification results. Furthermore it is necessary to discover and optimize the threshold for the confidence points. To access this questions a study was done in order to understand how the final classification differs with the variation of these variables. The method was to use the results of the discriminant functions with different values of weight of prior probability as well as the values for the establishment of confidence points in order to access the variations in the final result of the averaged kappa coefficient of agreement.

The results can be observed in Figure 38. The blue values correspond to lower values of the kappa coefficient of agreement while the darker red represent the higher values. Clearly, there is a maximum value for the kappa for the values of the weight of prior probability close to 0.4 and with the values for the threshold for the confidence point around 1.

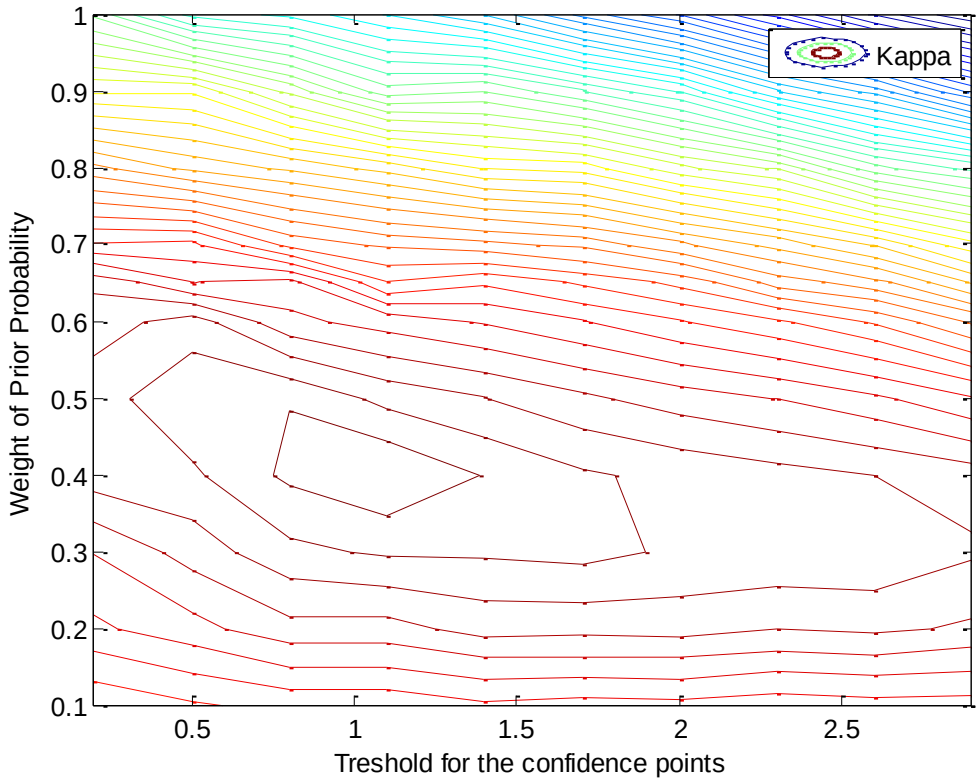


Figure 38. Averaged kappa coefficient of agreement for different weight values of prior probability and threshold values for the confidence points

5.10 Validation

In order to validate the results of the classification procedure, and access how the classifier will perform on an independent dataset, ten-fold cross validation scheme was used.

In ten-fold cross validation, the dataset is divided into 10 equal subsets. Then, iteratively, one of the subsets is used for testing, the testing data, and the remaining nine for training, the training data. The training set is used for feature selection, and to build and train the classifier model. After the classifier has been trained, it is used with the features from the testing data, to predict the class of the newly received data. Based on this prediction and the ground truth from the testing data, it is possible to validate the model and measure several performance statistics of the classification problem. Once the testing data has been classified, another part is selected to be used for testing, and the classification process

starts process starts again. This process is repeated until all parts have been used as the testing data.

The classification procedure can be observed in Figure 39.

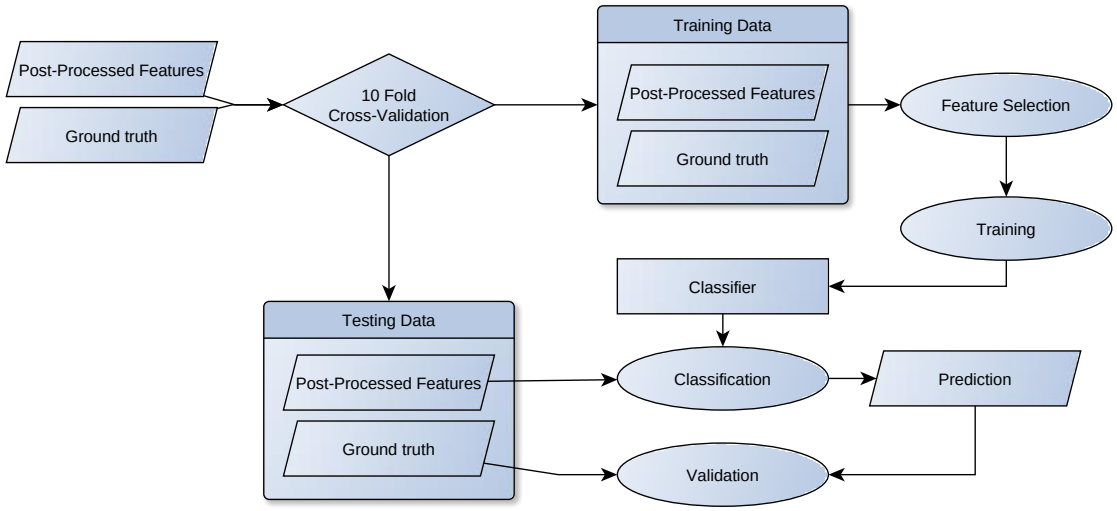


Figure 39. Diagram of the Classification procedure

5.11 Conclusions

The section described the classifier, the feature selection algorithm and the probabilistic post-processing used to improve the classification results as well as to establish final classification. The probabilistic post-processing aims at capturing the non-stationary temporal characteristics of sleep. Temporal prior probabilities calculation, as well as their application in the classifier output was described here. In the end of the section, the validation technique was presented as well as the diagram of the classification procedure.

6 Results and Discussion

In this section the classification results for the work performed will be presented.

The classification results, obtained using ten-fold cross validation, will be evaluated with the unprocessed data, the normalized data and with the transformed data. The impact of the temporal varying prior probability will be explored in each configuration. The evaluation of the performance of each individual discriminant will be made as well.

6.1 Cohens's kappa coefficient

The metric used for the evaluation of the classification results of this work, as well as the results of the studies from literature, is the Cohens's kappa coefficient. The seminal paper introducing kappa as a new technique was published by Jacob Cohen in the journal *Educational and Psychological Measurement* in 1960 [77].

The kappa coefficient (K) measures pairwise agreement among a set of coders making category judgments, correcting for expected chance agreement, where P(A) is the proportion of times that the coders agree and P(E) is the proportion of times that we would expect them to agree by chance, calculated along the lines of the intuitive argument presented below.

$$k = \frac{\mathrm{P(A)} - \mathrm{P(E)}}{1 - P(E)} \tag{17}$$

The values of K are constrained to the interval [-1, 1]. A K value of one means perfect agreement, a K value of zero means that agreement is equal to chance, and a K value of negative one means "perfect" disagreement.

It is considered to be a more robust measure than simple percent agreement calculation since it takes into account the agreement occurring by chance. Furthermore it compensates for class imbalance, penalizing the score whenever the classification performance for the underrepresented class is poor [43].

6.2 Results for with the original data

As seen in section 4.2, the original data has various problems regarding between subject variation and absence of features with high discriminative power. Therefore the classification results show low coefficient of agreement between the predicted classes as the actual ground truth. Another important issue is the standard deviation in the classification results with the original data. The high standard deviation shows that in fact, the normalization of the data, would lead to an improvement of the pooled data quality which would cause the standard deviation to drop and possibly increase the classification results.

The introduction of temporal prior probability information causes a substantial increase in the classification performance of the algorithm for both individual classes and overall.

This preliminary results are already very good as they are already comparable to values presented in literature, although this is a problem with a bigger number of classes.

Table 1. Results with the original data

Class	*No temporal prior probability*		*With temporal prior probability*	
	Kappa	Accuracy	Kappa	Accuracy
Overall	0.39 ∓ 0.13	0.58 ∓ 0.12	0.45 ∓ 0.20	0.67 ∓ 0.10
Deep Sleep	0.39 ∓ 0.17	0.80 ∓ 0.09	0. 41 ∓ 0.17	0.86 ∓ 0.06
Light Sleep	0.30 ∓ 0.16	0.64 ∓ 0.10	0.37 ∓ 0.15	0.69 ∓ 0.08
REM	0.52 ∓ 0.20	0.84 ∓ 0.08	0.55 ∓ 0.21	0.88 ∓ 0.07
Wake	0.43 ∓ 0.19	0.88 ∓ 0.08	0.52 ∓ 0.20	0.91 ∓ 0.07

6.3 Results for with the normalized data

Following the decrease in the between subject variability, and the consequent improvement of the data in section 4.3, perceptible with the increase of the number of features with a high ASMD, the classification results were expected to increase. When comparing Table 1 and Table 2 a very significant increase in the classification results is noticeable. The overall accuracy increased about 5% while the average kappa coefficient of agreement increase almost 0.1 which is very substantial. Another positive observation is the overall decrease of the standard deviation, both for accuracy and kappa, which can be explained by the minimization of the between subject variability.

The introduction of temporal prior probability information causes a substantial increase in the classification performance of the algorithm for both individual classes and overall. On the other hand, since the prior probability information added general information of the night, which worked on most subjects but degraded the results of other subjects, the standard deviation of the results increased as well.

Table 2. Results with the normalized data

Class	*No temporal prior probability*		*With temporal prior probability*	
	Kappa	Accuracy	Kappa	Accuracy
Overall	0.48 ∓ 0.12	0.65 ∓ 0.08	0.54 ∓ 0.12	0.72 ∓ 0.07
Deep Sleep	0.44 ∓ 0.15	0.84 ∓ 0.05	0. 49 ∓ 0.17	0.88 ∓ 0.04
Light Sleep	0.39 ∓ 0.13	0.69 ∓ 0.07	0.47 ∓ 0.13	0.74 ∓ 0.13
REM	0.56 ∓ 0.19	0.86 ∓ 0.06	0.60 ∓ 0.23	0.9 ∓ 0.06
Wake	0.58 ∓ 0.15	0.90 ∓ 0.04	0.63 ∓ 0.15	0.92 ∓ 0.04

6.4 Results for with the transformed data

The transformation step, as described in 4.4, aimed at creating new features, some of them with very high discrimination power, so that new information was added to the classification procedure. However since the decision of which features and transformations was useful for the classification of each class, was delayed for the feature selection procedure, a lot of redundant, indiscriminative and therefore unnecessary features were added to the pool. The results of this section show not only the improvements of the results with the newly added features, but the effectiveness of the feature selection algorithm, which was able to identify for each class, in a very short period time, the relevant features subset, in a set with 580 features (116 + 116 x 4).

The increase in the classification performance is again very substantial. Looking at classification without the introduction of any temporal prior probability the increase of the kappa coefficient of agreement was 0.1, and for the accuracy 8%. Comparing with the original data the differences are even bigger with an increase of approximately 0.2 in the kappa coefficient of agreement and 15% for accuracy.

The introduction of temporal prior probability information causes a slight increase in the classification performance of the algorithm. As in the results above, the standard deviation of the results increased as well. The decrease in the effectiveness of the introduction of prior probability information might have several explanations. The simplest one is the fact that with the gradual increase in the classification results, the increase in the classification results with the introduction of the general temporal information will decrease gradually, as the general temporal information will gradually introduce less new information. Another explanation might be concerned with the newly added features that might be representing the effects of the differences in time to the classifier.

Table 3. Results with the transformed data

Class	*No temporal prior probability*		*With temporal prior probability*	
	Kappa	Accuracy	Kappa	Accuracy
Overall	0.58 ∓ 0.11	0.73 ∓ 0.07	0.60 ∓ 0.12	0.76 ∓ 0.07
Deep Sleep	0.60 ∓ 0.16	0.88 ∓ 0.05	0. 62 ∓ 0.16	0.90 ∓ 0.04
Light Sleep	0.49 ∓ 0.13	0.75 ∓ 0.06	0.53 ∓ 0.13	0.77 ∓ 0.06
REM	0.66 ∓ 0.19	0.86 ∓ 0.06	0.67 ∓ 0.20	0.91 ∓ 0.05
Wake	0.60 ∓ 0.16	0.93 ∓ 0.03	0.60 ∓ 0.17	0.94 ∓ 0.03

6.5 Comparisons with literature results

In order to evaluate the results of this study, this section compares the results obtained in this work with the other similar studies from literature.

As this work is pioneer in the classification of four sleep stages based on cardiorespiratory data, it is harder to establish comparisons with the other results from literature. Nevertheless, since similar studies classified three sleep stages, Non REM, REM and Wake, with Non REM corresponding to the Deep sleep and Light sleep stages of this study, it is easy to use the same algorithm, converting only the labels to the three classes problem.

Another aspect to keep in mind is the number of subjects in this study is considerably higher than the other studies. This study was performed with a total of sixty one subjects, while other studies only performed classification on eighteen[39] and thirty one[37] subjects. These number of subjects might not be sufficient for this classification problem.

According to the computed learning curves of the algorithm, in order to have a stabilized classification results that are able to generalize well, at least approximately forty subjects are needed.

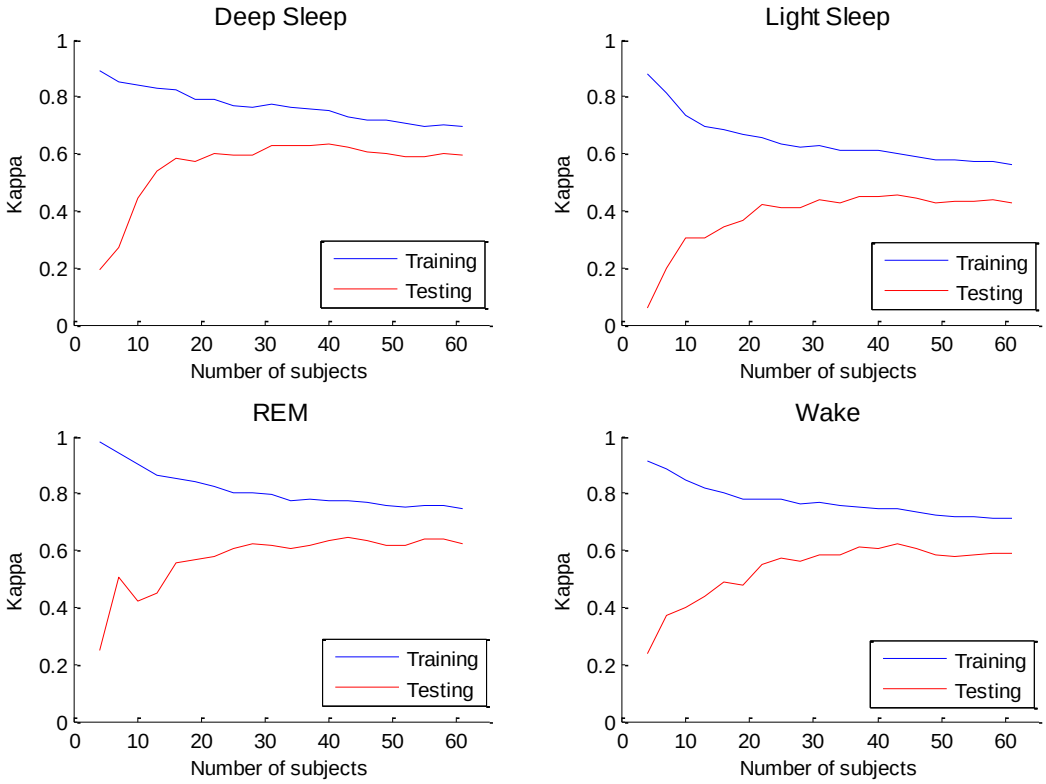


Figure 40. Learning curves for the different classes, for the classification algorithm

This figure shows that the number of subjects used in this study is appropriate.

Before establishing comparisons, one important factor should be taken into consideration. Truthfully, in [39] only a total of nine females, each with two nights, participated in the study. The cross validation methodology used in the study is unclear, but if when classifying a subject, data from the same subject is used for training, even if the data belongs to a different night recording, the classification results might be biased. Another issue to keep in mind are the total number of features. In [39] only three features were used while [37] used about thirty features, which means that both studies used significantly smaller feature sets than the one used in this work.

Table 4. Comparison with results from literature

	Classes	*Subjects*	*Kappa*	*Accuracy*
Kortelainen et al. [39]	3	18	0.44 ∓ 0.19	0.79 ∓ 0.10
Redmond et. al.[37]	3	31	0.45	0.76
Results Achieved	4	61	0.60 ∓ 0.12	0.76 ∓ 0.06
	3	61	0.66 ∓ 0.13	0.85 ∓ 0.06

When looking at the classification results, the differences are obvious. Considering the problem of classification regarding three classes, the results achieved in this study are increased by 0.2 for the coefficient of agreement, which reveals a much better agreement between classification results and ground truth, than the results from literature. The accuracy is significantly higher as well, and the standard deviation for both kappa and accuracy is much lower. For the four classes scenario, a more challenging task since there are more classes to distinguish from and therefore bigger chances of classification errors, the results are still remarkably better than the ones form previous studies. With similar accuracy and with an increase of 0.15 in the kappa coefficient of agreement.

When the differences between the subjects are taken in consideration, this values are even more noteworthy. The subjects used in this study have a mean age of 41 ± 17 years and are divided into 19 males and 41 females. In [37] a database of 31 male subjects with an mean age of 42 ± 7 years was used. In [39], nine woman with an age between 20 to 54 years. It is clear that not only the present study has subjects from different genders but also the standard deviation of the ages of the subjects are considerably larger. This means that subjects will have very different ages which, as seen in chapter 2.2.2, is a factor of variation in the subjects nigh architecture. With further investigation, the consequences of age in classification results, were found. They are resumed in the table below.

Table 5. The impact of age in the classification results

Age group	*20-29*	*30-39*	*40-49*	*50-59*	*60-79*	*>79*
Subjects	22	12	7	10	8	2
Average kappa	0,6100	0,605	0,6896	0,6220	0,5554	0,2650

This table shows that age is not only a cause of variation in subject's sleep architecture, but it also affects classification results. In fact there is a clear decrease in classification

performance with elderly subjects, 60 years or more. This fact is very interesting since when examining Figure 3, the most noticeable changes in the sleep architecture occur when the age of the subject is around 55 to 60 years old.

Nevertheless these differences are small and were probably minimized in the normalization procedure. For subjects with and age over 79 years old, the classification results were very different from the rest of the subjects. These results might suggest that this method may not be appropriate for subjects within this age group. On the other hand, there were only two subjects in that age group, so any conclusions regarding these results are unreliable because of the lack of sufficient subjects.

7 Conclusions

Sleep medicine is a relatively new field of study, which aims at study, diagnose and treat sleep disturbances. As seen in section 2, current methods for sleep assessment are expensive, very obtrusive and not widely available for the general population. Thus, reducing the personal burden and the societal cost of sleep disturbances has become one of the major challenges in the last decades. The system which has been presented in this thesis takes part in this challenge by aiming at automatic detection of multiple sleep stages, light sleep, deep sleep, REM and wake using cardiorespiratory data. The sensors used for the acquisition of the cardiorespiratory data, in future work can likely be substituted by non-obtrusive sensors like bed sensors. Section 3 describes the data collected from healthy subjects from various sleep laboratories using the previously mentioned sensors and full polysomnographic data. Sleep staging, was carried out by sleep professionals, using a 30 seconds epochs and the acquired polysomnographic data. From the original data 116 features were extracted regarding cardiac information, respiratory data or both. Since the scope of this project is to primarily assess the method for the classification of sleep stages on healthy subjects, a filter, with relaxed parameters defining healthy subjects characteristics obtained from literature, was applied in order to remove subjects with an unusual sleep architecture. This study, when compared against similar studies in literature, has a bigger number of subjects, from both genders, with a wide age distribution and measured in different sleep laboratories. These are all sources of variation and therefore normalization of the data had to be applied. In section 4, the normalization techniques are described. Also, in order to extract new information from the existent features, like statistical information, which might not be used by the classifier, a set of transformation techniques is presented. This two separate processes were able to improve the number of features with a higher discriminatory power for all classes and therefore increase the quality of the data, however the transformation step introduced a lot of irrelevant features as well. In section 5, the classifier algorithm is described, furthermore the feature selection algorithm that will deal with the high number of features, is presented and explained. In order to capture the time variations of sleep, section 5 also describes how the temporal prior information is computed and added to the results from the classifier. Moreover the establishment of the final classification, considering the results from the classifier and from the temporal

varying prior probabilities, is presented as well as the tune of the parameters. The performance and results obtained are discussed in section 6 for the different classes and configurations.

The proposed objectives for this work were all successfully accomplished. The research on the sleeping patterns and normal healthy adult sleep architecture, allowed the creation of newer probabilistic techniques to capture the non-stationary temporal characteristics of sleep. Furthermore the reported differences in the variability of the cardiorespiratory data, associated with different sleep stages, motivated the creation of probabilistic transformation methods, which created new highly discriminative features from the already existent ones.

When these new statistical transformation methods, inspired by the existent literature, are complemented with other transformation techniques, they were able to create new features with high discrimination power which improved the overall data quality, assuming that at classification time, the feature selection algorithm is able to select only the relevant features. Furthermore, in order to reduce the between subject variability, causing high standard deviation in the classification results and degrading data quality, the set of normalization techniques applied to the features. The discrimination power of the features increased and when classified the classification results improved as well with a decrease in the standard deviation.

As each class had a distinct discriminative set of features, the use of individual linear discriminants, trained in a one against the rest setup, allowed the selection and use of the relevant features for the discrimination of each class. The slightly modified version of correlation feature selection [72] has proven very effective in the problem of selecting the relevant features for each class, taking into account the correlation between the features. It is able to select a representative set of features for each class in a short period of time. The last step of the classification process is the introduction of temporal prior probabilities in order to capture the temporal non-stationary sleep properties. The probabilistic

techniques used in this work, extended others found in literature and were able to introduce new information to the classifier scores that further increased classification results. In the end the different scores are measured and the final classification is established.

The final and best configuration, taking into consideration, the normalizations and transformations of the data, the feature selection algorithm, the various linear classifiers and the temporal prior probabilities, achieved remarkably good results when compared against the existent results from literature. Cardiorespiratory signals provide significant accuracy and correlation for the classification of multiple sleep stages. This type of systems might be a useful addition for long term sleep monitoring purposes with the aptitude of being unobtrusive for the user.

8 Bibliography

[1] C. on S. M. and Research, *Sleep Disorders and Sleep Deprivation*. National Academies Press, 2006, p. 387.

[2] M. W. Report, "National Sleep Awareness Week Unhealthy Sleep-Related Behaviors," vol. 60, no. 8, pp. 2005–2008, 2011.

[3] X. Long and P. Fonseca, "Using dynamic time warping for sleep and wake discrimination," in *EEE-EMBS International Conference on Biomedical and Health Informatics*, 2012, vol. 25, no. Bhi, pp. 886–889.

[4] S. Devot, R. Dratwa, and E. Naujokat, "Sleep/wake detection based on cardiorespiratory signals and actigraphy.," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society Conference*, vol. 2010, pp. 5089–92, Jan. 2010.

[5] R. Silverburg, *The Dawn of Medicine*. New York: Putnam Publishing, 1966.

[6] S. Inoué, K. Honda, and Y. Komoda, "Sleep as neuronal detoxification and restitution," *Behavioural Brain Research*, vol. 69, no. 1–2, pp. 91–96, Jul. 1995.

[7] H. Berger, "Über das Elektrenkephalogramm des Menschen," *Archiv für Psychiatrie und Nervenkrankheiten*, vol. 87, no. 1, pp. 527–570, Dec. 1929.

[8] A. N. Rama, S. C. Cho, and C. A. Kushida, "Normal Human Sleep," *Sleep: A Comprehensive Handbook*, pp. 3–9, 2009.

[9] A. Loomis, E. Harvey, and G. Hobart, "Cerebral states during sleep, as studied by human brain potentials.," *Journal of Experimental …*, 1937.

[10] E. ASERINSKY and N. KLEITMAN, "Regularly occurring periods of eye motility, and concomitant phenomena, during sleep.," *Science (New York, N.Y.)*, vol. 118, no. 3062, pp. 273–4, Sep. 1953.

[11] E. Aserinsky and N. Kleitman, "Regularly occurring periods of eye motility, and concomitant phenomena, during sleep," *Science*, 1953.

[12] W. Dement and N. Kleitman, "Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming," *Electroencephalography and Clinical Neurophysiology*, vol. 9, no. 4, pp. 673–690, Nov. 1957.

[13] M. Jouvet, F. Michel, and J. Courjon, "Sur un stade d'activité électrique cérébrale rapide au cours du sommeil physiologique," *C.R. Soc. Biol. (Paris)*, vol. 153, pp. 1024–1028, 1959.

[14] J. W. Shepard, D. J. Buysse, A. L. Chesson, W. C. Dement, R. Goldberg, C. Guilleminault, C. D. Harris, C. Iber, E. Mignot, M. M. Mitler, K. E. Moore, B. a Phillips, S. F. Quan, R. S. Rosenberg, T. Roth, H. S. Schmidt, M. H. Silber, J. K. Walsh, and D. P. White, "History of the development of sleep medicine in the United States.," *Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine*, vol. 1, no. 1, pp. 61–82, Jan. 2005.

[15] H. Gastaut, C. Tassinari, and B. Duron, "Etude polygraphique des manifestations episodiques (hypniques et respiratoires) du syndrome de Pickwick," *Rev Neurol*, pp. 568–579, 1965.

[16] R. Jung and W. Kuhlo, "Neurophysiological Studies of Abnormal Night Sleep and the Pickwickian Syndrome," in *Sleep Mechanisms*, vol. 18, C. B. K. Akert and J. P. Schadé, Eds. Elsevier, 1965, pp. 140–159.

[17] J. Allan Hobson, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," *Electroencephalography and Clinical Neurophysiology*, vol. 26, no. 6, p. 644, Jun. 1969.

[18] W. C. Dement, "History of Sleep Physiology and Medicine," in *Principles and Practice of Sleep Medicine*, 2010, pp. 3–15.

[19] L. Rosenthal, "Physiologic Processes During Sleep," *Sleep: A Comprehensive Handbook*, pp. 19–23, 2009.

[20] M. A. Carskadon and J. G. Oskar, "Life cycles: Infants to Adolescents," in *Basics of Sleep Guide*, 2009, pp. 33–40.

[21] M. M. Ohayon, M. a Carskadon, C. Guilleminault, and M. V Vitiello, "Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan.," *Sleep*, vol. 27, no. 7, pp. 1255–73, Nov. 2004.

[22] Association of Sleep Disorders Centers and the Association for the Psychophysiological Study of Sleep, "Diagnostic Classification of Sleep and Arousal Disorders," *Sleep*, 1979.

[23] American Academy of Sleep Medicine, "International Classification of Sleep Disorders – Second Edition (ICSD-2)." p. 298, 2005.

[24] M. Daley, C. M. Morin, M. Leblanc, J. Grégoire, and J. Savard, "The Economic Burden of Insomnia : Direct and Indirect Costs for Individuals with Insomnia Syndrome , Insomnia Symptoms , and Good Sleepers," 1995.

[25] S. Schutte-Rodin, L. Broch, D. Buysse, C. Dorsey, and M. Sateia, "Clinical guideline for the evaluation and management of chronic insomnia in adults.," *Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine*, vol. 4, no. 5, pp. 487–504, Oct. 2008.

[26] T. Morgenthaler, C. Alessi, L. Friedman, J. Owens, V. Kapur, B. Boehlecke, T. Brown, A. Chesson, J. Coleman, T. Lee-Chiong, J. Pancer, and T. J. Swick, "Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007.," *Sleep*, vol. 30, no. 4, pp. 519–29, Apr. 2007.

[27] T. Young, M. Palta, J. Dempsey, J. Skatrud, S. Weber, and S. Badr, "The occurrence of sleep-disordered breathing among middle-aged adults.," *The New England journal of medicine*, vol. 328, no. 17, pp. 1230–5, Apr. 1993.

[28] T. Young, L. Evans, L. Finn, and M. Palta, "Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women.," *Sleep*, vol. 20, no. 9, pp. 705–6, Sep. 1997.

[29] T. Gislason and B. Benediktsdóttir, "Snoring, apneic episodes, and nocturnal hypoxemia among children 6 months to 6 years old. An epidemiologic study of lower limit of prevalence.," *Chest*, vol. 107, no. 4, pp. 963–6, Apr. 1995.

[30] N. A. Collop, "Portable monitoring for the diagnosis of obstructive sleep apnea.," *Current opinion in pulmonary medicine*, vol. 14, no. 6, pp. 525–9, Nov. 2008.

[31] N. A. Collop, W. M. Anderson, B. Boehlecke, D. Claman, and G. Rochelle, "Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients," *Journal of clinical sleep medicine JCSM official publication of the American Academy of Sleep Medicine*, vol. 3, no. 7, 2007.

[32] H. W. Agnew, W. B. Webb, and R. L. Williams, "The first night effect: an EEG study of sleep.," *Psychophysiology*, vol. 2, no. 3, pp. 263–6, Jan. 1966.

[33] A. Zaffaroni, P. de Chazal, C. Heneghan, P. Boyle, P. R. Mppm, and W. T. McNicholas, "SleepMinder: an innovative contact-free device for the estimation of the apnoea-hypopnoea index.," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2009, pp. 7091–4, Jan. 2009.

[34] N. D. Duffy, S. McGowan, and G. D. Drummond, "Optical noncontact measurement of abdominal volume change during anaesthesia," in *IEE Colloq. Opt. Tech. Biomed. Appl*, 1991, pp. 8/1–8/4.

[35] Y.-M. Kuo, J.-S. Lee, and P.-C. J. Chung, "A visual context-awareness-based sleeping-respiration measurement system.," *IEEE transactions on information technology in biomedicine*, vol. 14, no. 2, pp. 255–65, Mar. 2010.

[36] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data.," *Biometrics*, vol. 33, no. 1, pp. 159–74, Mar. 1977.

[37] S. J. Redmond, P. Chazal, C. O'Brien, S. Ryan, W. T. McNicholas, and C. Heneghan, "Sleep staging using cardiorespiratory signals," *Somnologie - Schlafforschung und Schlafmedizin*, vol. 11, no. 4, pp. 245–256, Oct. 2007.

[38] S. Redmond and C. Heneghan, "Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea," *IEEE transactions on bio-medical engineering*, vol. 53, no. 3, pp. 485–96, 2006.

[39] J. Kortelainen, M. Mendez, A. M. Bianchi, M. Matteucci, and S. Cerutti, "Sleep staging based on signals acquired through bed sensor," *IEEE transactions on information technology in biomedicine*, vol. 14, no. 3, pp. 776–785, 2010.

[40] M. O. Mendez, M. Matteucci, S. Cerutti, F. Aletti, and a M. Bianchi, "Sleep staging classification based on HRV: time-variant analysis.," *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, vol. 2009, pp. 9–12, Jan. 2009.

[41] R. Duda, P. Hart, and D. Stork, "Pattern Classification," *New York: John Wiley, Section*, 2001.

[42] J. P. M. de Sá, *Pattern recognition: concepts, methods and applications : with 197 figures (Google eBook)*. Springer, 2001, p. 318.

[43] Http://www.ofai.at/siesta/, "SIESTA," *(14-06-2013)*. .

[44] M. Carskadon and W. Dement, "Normal human sleep: an overview," in *Principles and Practice of Sleep Medicine*, 2010, pp. 16–26.

[45] O. Le Bon, L. Staner, G. Hoffmann, M. Dramaix, I. San Sebastian, J. R. Murphy, M. Kentos, I. Pelc, and P. Linkowski, "The first-night effect may last more than one night.," *Journal of psychiatric research*, vol. 35, no. 3, pp. 165–72, 2001.

[46] S. F. Q. Conrad Iber, Sonia Ancoli-Israel, Andrew L. Chesson Jr., "The AASM Manual for the Scoring of Sleep and Associted Events." 2007.

[47] R. A. and K. A., *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects.* Washington: Public Health Service, 1968.

[48] J. Tukey, "The future of data analysis," *The Annals of Mathematical Statistics*, 1962.

[49] W. Dixon, "Simplified estimation from censored normal samples," *The Annals of Mathematical Statistics*, 1960.

[50] R. J. Larsen and M. L. Marx, *Introduction to Mathematical Statistics and Its Applications (5th Edition)*. Pearson, 2011, p. 768.

[51] B. M. Bolstad, R. a Irizarry, M. Astrand, and T. P. Speed, "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.," *Bioinformatics (Oxford, England)*, vol. 19, no. 2, pp. 185–93, Jan. 2003.

[52] M. H. Bonnet and D. L. Arand, "Heart rate variability: sleep stage, time of night, and arousal influences.," *Electroencephalography and clinical neurophysiology*, vol. 102, no. 5, pp. 390–6, May 1997.

[53] S. Rostig, J. W. Kantelhardt, T. Penzel, W. Cassel, J. H. Peter, C. Vogelmeier, H. F. Becker, and A. Jerrentrup, "Nonrandom variability of respiration during sleep in healthy humans.," *Sleep*, vol. 28, no. 4, pp. 411–7, Apr. 2005.

[54] M. H. Bonnet and D. L. Arand, "Heart rate variability in insomniacs and matched normal sleepers.," *Psychosomatic medicine*, vol. 60, no. 5, pp. 610–5, 1998.

[55] T. Penzel, J. W. Kantelhardt, C.-C. Lo, K. Voigt, and C. Vogelmeier, "Dynamics of heart rate and sleep stages in normals and patients with sleep apnea.," *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, vol. 28 Suppl 1, pp. S48–53, Jul. 2003.

[56] T. Penzel and A. Bunde, "Sleep stage-dependent heart rate variability in patients with obstructive sleep apnea," *Computers in Cardiology*, pp. 249–252, 1999.

[57] M. H. Bonnet and D. L. Arand, "Heart rate variability: sleep stage, time of night, and arousal influences.," *Electroencephalography and clinical neurophysiology*, vol. 102, no. 5, pp. 390–6, May 1997.

[58] M. O. Mendez and P. Milano, "Sleep staging from Heart Rate Variability : time-varying spectral features and Hidden Markov Models Matteo Matteucci Vincenza Castronovo and Luigi Ferini-Strambi Sergio Cerutti and Anna Maria Bianchi," vol. 3, pp. 246–263, 2010.

[59] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967.

[60] J. R. Quinlan, *C4.5: Programs for Machine Learning*. 1993.

[61] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees*. Chapman and Hall/CRC, 1984, p. 368.

[62] A. Smola and B. Schölkopf, *Learning with kernels*, no. February. 1998.

[63] L. Bottou and C. Cortes, "Comparison of classifier methods: a case study in handwritten digit recognition," *International Conference on Pattern Recognition - ICPR*, 1994.

[64] U. H. G. Kressel, "Pairwise classification and support vector machines," in *Advances in kernel methods*, MIT Press, 1999, pp. 255–268.

[65] T. Hastie and R. Tibshirani, *The elements of statistical learning: data mining, inference and prediction*, vol. 27, no. 2. 2005, pp. 83–85.

[66] J. Friedman, "Another approach to polychotomous classification," 1996.

[67] J. C. Platt, N. Cristianini, and J. S. Taylor, "Large margin DAGs for multiclass classification," in *roceedings of Neural Information Processing Systems, NIPS'99*, MIT Press, 2000, pp. 547–553.

[68] G. B. Thomas G. Dietterich, "Solving multiclass learning problems via error-correcting output codes," *Journal of Artificial Intelligence Research*, vol. 2, no. 1, pp. 263–286, 1995.

[69] A. Klautau, E. Allwein, R. Schapire, and Y. Singer, "Reducing multiclass to binary: a unifying approach for margin classifiers," *The Journal of Machine Learning Research*, vol. 1, pp. 113–141, 2001.

[70] J. Weston and C. Watkins, "Support Vector Machines for Multi-Class Pattern Recognition," 1999.

[71] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, vol. 13, no. 4, pp. 1026–7, Jan. 2002.

[72] M. Hall, "Correlation-based feature selection for machine learning," 1999.

[73] P. Langley and S. Sage, "Induction of selective Bayesian classifiers," *Proceedings of the Tenth international conference on Uncertainty in artificial intelligence*, pp. 390–406, 1994.

[74] R. Kohavi and G. John, "Wrappers for feature subset selection," *Artificial intelligence*, vol. 97, no. 97, pp. 273–324, 1997.

[75] R. Kohavi and D. Sommerfield, "Feature subset selection using the wrapper method: Overfitting and dynamic search space topology," 1995.

[76] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[77] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and psychological measurement*, pp. 37–46, 1960.

[78] A. J. Camm, M. Malik, J. T. Bigger, G. Breithardt, S. Cerutti, R. J. Cohen, P. Coumel, E. L. Fallen, H. L. Kennedy, R. E. Kleiger, F. Lombardi, A. Malliani, A. J. Moss, J. N. Rottman, G. Schmidt, P. J. Schwartz, and D. H. Singer, "Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[79] M. C. Teich, S. B. Lowen, B. M. Jost, and K. Vibe-rheymer, "Heart Rate Variability: Measures and Models," 2000.

[80] M. Costa, A. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of biological signals," *Physical Review E*, vol. 71, no. 2, p. 021906, Feb. 2005.

[81] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series.," *Physical review letters*, vol. 89, no. 6, p. 068102, Aug. 2002.

[82] T. Penzel, J. W. Kantelhardt, L. Grote, J.-H. Peter, and A. Bunde, "Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in

sleep and sleep apnea.," *IEEE transactions on bio-medical engineering*, vol. 50, no. 10, pp. 1143–51, Oct. 2003.

[83] J. Kantelhardt and E. Koscielny-Bunde, "Detecting long-range correlations with detrended fluctuation analysis," *Physica A: Statistical Mechanics and its Applications*, vol. 295, no. 3–4, pp. 441–454, 2001.

[84] M. Staudacher, S. Telser, a. Amann, H. Hinterhuber, and M. Ritsch-Marte, "A new method for change-point detection developed for on-line analysis of the heart beat variability during sleep," *Physica A: Statistical Mechanics and its Applications*, vol. 349, no. 3–4, pp. 582–596, Apr. 2005.

[85] S. Telser, M. Staudacher, and Y. Ploner, "Can One Detect Sleep Stage Transitions for On-Line Sleep Scoring by Monitoring the Heart Rate Variability?," *Somnologie - Schlafforschung und Schlafmedizin*, vol. 8, no. 2, pp. 33–41, 2004.

[86] M. Mendez, a M. Bianchi, O. Villantieri, and S. Cerutti, "Time-varying analysis of the heart rate variability during REM and non REM sleep stages.," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, pp. 3576–9, Jan. 2006.

[87] M. Adnane and Z. Jiang, "Automatic sleep-wake stages classifier based on ECG," *ICCAS-SICE, 2009*, vol. 2, pp. 493–498, 2009.

[88] Y. Ichimaru, K. P. Clark, J. Ringler, and W. J. Weiss, "Effect of sleep stage on the relationship between respiration and heart rate variability," *[1990] Proceedings Computers in Cardiology*, pp. 657–660, 1991.

[89] A. Noviyanto, S. Isa, I. Wasito, and A. Arymurthy, "Selecting features of single lead ecg signal for automatic sleep stages classification using correlation-based feature subset selection," *International Journal of Computer Science Issues*, vol. 8, no. 5, pp. 1178–1181, 2011.

[90] M. Wiggins, A. Saad, B. Litt, and G. Vachtsevanos, "Evolving a Bayesian classifier for ECG-based age classification in medical applications," *Applied soft computing*, vol. 8, no. 1, pp. 599–608, 2008.

[91] B. O. Hjorth, "EEG analysis based on time domain properties," *Electroencephalography and Clinical Neurophysiology*, vol. 29, no. 3, pp. 306–310, 1970.

[92] F. S. Bao, X. Liu, and C. Zhang, "PyEEG: an open source Python module for EEG/MEG feature extraction.," *Computational intelligence and neuroscience*, vol. 2011, p. 406391, Jan. 2011.

[93] T. Willemen, D. Van Deun, V. Verhaert, S. Pirrera, V. Exadaktylos, J. Verbraecken, B. Haex, and J. Vander Sloten, "Automatic sleep stage classification based on easy to register signals as a validation tool for ergonomic steering in smart bedding systems.," *Work (Reading, Mass.)*, vol. 41 Suppl 1, pp. 1985–9, Jan. 2012.

[94] B. Yilmaz, M. H. Asyali, E. Arikan, S. Yetkin, and F. Ozgen, "Sleep stage and obstructive apneaic epoch classification using single-lead ECG.," *Biomedical engineering online*, vol. 9, p. 39, Jan. 2010.

[95] H. Bettermann, D. Cysarz, and P. Van Leeuwen, "Detecting Cardiorespiratory Coordination By Respiratory Pattern Analysis of Heart Period Dynamics — the Musical Rhythm Approach," *International Journal of Bifurcation and Chaos*, vol. 10, no. 10, pp. 2349–2360, Oct. 2000.

[96] D. Cysarz, H. Bettermann, S. Lange, D. Geue, and P. van Leeuwen, "A quantitative comparison of different methods to detect cardiorespiratory coordination during night-time sleep.," *Biomedical engineering online*, vol. 3, no. 1, p. 44, Nov. 2004.

[97] J. Richman and J. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *merican journal of physiology. Heart and circulatory physiology*, vol. 278, no. 6, pp. H2039–49, 2000.

[98] X. Long, J. Foussier, P. Fonseca, R. Haakma, and R. M. Aarts, "Respiration Amplitude Analysis for REM and NREM Sleep Classification," *35th Annual International Conference of the IEEE EMBS*, 2013.

9 Appendix

Since available feature set is quite extensive, the information was resumed in a table with the corresponding feature number, name of the feature and a small description. The features number is not sequential for several reasons. Some features are not available to all datasets in this study. Other features are being implemented, have errors or were removed for other reasons.

Table 6. Features used

N	Name	Description
13	resp_freq_periodogram	The respiratory frequency [38]
14	resp_power_freq_periodogram	The power of the respiratory frequency (in frequency domain) [38]
15	resp_vlf_periodogram	Normalized total power in the Very Low Frequency band [38]
16	resp_lf_periodogram	Normalized total power in the Low Frequency band [38]
17	resp_hf_periodogram	Normalized total power in the High Frequency band [38]
18	resp_lf_hf_periodogram	Ratio between the low and high frequency band power [38]
19	resp_v_5_epochs	Logarithm of the standard deviation of the respiratory frequency over a 5 epochs sliding window [37]
20	resp_v_7_epochs	Logarithm of the standard deviation of the respiratory frequency over a 7 epochs sliding window [37]
21	resp_v_9_epochs	Logarithm of the standard deviation of the respiratory frequency over a 9 epochs sliding window [37]
22	resp_mean_breath_by_breath_corr	Mean breath-by-breath correlation [38]
23	resp_std_breath_by_breath_corr	Standard deviation of breath-by-breath correlation [38]
24	resp_std_breath_length	Standard deviation of breath length[38]
25	resp_freq_td	Respiratory frequency estimation from the time-domain [38]
26	ecg_hr_mean	Mean heart rate, reciprocal of RR mean [38],[78]
27	ecg_rr_mean	Mean RR-interval [38]

28	ecg_sdnn	Logarithm of the standard deviation of the normalized RR interval [79]
29	ecg_rr_range	Range of the normalized RR interval [79]
30	ecg_pnn50	Percentage of NN intervals with a >50ms difference [79]
31	ecg_rmssd	Square root of the mean squared differences between adjacent normal-to-normal heart beats (NN) intervals [79]
32	ecg_sdsd	Standard deviation of differences between adjacent NN intervals [79]
33	ecg_vlf_norm	Logarithm of ratio of Very Low Frequency power [78],[79]
34	ecg_lf_norm	Logarithm of ratio of Low Frequency power [78], [79]
35	ecg_hf_norm	Logarithm of ratio of High Frequency power [78], [79]
36	ecg_lf_hf_ratio	Ratio between low and high power [78], [79]
37	ecg_sampen1_scale1	Computes Sample Entropy, over the original detrended RR interval using scale=1 and template length = 1 [80], [81]
38	ecg_sampen2_scale1	Computes Sample Entropy, over the original detrended RR interval using scale=1 and template length = 2 [80], [81]
39	ecg_sampen1_scale2	Sample Entropy with scale=2 and template length = 1
40	ecg_sampen2_scale2	Sample Entropy with scale=2 and template length = 2
41	ecg_sampen1_scale3	Sample Entropy with scale=3 and template length = 1
42	ecg_sampen2_scale3	Sample Entropy with scale=3 and template length = 2
43	ecg_sampen1_scale4	Sample Entropy with scale=4 and template length = 1
44	ecg_sampen2_scale4	Sample Entropy with scale=4 and template length = 2
45	ecg_sampen1_scale5	Sample Entropy with scale=5 and template length = 1
46	ecg_sampen2_scale5	Sample Entropy with scale=5 and template length = 2
47	ecg_sampen1_scale6	Sample Entropy with scale=6 and template length = 1
48	ecg_sampen2_scale6	Sample Entropy with scale=6 and template length = 2
49	ecg_sampen1_scale7	Sample Entropy with scale=7 and template length = 1
50	ecg_sampen2_scale7	Sample Entropy with scale=7 and template length = 2

51	ecg_sampen1_scale8	Sample Entropy with scale=8 and template length = 1
52	ecg_sampen2_scale8	Sample Entropy with scale=8 and template length = 2
53	ecg_sampen1_scale9	Sample Entropy with scale=9 and template length = 1
54	ecg_sampen2_scale9	Sample Entropy with scale=9 and template length = 2
55	ecg_sampen1_scale10	Sample Entropy with scale=10 and template length = 1
56	ecg_sampen2_scale10	Sample Entropy with scale=10 and template length = 2
57	ecg_alpha_1	Computes short-term DFA correlation coefficient [82]
58	ecg_alpha_2	Computes long-term DFA correlation coefficient [82]
59	ecg_alpha	Computes overall DFA correlation coefficient [83]
60	ecg_alpha_al	Computes non-normalized DFA correlation coefficient
61	ecg_pdfa	Computes progressive DFA [84], [85]
62	ecg_mean_resp_freq	Estimates the respiratory frequency from ECG
63	ecg_power_mean_resp_freq	Power of the mean respiratory frequency (in frequency domain)
64	ecg_phase_hf_pole	Estimate the phase of the HF pole [86]
65	ecg_module_hf_pole	Estimate the module of the HF pole [86]
66	ecg_rr_mean_detr	Mean RR-interval of detrended series [38]
67	ecg_wdfa	Computes windowed DFA [4], [87]
88	resp_sampen	Computes the SampEn over the original respiratory signal
89	x_resp_ecg_copower	Spectral coherence between respiratory and ecg (above VLF band) [88]
90	resp_activity	Variance over the respiratory signal of which only the peaks are preserved (through a median filter)
91	resp_dtw_dist	Constrained Dynamic Time Warp distance [3]
94	ecg_power	Power of the ecg signal (x^2) [89]
95	ecg_4th_power	4th power of the ecg signal (x^4) [89], [90]
96	ecg_curve_length	Length of the signal curve [89]
97	ecg_nonlin_energy	Nonlinear energy [89]
100	ecg_hjorth_mobility	Hjorth Mobility [89], [91]
101	ecg_hjorth_complexity	Hjorth Complexity [91]

103	ecg_psd_peak_power	Peak power of the ecg spectral density [89]
104	ecg_psd_peak_frequ	Frequency corresponding to the peak power of the ecg [89]
105	ecg_psd_mean	Mean power of the ecg spectral density [89]
106	ecg_psd_median	Median spectral power of the ecg [89]
107	ecg_psd_entropy	Spectral entropy [89]
109	ecg_hurst_exponent	Computes the Hurst exponent over the ecg signal (also Rescaled Range Statistics) [92]
116	x_cwt_activity	Estimation of the number of movements (activity), over an epoch
119	ecg_rr_percentile10	10^{th} percentile of the RR interval distribution [93]
120	ecg_rr_percentile25	25^{th} percentile of the RR interval distribution [93]
121	ecg_rr_median	Median RR interval [94]
122	ecg_rr_percentile75	75^{th} percentile of the RR interval distribution [93]
123	ecg_rr_percentile90	90^{th} percentile of the RR interval distribution [93]
124	ecg_rr_MAD	Mean absolute deviation of the RR interval [94]
125	ecg_rr_percentile10_detr	10^{th} percentile of the detrended RR interval distribution
126	ecg_rr_percentile25_detr	25^{th} percentile of the detrended RR interval distribution
127	ecg_rr_median_detr	Median of the detrended RR interval
128	ecg_rr_percentile75_detr	75^{th} percentile of the detrended RR interval distribution
129	ecg_rr_percentile90_detr	90^{th} percentile of the detrended RR interval distribution
130	ecg_rr_MAD_detr	Mean absolute deviation of the detrended RR interval
131	ecg_hr_percentile10	10^{th} percentile of the heart rate
132	ecg_hr_percentile25	25^{th} percentile of the heart rate
133	ecg_hr_median	Median heart rate
134	ecg_hr_percentile75	75^{th} percentile of the heart rate
135	ecg_hr_percentile90	90^{th} percentile of the heart rate
136	ecg_hr_MAD	Mean absolute deviation of the heart rate
137	ecg_hr_percentile10_detr	10^{th} percentile of the detrended heart rate

138	ecg_hr_percentile25_detr	25th percentile of the detrended heart rate
139	ecg_hr_median_detr	Median detrended heart rate
140	ecg_hr_percentile75_detr	75th percentile of the detrended heart rate
141	ecg_hr_percentile90_detr	90th percentile of the detrended heart rate
142	ecg_hr_MAD_detr	Mean absolute deviation of the detrended heart rate
159	x_resp_ecg_phase_coordination_long	Coordination between resp and ecg long term [95], [96]
160	x_resp_ecg_phase_coordination_short	Coordination between resp and ecg short-term [95], [96]
161	ecg_phase_coordination_long	Coordination in ecg long term
162	ecg_phase_coordination_short	Coordination in ecg short-term
163	resp_dfw_dist	Constrained dynamic time warp distance measure for respiration, over the power spectral density [3]
164	resp_amp_peak_ApEn	Approximate entropy for peaks for respiration amplitude [97], [98]
165	resp_amp_trough_ApEn	Approximate entropy for troughs for respiration amplitude [71]
166	resp_amp_peak_sd_mean	Standardized mean peak value for respiration amplitude [71]
167	resp_amp_trough_sd_mean	Standardized mean trough value for respiration amplitude [71]
168	resp_amp_peak_sd_median	Standardized median peak value for respiration amplitude [71]
169	resp_amp_trough_sd_median	Standardized median trough value for respiration amplitude [71]
170	resp_amp_pt_dist_median	Median peak/trough value ratio. Ratios of zero are ignored [71]
171	resp_amp_pt_dtw_dist	Minimum constrained dynamic time warp distance between Zscore normalized peak and trough series [71]
172	resp_breath_vol_median	Median breath volume. Zeros are ignored [71]
173	resp_breath_in_vol_median	Median inhale volume (ignoring zeros) [71]
174	resp_breath_ex_vol_median	Median exhale volume (ignoring zeros) [71]
175	resp_breath_fr_median	Median of the breath volume over time (ignoring zeros) [71]

176	resp_breath_in_fr_median	Median of the inhale volume over time (ignoring zeros) [71]
177	resp_breath_ex_fr_median	Median of the exhale volume over time (ignoring zeros) [71]
178	resp_breath_in_ex_fr_ratio	Median of the ratio of inhale and exhale volume over time (ignoring zeros) [71]
179	resp_breath_in_ex_time_ratio	Median inhale/exhale time ratio [71]