Aalto University
School of Science
Degree Programme in Life Science Technologies

Jukka Ranta

Sleep classification in infants

using a bed mattress sensor and analysis of ECG

Master's Thesis
Espoo, May 5, 2018

Supervisor: Professor Jari Saramäki, Aalto University
Advisor: Nathan Stevenson Ph.D., University of Helsinki
Professor Sampsa Vanhatalo, University of Helsinki

Aalto University
School of Science
Degree Programme in Life Science Technologies

ABSTRACT OF
MASTER'S THESIS

Author: Jukka Ranta

Title:
Sleep classification in infants using a bed mattress sensor and analysis of ECG

Date: May 5, 2018 **Pages:** vii + 65

Major: Complex Systems **Code:** T-110

Supervisor: Professor Jari Saramäki

Advisor: Nathan Stevenson Ph.D.
Professor Sampsa Vanhatalo

Infants spend the majority of their time asleep. Although extensive studies have been carried out, the role of sleep for infant cognitive, psychomotor, temperament and developmental outcomes is not clear. The current contradictory results may be due to the limited precision when monitoring infant sleep for prolonged periods of time, from weeks to even months. Sleep-wake cycle can be assessed with sleep questionnaires and actigraphy, but they cannot separate sleep stages. The gold standard for sleep state annotation is polysomnography (PSG), which consist of several signal modalities such as electroencephalogram, electrooculogram, electrocardiogram (ECG), electromyogram, respiration sensor and pulse oximetry. A sleep clinician manually assigns sleep stages for 30 sec epochs based on the visual observation of these signals. Because method is obtrusive and laborious it is not suitable for monitoring long periods. There is, therefore, a need for an automatic and unobtrusive sleep staging approach.

In this work, a set of classifiers for infant sleep staging was created and evaluated. The cardiorespiratory and gross body movement signals were used as an input. The different classifiers aim to distinguish between two or more different sleep states. The classifiers were built on a clinical sleep polysomnography data set of 48 infants with ages ranging from 1 week to 18 weeks old (a median of 5 weeks). Respiration and gross body movements were observed using an electromechanical film bed mattress sensor manufactured by Emfit Ltd. ECG of the PSG setup was used for extracting cardiac activity. Signals were preprocessed to remove artefacts and an extensive set of features (N=81) were extracted on which the classifiers were trained.

The NREM3 vs other states classifier provided the most accurate results. The median accuracy was 0.822 (IQR: $0.724 - 0.914$). This is comparable to previously published studies on other sleep classifiers, as well as to the level of clinical interrater agreement. Classification methods were confounded by the lack of muscle atonia and amount of gross body movements in REM sleep.

The proposed method could be readily applied for home monitoring, as well as for monitoring in neonatal intensive care units.

Keywords: Infant, sleep stage classification, bed mattress sensor, electrocardiogram, polysomnography, electromechanical film, support vector machine

Language: English

Aalto-yliopisto
Perustieteiden korkeakoulu
Tietotekniikan koulutusohjelma

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä:	Jukka Ranta		
Työn nimi:	Vauvojen unen luokittelu patja-sensorilla ja EKG:lla		
Päiväys:	5. toukokuuta 2018	**Sivumäärä:**	vii + 65
Pääaine:	Kompleksiset Järjestelmät	**Koodi:**	T-110
Valvoja:	Professori Jari Saramäki		
Ohjaaja:	Tohtori Nathan Stevenson Professori Sampsa Vanhatalo		

Vauvat nukkuvat suurimman osan vuorokaudesta. Vaikkakin laajasti on tutkittu unen vaikutusta lapsen kognitioon, psykomotoriikkaan, temperamenttiin ja kehitykseen, selkeää kuvaa ja yhtenäistä konsensusta tiedeyhteisössä ei ole saavutettu. Yksi syy tähän on että ei ole olemassa menetelmää, joka soveltuisi jatkuva-aikaiseen ja pitkäkestoiseen unitilan monitorointiin. Vauvojen uni-valve- sykliä voidaan selvittää vanhemmille suunnatuilla kyselyillä ja aktigrafialla, mutta näillä ei voi havaita unitilojen rakennetta. Kliinisenä standardina unitilojen seurannassa on polysomnografia, jossa samanaikaisesti mitataan mm. potilaan elektroenkelografiaa, elektro-okulografiaa, elektrokardiografiaa, electromyografiaa, hengitysinduktiivisesta pletysmografiaa, happisaturaatiota ja hengitysvirtauksia. Kliinikko suorittaa univaiheluokittelun signaaleista näkyvien, vaiheille tyypillisten, hahmojen perusteella. Työläyden ja häiritsevän mittausasetelman takia menetelmä ei sovellu pitkäaikaiseen seurantaan. On tarvetta kehittää tarkoitukseen sopivia automaattisia ja huomaamattomia unenseurantamenetelmiä.

Tässä työssä kehitettiin ja testattiin sydämen syke-, hengitys ja liikeanalyysiin perustuvia koneluokittimia vauvojen unitilojen havainnointiin. Luokittimet opetettiin kliinisessa polysomnografiassa kerätyllä datalla 48 vauvasta, joiden ikä vaihteli 1. viikosta 18. viikkoon (mediaani 5 viikkoa). Vauvojen hengitystä ja liikkeitä seurattiin Emfit Oy:n valmistamalla elektromekaaniseen filmiin pohjatuvalla patja-sensorilla. Lisäksi ECG:lla seurattiin sydäntä ja opetuksessa käytettiin lääkärin suorittamaa PSG-pohjaista luokitusta. Esikäsittelyn jälkeen signaaleista laskettiin suuri joukko piirrevektoreita (N=81), joihin luokittelu perustuu.

NREM3-univaiheen tunnistus onnistui parhaiten 0.822 mediaani-tarkkuudella ja $[0.724, 0.914]$ kvartaaleilla. Tulos on yhtenevä kirjallisuudessa esitettyjen arvojen kanssa ja vastaa kliinikkojen välistä toistettavuutta. Muilla luokittimilla univaiheet sekoituivat keskenään, mikä on oletettavasti selitettävissä aikuisista poikeavalla REM-unen aikaisella lihasjäykkyydellä ja kehon liikkeillä.

Työ osoittaa, että menetelmällä voi seurata vauvojen uniluokkien oskillaatiota. Järjestelmää voisi käyttää kotiseurannassa tai vastasyntyneiden teholla unenvalvontaan.

Asiasanat:	Vauvat, unenluokittelu, patja-sensori, elektrokardiografia, polysomnografia, elektromekaaninen filmi, tukivektorikone
Kieli:	Englanti

Acknowledgements

Thank you to my instructors, Nathan Stevenson and Sampsa Vanhatalo, for introducing me to the fascinating world of infant sleep analysis and also for your never ending patience.

Espoo, May 5, 2018

Jukka Ranta

Abbreviations and Acronyms

AAS	Activating arousal system
ACC	Accuracy
Acti	Actigraphy
ADASYN	Adaptive Synthetic sampling
ANS	Autonomic nervous system
AUC	Area under ROC curve
AS	Active sleep
BCG	Ballistocardiography
BMS	Bed mattress sensor
CSA	Central sleep apnea
ECG	Electrocardiogram
EEG	Electroencephalogram
EMG	Electromyogram
EMFi	Electromechanical film
EOG	Electroculogram
FFT	Fast fourier transform
HBI	Heart beat interval
HF	High frequency
HR	Heart rate
HMM	Hidden Markov model
IQR	Inter quartile range
k-NN	k - Nearest Neighbor classifier
LOPO-CV	Leave-one-patient-out cross validation
NN	Neural network
NREM	Non rapid eye movement sleep
OSA	Obstructive sleep apnea
PSD	Power spectral density
PSG	Polysomnography
QS	Quiet sleep
RCL	Respiratory cycle length
REM	Rapid eye movement sleep
RIP	Respiratory inductance plethysmography
ROC	Receiver operating characteristic
RR	Respiratory rate
SVM	Support vector machine

Contents

Abbreviations and Acronyms v

1 Introduction 1
- 1.1 Sleep . . . 2
  - 1.1.1 Neurobiology of sleep . . . 3
  - 1.1.2 Heart and respiration rate control . . . 5
  - 1.1.3 Infant sleep related characteristics and their development 6
  - 1.1.4 Effects of infant sleep problems . . . 7
- 1.2 Infant sleep staging . . . 7
  - 1.2.1 Gold Standard - polysomnography . . . 8
  - 1.2.2 Unobtrusive approaches for sleep tracking . . . 9
  - 1.2.3 Automated cardiorespiratory and movement based sleep classification . . . 10
- 1.3 Proposed method . . . 13

2 Materials and methods 14
- 2.1 Data . . . 16
- 2.2 Preprocessing . . . 17
  - 2.2.1 Identification of anomalies . . . 17
  - 2.2.2 Infant presence . . . 18
  - 2.2.3 Identification of gross body movements . . . 18
  - 2.2.4 Identification of respiration cycle . . . 20
  - 2.2.5 Determination of heart beat intervals . . . 21
- 2.3 Feature extraction . . . 21
  - 2.3.1 Waveform . . . 23
  - 2.3.2 Spectral . . . 25
  - 2.3.3 Non-linear dynamics . . . 26
  - 2.3.4 Miscellaneous . . . 29
  - 2.3.5 Postprocessing of features . . . 30
- 2.4 Support vector machine . . . 31
  - 2.4.1 Margin optimization . . . 31
  - 2.4.2 Multiclass SVM . . . 33
  - 2.4.3 Hyperparameter optimization . . . 34
- 2.5 Postprocessing . . . 34

3 Evaluation **35**
3.1 NREM3 - Rest 35
3.2 Awake - NREM - REM 37
3.3 Awake - REM - NREM1-2 - NREM3 38
3.4 Postprocessed NREM3 vs Rest 39

4 Discussion **45**
4.1 Comparison to previous work 45
4.2 Inter-observer agreement 46
4.3 Future directions 47

5 Conclusions **48**

A PSG and EEG infant sleep classifiers **63**

B Postprocessed performances and hypnogram examples **64**

Chapter 1

Introduction

Longstanding sleep monitoring is required to assess the role of sleep for infant cognitive, psychomotor, temperament and developmental outcomes. Currently infant sleep behavior can be assessed via sleep questionnaires or actigraphy, which provide knowledge of awake/sleep cycles but lack insight as to what happens within sleep. To acquire understanding on the infant sleep stages one must rely on polysomnography (PSG), where several obtrusive sensors are attached to an infant to record the changing cortical activity and the effects of the autonomous nervous system on circulation and respiration in addition to eye movements and gross body movements. The PSG is the gold standard in sleep assessment but requires a clinician or sleep technician to manually annotate the sleep stages from the raw signals. Because of obtrusiveness and laboriousness it is not suitable for long term monitoring.

There has been great interest in scientific community to circumvent the tediousness of PSG and several approaches to assess the sleep stages based on only limited selection of signal modalities have been proposed, that take an advantage of automated signal analysis and machine classification. This work considers the applicability of the cardiorespiratory phenomena and gross body movements for infant sleep classification. Clinically gathered infant sleep monitoring data of a bed mattress sensor and electrocardiogram (ECG) along side with the manually annotated hypnograms are used to train a support vector machine classifiers with different sleep state partitions.

Because sleep originates from the brain this Chapter continues with an overview of the current neurobiological understanding of sleep and summarizes also the relevant infant sleep related aspects. In addition, the manual infant PSG is reviewed with a discussion of previously proposed cardiorespiratory and gross body movements based classifiers. The Methods Chapter provides full breakdown of the proposed signal processing and machine learning pipeline. In the final Chapters the infant sleep classifiers performances are evaluated and discussed.

1.1 Sleep

Sleep is characterized with a lack of resposiveness to external stimuli [1] and reduced state of conciousness. Several causes for sleep has been hypothesized such as sleep being a recovery process, maintaining brain plasticity and associated with learning in addition to being an adaptation to a circadian environment [2, 3]. Sleep has been traditionally thought as an endogenous phenomena of the central nervous system [4]. However, sleep deprivation studies have shed light on the importance of sleep on immune, cardiovascular [5], metabolic [6] and musculoskeletal function[7]. In animal models, wake and sleep are characterized with different transcriptional profiles [8, 9]. One interesting finding is that in the epidermal stem cells of young mice the maximal energy metabolism (oxidative phosphoration) and DNA replication are diurnally segregated [10, 11]. This has been argued to be critical for avoiding DNA damage due to oxidative stress and thus minimizing cellular damage, which is though to be the general cause for senecense [12]. In this paradigm, sleep is a body-wide synchronization mechanism to limit DNA synthesis so that during high activity periods all cells are collectively capable of tasks requiring high performance and energy consumption. Despite the fact that extensive sleep research has uncovered interesting observations and provided several hypotheses, the complete picture is still unknown. The observed necessity for sleep might only be solvable when the definitive explanation of conciousness is reached and associated with the holistic picture of physiology.

As wake and sleep are characterized differently so can sleep be further divided into distringuishable stages. In 1950s Aserinsky et al. reported cycling periods of jerky eye movements within sleep [13], which have lead to the partition of sleep into rapid eye movement (REM) and non-REM (NREM) stages. Ultradian REM/NREM cycle takes approximately from 60 min in children [14] to 90 min in adults [2]. The proportion of sleep stages changes due maturation. Infants spend typically 50% of the total sleep time in REM whereas in adults this is reduced to 20% [15]. Besides the characteristic eye movements REM is associated with more irregular respiration and heart rate in addition to penile erection and an electroencephalography (EEG) pattern that closely resembles the wake state [2]. Also parasympathetic versus sympathetic dominance is different between NREM and REM, respectively. In the endocrine system level the release of growth hormone peaks and cortisol levels decrease during the deepest stage of NREM [16]. Dream recollection studies have shown that REM sleep produces significantly more vivid dreams, which can be recalled more precisely after waking up [2].

Sleep also has a significant impact on memory and learning. During sleep synaptic potentiation affects memory consolidation and de-potentation acts as a re-balancing method for forgetting irrelevant information enabling brain plasticity and learning during the next awake period [17]. It has been proposed that information is selectively tagged during awake for sleep time consolidation and then can be even further processed by integrating it with previous knowledge or generilized to find underlying rules or patterns [18]. The dual process hypothesis assumes that REM sleep would be beneficial for non-declarative memories (procedural and emotional), whereas NREM sleep would have higher impact on

declarative memories (episodic and semantic), but recent studies suggest that memory encoding is more intertwined between the states [19].

1.1.1 Neurobiology of sleep

Since the early observations of sleeping sickness epidemic *encephalitis lethargica* in 1910s conducted by Constantin von Economo [20], which have been followed by several brain lesion, stimulation, pharmagolocial and optogenetic studies, to mention but a few techniques, have provided insight into brain regions and circuitry that governs wakefulness and sleep. Many important nuclei, their connections and neurotransmitters have been proposed and identified. Ascending arousal system (AAS, also known as reticular formation) spanning from pons, midbrain and posterior hypothalamus activates the cortex and is responsible for wakefulness. Sleep, both NREM and REM, have also their own characteristic nuclei and networks. The circuitry is characterized with self-reinforcement and mutual inhibition to achieve flip-flop type switch between states to ensure rapid transitions and stability. At macroscopic level sleep wake cycle is commonly thought to be regulated by two parallel processes managed by the circadian and homeostatic factors [2].

The AAS consists of several lower brain regions that are thought to promote specific aspects of behavioral arousal independently but they are, nevertheless vastly interconnected. In the pons, close to the fourth ventricle, is the locus coeruleus (LC) nucleus which is the major source of norepinephrine to the forebrain and is associated with stress and cognitive challenge related to stimuli. In the midbrain, there are several important nuclei. The dorsal raphe nucleus (DR) and other seratonergic neurons of different raphe nuclei scattered around brain stem innervate preoptic area, basal forebrain, hypothalamus and thalamus, to name but a few targets, are the most active during awake. Cholinergic latero dorsal and pedunculopontine tegmental nuclei (LDT/PPT) project e.g. into thalamus, where they enhance the relay of information from peripheral nervous system to cortex and thus are responsible for thalamocortical activation. Dopamine (DA) is thought to promote arousal when a subject is highly motivated and physically active and the important sources are substantia nigra (SN), ventral tegmental area (VTA) and ventral aqueductal gray (vPAG). In the posterior hypothalamus is the histaminergic tuberomammillary nucleus (TMN) and it is thought to be important in initiating arousal. The lateral hypothalamus contains hypocretin (orexin) neurons and with their vast connections to other AAS nuclei they are though to maintain wakefulness. These cells have been observed to be lower in numbers in narcolepsy, a sickness where an individual suddenly falls asleep. Several regions of basal forebrain (BF) have cholinergic and γ-aminobutyric acid (GABA) neurons that promote cortical activation by targeting different cell types. [21, 22]

For NREM sleep, two important regions are the median preoptic area (MNPO) and ventrolateral preoptic area (VLPO). The former has been suggested to help in sleep initiation, whereas the latter is thought to be required for maintenance of sleep. They project all around the arousing system with inhibitory neurotransmitters GABA and galanin. The BF is also important in cortical NREM phenomena. [21, 22]

The pons has been recognized to be paramount for REM. The sublaterodorsal tegmental nucleus (SLD) play a central role in manifestations of several REM

phenomena. Also some brain regions associated with AAS contribute to REM sleep. LDT/PPT have REM active cells, whereas in lateral hypothalamus in same area with hypocretin cells are melanin-concentrating hormone/ GABAergic neurons that inhibite the same targets that hypocretin cells excite. [22, 23] Figure 1.1 shows the sagittal plane of the brain with the most prominent sleep associated nuclei.

Macrolevel phenomena are responsible for cyclic behavior of the nuclei mentioned above. The circadian process is an endogenously generated biological rhythm and besides its most obvious phenomena, sleep wake cycle, it influences daily oscillation of alertness, temperature, hormonal secretion and blood pressure. The circadian system has a pacemaker suprachiasmatic nucleus (SCN) located in the anteroventral hypothalamus and its cells oscillate autonomously because of the delayed transcriptional negative feedback loop of PER/CRY genes. Cells are also coupled together to assure in-phase oscillation of the whole nucleus, but this rhythm can be shifted by external cues called zeitgebers. Light is the most powerful zeitgeber and retinal ganglion cells directly synapse SCN. Light also affects the secretion of sleep-promoting hormone melatonin from pineal gland. [24, 25]

The homeastatic process is responsible of the internal neurophysiological drive towards sleep. During awake periods sleep promoting factors, somnogens, accumulate in central nervous system increasing the drive to sleep and dissipate during sleep. Adenosine has been thought to be the most prominent somnogen. It is a by-product of cell metabolism and activates ventrolateral preoptic nucleus (VLPO) and thus promotes sleep. Cytokines and prostaglandines among other potential substances have been also argued to act as somnogens. [22]

Besides the sleep control network, the brain has specific regions for respiratory, motor and circulatory system regulation. Sleep manifests in these systems, although the exact neural link from sleep areas is unknown.

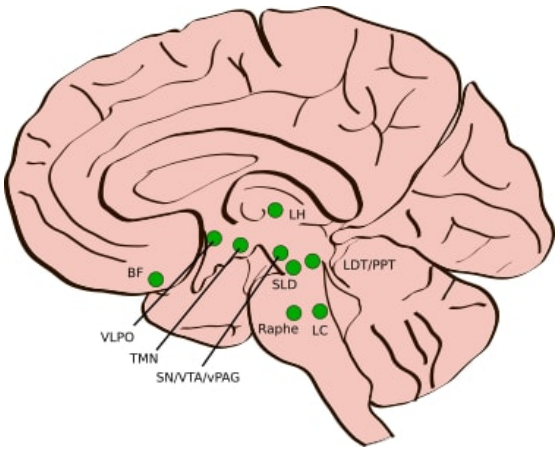


Figure 1.1: Deep brain nuclei and areas relevant for sleep. Basal forebrain (BF), ventolateral preoptic area (VLPO), tuberomamillary nucleus (TMN), lateral hypothalamus (LH) substantia nigra (SN), vetral tegmental area (VTA), ventral periaquaductal gray (vPAG), sublaterodorsal nucleus (SLD), Raphe nuclei, laterodorsal and pedunculopontine tegmental nuclei (LDT,PPT) and locus coeruleus (LC).

1.1.2 Heart and respiration rate control

The pacemaker potential created by the sinoatrial (SA) node in the right atrial wall initiates the contraction of heart and its firing frequency determines the heart rate (HR). The SA node has a capability of autonomous depolarization but it is modulated by several factors such as autonomic nervous system (ANS), hormones (e.g. epinephrine) and cation levels in the circulatory system. Both sympathetic and parasympathetic divisions of ANS influence HR via thoracic spinal nerves (norepinephrinergic) and vagus nerves (cholinergic), respectively. The former stimulation increases and the latter decreases HR. Both of these originate from the cardiovascular center (CV) in the medulla oblongata, which is also responsible for heart contractility and vasomotor tone. More specifically sympathetic control arises from the restroventrolateral medulla and parasympathetic control from the nucleus ambiguus, dorsal motor nucleus of the vagus and intermediate zone of medulla. The CV receives input from several sensory receptors (proprioceptors, chemoreseptors, barorespetors), limbic system and cortex. [26, 27]

Change in the size of the thoracic cavity enables in- and outflow of gases from the lungs; respiration. The change of thorax size is controlled by the surrounding muscles and elastic recoil of tissues. During inspiration, the diaphragm and external intercostal muscles contract, whereas exhalation is a result of diaphragm relaxation and passive elastic recoil. This is typical in quiet breathing and is different in forceful breathing, e.g., in exercise where more muscles are recruited and expiration becomes an active process. Respiration is controlled by the respiratory center, which involves clusters of neurons in the medulla oblongata and pons. It can be functionally divided into three areas. Respiration rhythm is generated in medullary rhythmicity area, where inspiratory area (pre-Bötzinger complex) neurons link to external intercostal muscles via intercostal nerves and to the diaphragm via phrenic nerves and where expiratory area (retrotrapezoid nucleus/parafacial respiratory group) influences internal intercostal and abdominal muscles during forceful exhalation. The medullary rhythmicity area is influenced by pontine pneumotaxic (inhibits) and apneustic (stimulates) areas. The respiratory center receives input from chemoreceptors in the carotid bodies, aortic bodies and medulla monitoring levels of O_2, CO_2, and pH in addition to bronchial stretch receptors and multiple other brain areas (e.g., limbic system and cortex). Chemoreceptors are responsible for hypoxic and hypercapnic ventilatory drives.[26, 28]

The linkage between sleep associated brain areas and cardiovascular and respiratory centers is elusive. Although there are several cardiorespiratory signs that differ between the sleep stages. Both heart and respiratory rates are typically lower in sleep compared to awake and further decreased in NREM [2]. See Table 1.1 for basic rates during the first six months of life reported by Heimann et al. [29]. Although they obtained slighty higher median respiration rates during NREM sleep for the first 2 months of infancy. The other major difference between REM and NREM is heart and respiratory rate regularity. Heart rate variability derived from heart beat intervals are more variable in REM due to sympathatic tone and the influence of behavioral brain areas [30]. NREM is charactericed by parasympathetic vagal tone and besides lowering the heart rate it is responsible for creating respiratory sinus arrhytmia (RSA). RSA is characterized by decreased heart beat intervals (HBI) during inspiration and

prolonged during expiration which improves pulmonary gas exchange and saves energy [31]. Oscillating higher vagal tone in synchrony with expiration suppresses unnecessary heart beats. By doing spectral analysis on the HBI series this phenomena can be seen in the frequency range of 0.15-2Hz, which is typically narrowed down depending on the age [32]. In NREM, regular respiration is primarily controlled by carbon dioxide levels. In REM sleep, tone of intercostal muscles decreases so that respiration relies more on the diaphragm. Because the infant chest has high compliance this might result in paradoxical chest and abdominal movement. Also the decreased tone in the muscles of the upper airway increases the risk for obstructive apnea events and periodic breathing during REM. [31, 33]

Table 1.1: Basic heart rates (HR) and respiratory rates (RR) and their inter-subject variations during the first 6 months of infancy. Age is in months and N indicates sample size. AS and QS are active (REM) and quiet (NREM) sleep stages, respectively. Adapted from [29].

		Median HR (IQR) beats/min		**Median RR (IQR) breaths/min**	
Age	**N**	**AS**	**QS**	**AS**	**QS**
1	47	134.7 (129.8-141.8)	132.1 (126.2-138.7)	35.8 (32.5-44.0)	37.9 (33.0-44.0)
2	82	131.0 (125.8-137.0)	127.0 (121.4-133.9)	32.4 (29.0-36.0)	33.6 (28.9-38.0)
3-4	68	126.1 (120.6-130.9)	120.6 (109.4-120.6)	29.9 (26.5-33.1)	29.0 (25.0-32.2)
5-6	33	120.0 (114.3-123.6)	113.3 (107.8-120.0)	27.0 (24.0-29.8)	25.3 (22.3-28.3)

1.1.3 Infant sleep related characteristics and their development

Besides cardiorespiratory characteristics, infant sleep shows significantly different patterns in other physiological measurements too. Electroencephalography (EEG) opens a window on the cortical function and it has a pivotal role in clinical sleep studies. The current polysomnography based clinical sleep stage annotation heavily relies on EEG and it's stage specific characteristics. During infancy the sleep structure and faces major changes.

Preterm studies indicate that the sleep as a behavioral state emerges by 28-32wk gestational age (GA). In utero, the fetus is under the influence of the mother and it has been observed that the fetus is synchronised with the mother's circadian rhythm. Although after birth the rhythm becomes ultradian due to the immature circadian clock, high caloric demand and inability to retain calories relative to size. EEG-based sleep-stage-specific patterns are visible around 28 wk GA. REM (active) sleep is characterized at this point with continuous mixed-frequency low amplitude EEG activity including rapid bursts in EOG and low-amplitude EMG superimposed with twitches and phasic jerky movements. In NREM (quiet) sleep EEG has continuous slow frequency pattern or so-called trace alternant where 2-8 sec low-frequency pattern is intertwined with 4-10 sec low-voltage mixed-frequency pattern. In EOG there are no eye movements

and the EMG has low amplitude. The sleep cycle is approximately 45 min. [15, 24, 34]

Infants born at term sleep around 16-18 hours a day and wake up for feeding. Sleep onset is characterised by the REM stage for the first 3 months which is atypical compared to adults. The cycle length increases to 70 min, where the distribution of NREM and REM are equal. Around 6-8 weeks post-natal age the first major EEG feature, sleep spindles, begins to appear. This activity is characterised by a 12-14 Hz burst that reflects the development of thalamocortical structures and is associated with memory consolidation. Also around 2-3 months slow wave activity (delta) appears for the first time. This activity is associated with the deepest adult NREM state. Slowly the proportion of REM in total sleep time begins to drop and is about 40% around 3-5 months. K complexes emerge around 5 to 6 months with vertex sharp waves. The total sleep time drops and circadian rhythm becomes diurnal. Awake state characteristic dominant posterior rhythm increases and approaches adult frequencies. [15, 24, 34]

1.1.4 Effects of infant sleep problems

The infant brain doubles its volume from 420 ml to 860 ml during the first year of life [35] with growth depending on the brain area [36]. Because growth is typically performed during the sleep, there has been great interest on how sleep problems and disturbances affect infant development. The typical interest is how the total deprivation, sleep onset time, duration, number of arousals and efficiency/fragmentation influence on physiological and developmental factors. In practice most often these studies have been conducted using sleep questionnaires and more seldom with direct physiological measures (e.g., polysomnography or actigraphy) or video recordings. Differences in study methods have led to contradictory results, as such the causal link between sleep and cognitive, psychomotor and temperament developmental outcomes have not been established [37–39].

Sleep disturbances correlate with some developmental outcomes. Reduced sleep duration increases the risk of childhood obesity [40, 41] which is in line with observations from adult sleep studies and animal models [42, 43]. It has been proposed that this could be due to disturbances in levels of appetite and energy metabolism related hormones leptin and ghrelin [44]. Lower sleep quality has also shown to increase the risk of social-emotional, attention regulation and behavior problems [45, 46]. Interestingly it has also been reported in adult patients that growth hormone secretion is impaired when the amount of deepest NREM sleep is reduced due to obstructive sleep apnea [47].

1.2 Infant sleep staging

The earliest proposed infant sleep staging was based on behavior and arousal thresholds [48] and since the discovery of REM sleep in the 1950s assessments shifted towards physiological measurements. Aserinsky observed cycling of jerky eye movement periods in adults and infants which led to the definition of sleep stage specific characteristics based on the electro-oculogram (EOG), EEG, gross body movements and respiration [13, 49–51]. Studies by Dreyfus-Brisac and

Monad [52] and Parmelee et al. [53, 54] defined corresponding characteristics for infants and discussed their evolution during development.

This incorporation of multiple different physiological measurements in sleep staging is called polysomnography (PSG) with chin electromyography (EMG) and ECG also typically included. Based on patterns in these signals, the sleep stage is visually scored by a trained clinician or a sleep technician.

1.2.1 Gold Standard - polysomnography

Polysomnography has remained the gold standard in clinical sleep staging and it is used as a diagnostic tool for sleep apnoea, neuromuscular disorders and chronic lung diseases [55]. Although various criteria and stage divisions have been proposed, it is typically up to cultural preferences and the clinicians' personal experiences when defining sleep state.

For adult sleep staging the most influential manual was written by Rechtschaffen and Kales [56] in 1967. An infant sleep staging manual was published in 1971 "*A Manual for Standardized Teminology, Techniques, and Criteria for Scoring of States of Sleep and Wakefulness in Newborn Infants*" by Anders et al. [57] and it has remained influential in the field. The manual was directed for 0-2 month olds and they defined three sleep stages and three awake stages. Active and quiet sleep were thought to be precursors of REM and NREM sleep, respectively. An indeterminate state was assigned when characteristics from both states were apparent. The awake stage was divided into crying, active and quiet awake. Several updated manuals and alternative criteria have been proposed by authors such as Prechtl [58], Hoppenbrouwers [59], Quilleminault and Souquet[60], Crowell et al [61] and Scholle and Schafer [62].

Today one of the most impactful authority in the field is the American Academy of Sleep Medicine and they publish updated manuals on a regular basis. In 2007 they published a comprehensive literature review and assessed infant sleep scoring criteria [63]. Their recommendation were for children over 2 months, although the scoring criteria could be extended to younger infants by incorporating Anders criteria. In their criteria, sleep is assessed in 30 sec epochs and divided into REM and three different NREM classes based on increased arousal threshold: N1-N3. Criteria are summarized in Table 1.2.

Because PSG is methodologically obtrusive and laborious several unobtrusive observational methods suitable for long-term sleep monitoring in addition to automated scoring approaches have been proposed in the past decades.

Table 1.2: American Academy of Sleep Medicine sleep scoring criteria for infants > 2 months. DPR - dominant posterior rhythm and SWA - slow wave activity.

	REM	NREM		
		N1	N2	N3
EEG	low voltage mixed activity	DPR < 50% and one of the flowing: 1) diffuse low voltage mixed frequency, 2) hypnogogic hypersynchrony 3) rhytmic anterior theta of drowsiness, 4) diffuse high voltage occipital delta slowing 5) burst of diffuse rhytmic 3-5hz slowing 6) vertex sharp waves 7) post-postarousal hypersynchrony		
		-	K complexes sleep spindles SWA< 20%	20% >SWA
EKG	irregular	-		
Respi	irregular	mostly regular		
EMG	decreased/absent	mostly present		
EOG	rapid eye movements	-		

1.2.2 Unobtrusive approaches for sleep tracking

There has been a progressive development of unobtrusive sleep tracking systems, which has led to a rise of consumer grade applications for adult vigilance state monitoring in the past years. Unobtrusive methods typically record cardio-respiratory measures and body movements and proposed applications to observe these include ballistocardio-, seismocardio-, acti-, photoplethysmo-, thermography, capacitive electrodes, radars, lazers, electrical/ magnetic impedance and video recordings [64, 65].

Besides the PSG, actigraphy is the only method that has been widely accepted and deployed in clinical setting [66, 67]. Actigraphy measures physical motion using a miniaturized acceleration sensor and it is mainly used to discriminate wake from sleep state in adults and infants. Many actigraphy-based consumer products have been also released into the market for activity monitoring and determining sleep duration.

Ballistocardiography (BCG) is a method of detecting the fine movement of the body that is generated by the pumping heart and blood circulation [68]. The signal is typically superimposed on the respiratory signal and cross body movements. A bed mattress sensor (BMS) is one solution among several to record the respiratory signal, body movements and BCG and there exist different technical implementations such as static charge sensitive bed (SCSB), polyvinylidene fluoride and electromechanical films (EMFi)[69–71].

The BMS has been previously reported to provide enough information for infant sleep staging. In 1979, Thoman and co-workers reported a study where they used a mattress sensor to record cross-body movements and breathing in 9 infants during a 2 hour sleeping session [72]. Four manual annotators inspected the signal based on breathing regularity and amount of movements. Total, active and quiet sleep inter-observer agreements had values of 84.0%, 82.9%

and 80.0%, respectively. There were full agreement on awake state. They also compared the method to behaviour based scoring which yielded 81.9%, 79.0% and 81.3% agreements, respectively.

In 1990, Erkkijuntti et al. defined a scoring criteria for their SCSB based approach [73]. In their scoring schema points were given based on respiration amplitude and cycle length variability, regularity of ballistocardiogram and movement duration for each 20 sec epoch. Two clinical neurophysiologists applied scoring for 8 infant full night recordings and achieved 68.1% and 64.1% concordance to Anders criteria.

The work of Kirjavainen et al. devised 1-min epoch scoring criteria that utilized novel manual detection of paroxysmal phasic activation events of motor and autonomic nervous system that could be detected in 6-16 Hz SCSB signal band as high amplitude - short duration phenomena [74]. In their criteria, an epoch was scored awake, if there was more than 30 sec of movement. NREM was scored if the spiking was absent or regular and for REM if the required presence of irregular spiking with simultaneous chaotic respiration. If the breathing was regular but the spiking was irregular, the epoch was scored intermediate (I). They evaluated the scoring performance in 22 babies against Guilleminault & Souquet criteria (< 1 year) and Rechtschaffen & Kales rules. If the I stage scoring (totaling 12.7% of epochs) was interpreted as NREM they achieved the total agreement of 85% with PSG. Similarly the sensitivities had the values of 90% (NREM), 61% (REM) and 85% (Awake).

1.2.3 Automated cardiorespiratory and movement based sleep classification

Because manual sleep scoring is a laborious process there is great interest in developing automatic sleep assessment approaches. Several automatic PSG based sleep classifiers have been proposed in the literature for both infants and adults alike. See Table A.1 in Appendix A for a short collection of proposed infant PSG sleep classifiers. There are also several adult PSG equipment manufactures that provide automatic decision support systems. For actigraphy the awake/sleep identification can be accomplished using algorithms such as those proposed by Redmond [75], Cole [76], Sadeh [77] or Galland [78].

Although this work utilizes a bed mattress sensor signal and ECG, it is meaningful to cover previously proposed classification approaches that use cardiorespiratory and gross-body movements irrespective the actual signal acquisitional method. There exist several approaches for adult sleep classification in literature [79–93], which are summarized in Table 1.3. Similarly, multiple methods have been proposed for infants (Table 1.4).

Table 1.3: Collection of proposed automatic adult sleep classifiers for cardiorespiratory and movement signals. Abbreviations: N - number of subjects, Perf - reported performance (method might vary), BMS - bed mattress sensor, RIP - respiratory inductance plethysmography, BCG - ballistocardiography, Acti - actigraphy.

Year	Authors	Signals	Method	N	Classes	Perf
1993	Jansen et al. [79]	BMS	Quadratic maximum likelihood	6	awake, REM, N1, N2, N3/4 and awake, REM,NREM	52-75% and 86-98%
2007	Redmond et al. [80]	ECG, RIP	Linean and quadratic discriminant	31	awake, sleep,REM	89%
2008	Karlen et al. [81]	ECG, RIP, Acti	Neural network	3	awake,sleep	96%
2009	Mendez et al. [82]	BMS (BCG)	HMM	6	awake,REM,NREM	83%
2009	Mack et al. [83]	BCG, resp, movement	NAPS algorithm	8	awake,sleep	84%
2009	Mendez et al. [84]	ECG	HMM	17	awake,REM, NREM	78%
2010	Migliorini et al. [85]	BMS (BCG)	Linear and quadratic discriminant	17	awake,REM,NREM	77%
2010	Mendez et al. [86]	BMS (BCG)	k-NN (Neural network)	17	awake,NREM,REM	72% (67%)
2010	Kortelainenet al. [87]	BMS (BCG, movement)	HMM	18	awake,NREM,REM	79%
2010	Chung et al. [88]	Load Cell (BCG)	Threshold	10	awake,sleep	72%
2011	Isa et al. [89]	ECG	k-NN, Random forest, SVM	16	awake,REM,N1, N2,N3,N4 and awake,REM,N1/2, N3/4	(4-stage) 53%, 60%, 57%
2012	Austin et al. [90]	Load Cell (movement)	SVM	27	awake,sleep	81%
2012	Guerrero-Mora et al. [91]	BMS (resp)	Neural network, random forest, regression	20	awake,sleep	82%
2015	Long [92]	ECG, RIP	several	N/A	several	N/A
2017	Kaji et al. [93]	resp, movement	LSPC	10	awake,REM,N1/2, N3/4	68%

In 1987 Thoman et al. created the first reported automated infant sleep detection algorithm for the mattress sensor signal [94]. Their data contained respiration and body movements from 10 infants recorded over a period of 2 hours. Their method was based on a comparison of feature vector histograms. From 30 sec epochs they calculated the histograms for the amplitude, the intervals for the zero crossings and slope changes along side with the amplitude difference between slope change intervals. The histogram bins were compared to the pool of prototype epochs and the closest (Manhattan distance) prototype class was chosen. Automatic active sleep (AS), quiet sleep (QS), AS-QS transition, sleep-wake transition and wake classification was compared to a manual behavior based scoring [95]. Overall mean accuracy had a value of 80.6% and standard deviation 5.1%.

In same year Haddad et al. published classification results using cardiorespiratory measures derived via ECG and barometric plethysmography from 14 infants (1 and 4 months) [96]. From the former they derived HBIs and from the latter respiratory cycle times and tidal volumes. They calculated the mean and coefficient of variations for 30 sec epochs and by comparing the distributions between REM and QS sleep (omitting intermediate sleep and noisy epochs) they derived a threshold for each variable that maximizes the Kolmogorov-Smirnov test statistic. They concluded that the coefficient of respiratory cycle time outperforms the others and with extended 5 min epochs they reported that 21/22

and 20/21 of the epochs, in QS and REM, respectively, were correctly classified in QS and REM, respectively.

Harper et al. created a set of discriminant function classifiers based on heart and respiratory rates and variabilities in addition to a HBI spectral analysis [97]. The features they calculated from ECG and CO_2 measurements. Their data set contained 25 infants and they built an age specific classifier for each month from 1 to 6 months. The overall reported accuracy of 84.8% was determined against the Anders scoring criteria.

In 2002, Hukka evaluated commercial Biomatt BR99 software designed for the static charge sensitive bed [98]. Three parameters were chosen for classification, namely, the amount of small movements, variation of respiratory frequency and variation of respiratory amplitude. The data set contained 15 infants and results yielded 59% agreement with the manual PSG annotation.

In 2004 Lewicke et al. created and compared actigraphy and ECG classifiers for sleep/awake detection [99]. From the ECG, the group selected mean HBI of the current and previous 8 epochs (30 sec) for feature vectors. From actigraphy, they used maximum value of 9 epochs as features. They used collaborative Home Infant Monitoring Evaluation (CHIME) data base and trained Learning Vector Quantization neural network (LVQ-NN) with the first half of the patients (13) and the other half (12) was left for testing. Actigraphy yielded better sleep prediction with an accuracy of 92.3% compared to ECG, which had a value 89.5%. Although mean HBI recognized awake better 56.5% compared to 42.4%.

The same group published more extensive analysis that included 190 infants and multiple ECG classifiers such as the multilayer perceptron neural network, support vector machine and LVQ-NN [100]. Using the same feature vector, SVM outperformed other methods only by a small margin, detecting 78% (78.5%) sleep (awake) epochs correctly. They also compared multiple single feature binary SVM classifiers for detecting awake from sleep and active from quiet sleep [101]. As signal sources they used actigraphy and respiratory inductive plethysmography. They reported that the coefficient of variation of tidal volume discriminated awake/sleep best with 74.8% accuracy, whereas for AS/QS detection the coefficient of variation of respiratory cycle length was the best feature with 70.1% detection rate.

Terrill et al. [102] provided a fresh perspective to the infant sleep classification by investigating the non-linearities of infant respiratory cycle length series with recurrence quantification analysis. This method evaluates recurrence plots that capture the repeating behavior of phase space embedded signal [103]. By determining the threshold for the percentage of recurrent points in epochs of 25-400 cycles they claimed to correctly classify 88.2-94.4% of the AS and QS epochs.

In 2014, Cohen and Chazal published article on their sleep/awake linear discriminant classifier that was based on multiple features calculated from ECG, pulse oximetry (SpO2) and actigraphy [104]. With data set (CHIME) of 402 infants they achieved an accuracy of 74.1%, a sensitivity of 82.0% and a specificity of 60.9%.

Using simple thresholding for respiratory inductive plethysmography derived respiration cycle length variance Isler et al. achieved 80% AS and 87% QS agreements for 8 infants against manual PSG scoring [105].

Werth et al. calculated a set of HRV features from ECG and trained SVM

based classifier [106]. With a data set of 8 preterm infants they achieved mean AUC of 0.87[1].

Table 1.4: Collection of proposed automatic infant sleep classifiers for cardiorespiratory and gross-body movement signals. Bar. pleth. - barometric plethysmography, CO2 - nasal CO_2 flow, SpO2 - oxygen saturation, AS - active sleep (REM), QS - quiet sleep (NREM). See Table 1.3 for more abbreviations.

Year	Author	Signals	Method	N	Classes	Perf
1987	Thoman et al. [94]	BMS	Template matching	10	AS,QS, Awake, transitions	81%
1987	Haddad et al. [96]	ECG Bar. pleth.	Decision rule	9	REM, QS	84/93%
1987	Harper et al. [97]	ECG CO2	Discriminant analysis	25	AS, QS Awake	85%
2002	Hukka [98]	BMS	Commercial algorithm BR99	15	AS, QS Awake	59%
2004	Lewicke et al. [99]	ECG Acti.	LVQ-NN	13	Sleep, Awake	90/57% vs 92/42%
2006	Sazonova et al. [101]	Acti. RIP	SVM	26	Sleep vs. Awake, AS vs. QS	68-74% 53-70%
2008	Lewicke et al. [100]	ECG	LVQ-NN, NN, SVM	190	Sleep, Awake	78.0/78.5% 77.5/79.0% 78.0/78.5%
2010	Terrill et al. [102]	RIP	Recurrence plot threshold	24	AS,QS	88-94%%
2013	Galland et al. [78]	Acti.	Multiple	31	Sleep, Awake	86/84%
2014	Cohen et al. [104]	ECG Sp02 Acti.	Linear discriminant	402	Sleep, Awake	82/61%
2016	Isler et al. [105]	RIP	Threshold	49	AS, QS	80/87%
2016	Werth et al. [106]	ECG	SVM	8	AS, QS	AUC 0.87

1.3 Proposed method

The development of cardiorespiratory and gross-body movement based automated classifiers for infants have excited great interest. However, there has been quite a long break from utilization and development of a BMS based infant classifiers. In this work, support vector machine classifiers are built on data collected during a normal clinical sleep study which provides authentic challenges to signal processing. These challenges include several artefact sources and noise factors. Respiration and gross-body movements were recorded with an electromechanical film (EMFi) BMS manufactured by Emfit Ltd [107]. The EMFi consists of several charged polypropylene layers separated by air voids and it acts as an electret [108]. ECG was included to provide heart rate.

[1] Area under curve, also known as AUROC - Area under Receiver Operating Characteristic

Chapter 2

Materials and methods

The task at hand was to create and evaluate the performances of 30 second epoch sleep classifiers based on the signals from the BMS and ECG. The classifiers would be trained to recognize several different classifications: 1) NREM3 vs rest, 2) awake vs REM vs NREM and 3) awake vs REM vs NREM1/2 vs NREM3. This ascending order of difficulty would yield thorough analysis of stage separability and method applicability.

Time series classification can be handled via several approaches. One approach is to compare a new time series of interest to a set of class prototypes via some distance metric and select the class based on the nearest prototype. Although the method is fairly intuitive it has pitfalls. It is prone to noise, an appropriate distance measure must be selected and a relatively large prototype library needs to be maintained.

An other approach is to compress the information of the whole time sequence by calculating features and optimize some classification algorithm which discriminates the classes in the feature space. This requires domain specific knowledge to derive useful feature vectors that have sufficient discrimination power. There are several classification algorithms such as k-nearest neighbors, neural networks, random forests and Support Vector Machines (SVM). Feature based methods seem to be the most popular approach in sleep classification literature due the fact that features are derived from knowledge of the underlying physiological function. This approach has been selected in this work with SVM selected as a classifier because it has a well established theory and algorithms, and performs well on small to moderately sized datasets.

When feature based classifiers are utilized, the importance of signal prepocessing is critical. Biological data is inherently noisy, contains disruptions due to lack of contact between patient and sensor and is contaminated with a wide variety of artefacts. In this chapter the data is described and the necessary preprocessing steps are discussed. Also the extraction of respiratory cycle length (RCL) and HBI time series from the sleep mattress signal and ECG, respectively, are explained. These are followed by a thorough introduction to the features which are combined in the SVM classifier. The principles of SVM are discussed with its applicability in multiclass classification. Because the SVM classifier implemented here is epoch-based it lacks temporal insight on the sleep trend. A simple post-processing stage is applied to incorporate temporal information into the classification output. The processing pipeline is illustrated in Figure 2.1.

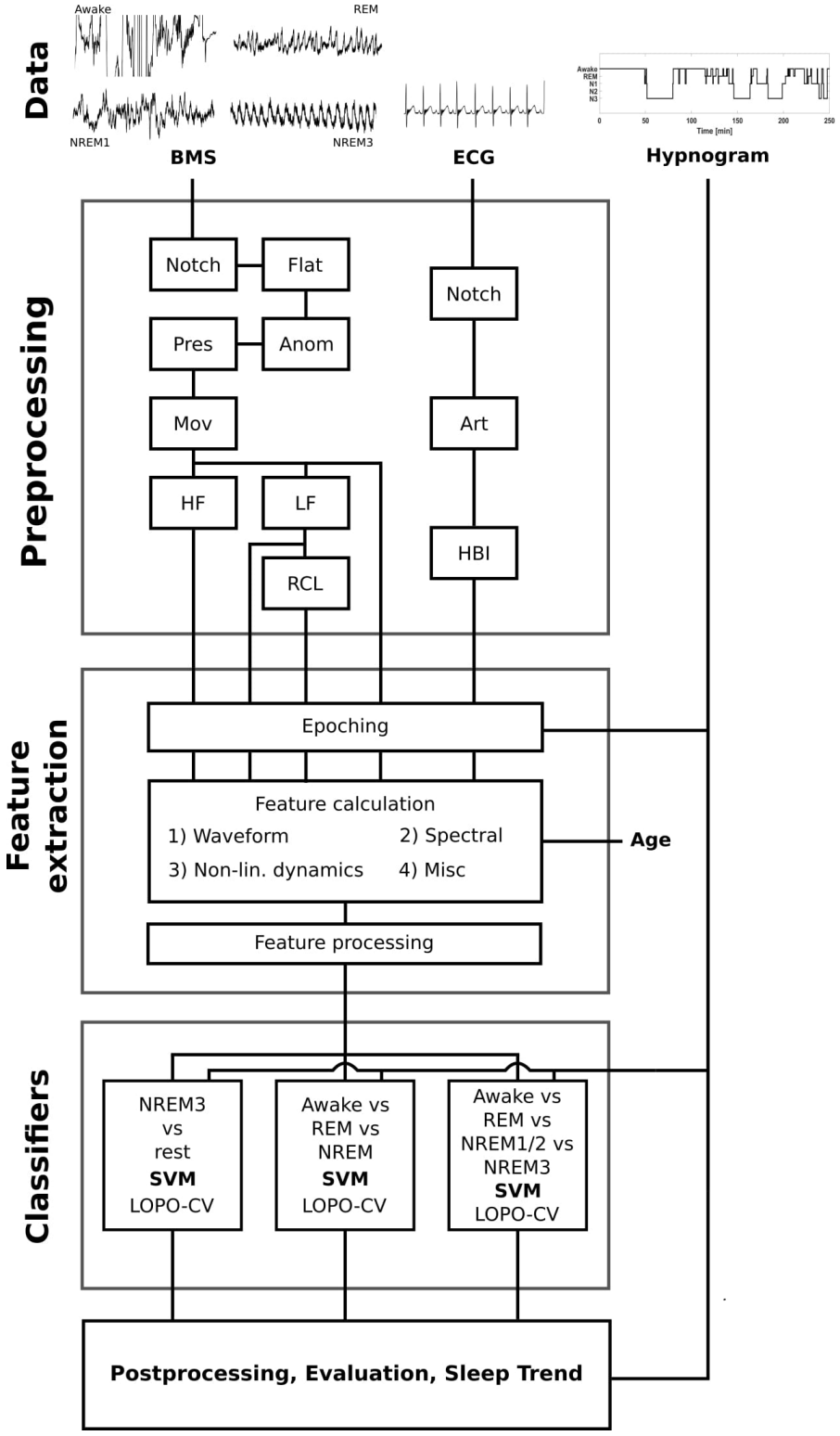
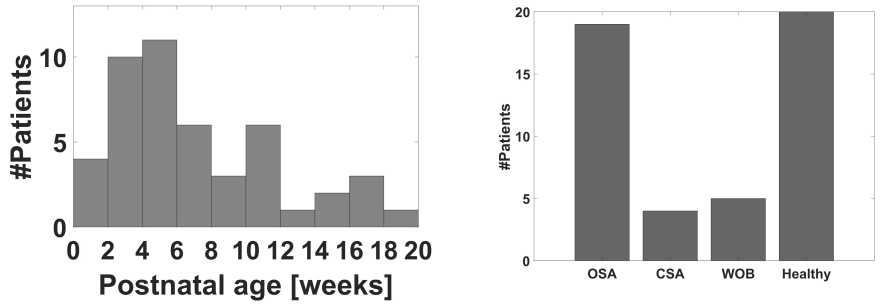


Figure 2.1: Flow chart of the machine learning pipeline applied to train infant sleep classifiers. Abbreviations: Flat - flatline, Anom - anomaly, Pres - infant presence, Mov -Gross body movement, HF - high pass band filter, LF -low pass band filter, RCL - respiration cycle length, Art -artefact, HBI - heat beat interval.

2.1 Data

Data contains sleep mattress, PSG and sleep state annotation recordings of 51 infants which were gathered during a sleep apnoea study. Three patients had corrupted signals and were excluded from further study. The measurement period varied from 2h to 9h with an average of 3h 45 min. The infants were between 1-18 weeks of postnatal age. See Figure 2.2a for age distribution. Over half of the infants included in the study were diagnosed with some breathing difficulty. The numbers of healthy, obstructive sleep apnoea (OSA), central sleep apnoea (CSA) and the patients which had otherwise increased work of breathing (WOB) can be seen in Figure 2.2b.



(a) Age distribution

(b) Breathing problems

Figure 2.2: Age of patients **(a)** and diagnosed breathing problems **(b)**. Note only the 48 patients with successful recordings included. OSA - obstructive sleep apnoea, CSA - central sleep apnoea and WOB - increased work of breathing (WOB cases without OSA or CSA diagnosis).

Manually annotated sleep stages were determined for 30 second epochs. The annotation is based on the visual inspection of the full polysomnographic recording by a skilled clinician and sleep is divided into awake, REM, NREM1, NREM2 and NREM3. The sleep stage epoch distribution is shown in Figure 2.3. Awake is the most common stage whereas REM, NREM1 and NREM3 are well represented. NREM2 is the least common due the fact that it is a transitional state characterized by particular EEG patterns, namely K-complexes and spindles, which can be detected only in older infants (> 2 months) and thus this class is missing in younger infants.

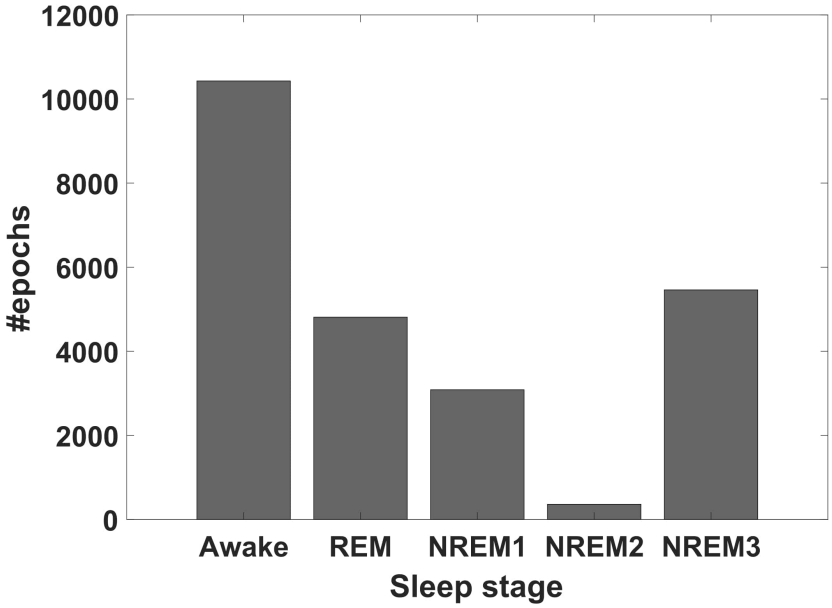


Figure 2.3: Combined sleep stage epoch distribution of 48 patients.

2.2 Preprocessing

Because the data was gathered during a normal clinical study it faces several realistic challenges and need to be addressed with multiple preprocessing steps. The purpose of preprocessing is to remove anomalies and artefacts and set the signal into a state where the feature vectors can be robustly derived. The preprocessing utilized here include band-stop power outlet filtering, signal rejection, signal decomposition into LF and HF sub-bands, and a RCL and HBI signal calculations.

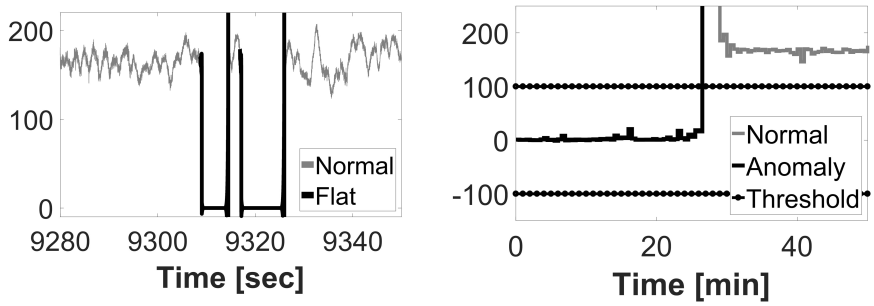
The 50 Hz power outlet signal contaminated every sleep mattress signal and the first step was to remove it with a corresponding band-stop Notch filter. The tasks of detecting the bad quality and cross-body movement time segments are discussed thoroughly in following Sections 2.2.1-2.2.3. The ECG signals were visually inspected for artefacts.

The raw signal was decomposed into two bands using Butterworth bandpass filters. The LF band contains a respiration signal and was separated with 0.2-1 Hz pass band. The HF signal contains ballistocardiographic and increased respiratory resistance components and was set to 6-16 Hz. In addition to these two sub-signals RCL and HBI time series were also calculated. More about these in Sections 2.2.4 and 2.2.5.

2.2.1 Identification of anomalies

The interruption between sensor and measurement device causes a flat line that is characterized with a sudden drop of the signal to near zero. This is illustrated with an example from the data set in Figure 2.4a. The signal level dropped below

the threshold of $\theta_f = 0.01$ for at least a second. The segment was extended for 1 sec down and upstream to include the level drop elements.



(a) Flatline **(b)** Anomaly

Figure 2.4: Signal artefacts. **(a)** Flatline is caused by a loss of connection between the sensor and recorder whereas **(b)** anomalies were characterized by a longer duration low (or extremely high) offsets in median filtered signal.

The anomalies were seen at the beginning of the measurements. There is a recognisable change in the offset of signal, which might be caused by the change in pressure applied to the mattress. Also in some patients it seemed that the machine side filtering worked partially at the beginning causing an anomaly. A simple threshold based anomaly detection was applied. When an epoch median drops below or rises above set thresholds of $\theta_l = 100$ or $\theta_u = 1000$, respectively, the epoch is classified as an anomaly and is not included in further analysis. See Figure 2.4b for an example of epoch wise median filtered signal. Because the offset could get out of the boundaries during excessive cross body movements yet an another rule was added stating that the signals interquartile range should be below $\theta_{iqr} = 150$ during these anomalies.

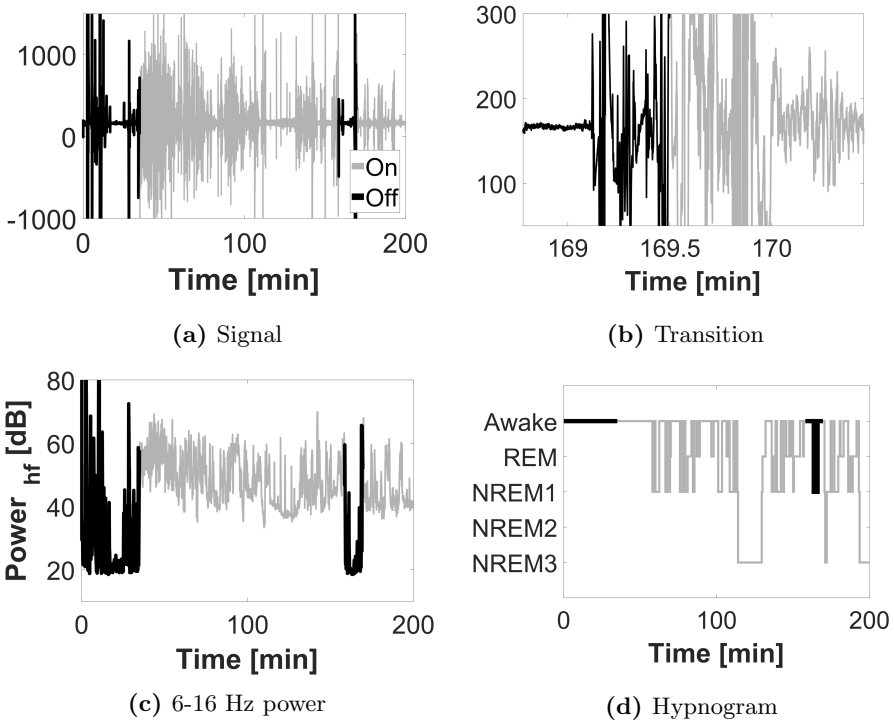
2.2.2 Infant presence

The clinical setting allows the infant to be lifted from the bed for care taking during the PSG study. Because there was no record of these care taking sessions they were deduced from signal. When an infant is removed from the mattress there should be high amplitude spiking followed by a period with very low amplitude without detectable respiration signal and a significant drop of power in 6-16 Hz ballistocardiographic band. An example is shown in Figure 2.5.

Due to individual differences there did not exist a suitable global HF power threshold that could be applied for every patient. Thus the infant absence was deduced visually by inspecting clues for the lack of respiratory oscillation in the raw signal accompanied with a very low signal HF power.

2.2.3 Identification of gross body movements

Because whole body movements are characterized by high amplitude phenomena in the mattress sensor signal and mask all the other signal components such as those induced by respiration, the next step was to identify the gross body movement segments.



(a) Signal

(b) Transition

(c) 6-16 Hz power

(d) Hypnogram

Figure 2.5: Presence of infant on mattress. **(a)** Long duration PSFM signal where two absent periods are highlighted as black and **(b)** higher resolution example of the latter segment where absence is characterized with low amplitude random noise. **(c)** The ballistocardiographic (6-16 Hz) power drops significantly during these periods. **(d)** Corresponding periods in hypnogram.

The approach used is a modified version of the algorithm presented by Schetin and Jakaite where they used adaptive segmentation to derive a discontinuity feature for the EEG signal [109]. The Kolmogorov-Smirnov test applied to power spectral densities of adjacent windows is the basis of segmentation. If the test yields a significant outcome, a border is assigned between them. The maximum observed value of a segment is later compared to a standard deviation based threshold determined from a history window. The algorithm outlined as follows:

1. The signal is filtered with Butterworth high pass filter with a pass band frequency of 0.4Hz to remove slow trends from the data.

2. A FFT based PSD estimate is taken from adjacent non-overlapping short duration windows. The PSD over the frequencies of interest f_b is specified as PSD_b.

3. The Kolmogorov-Smirnoff test is applied for PSD_b of adjacent windows with a critical level of d_0. If the test yields significant result a segment border between the windows is assigned.

4. The segment is deemed to contain gross-body movements if the maximal

amplitude of the time-series segment exceeds the threshold of

$$\lambda_h = \tau(w * \sigma_h + (1 - w) * \sigma_c), \tag{2.1}$$

where σ_h is an adaptive standard deviation calculated from a history window and σ_c is a constant which stops the threshold converging to zero. Note that the history window only covers the previous segments not deemed to be a gross-body movement related. A weight parameter $w \in [0, 1]$ can be selected to adjust the threshold adaptation level and τ is a scaling factor to alter the sensitivity.

The parameters were set experimentally to $f_b = [5 - 25Hz]$, critical level $d_0 = 0.1$, $\tau = 3$, $w = 0.4$, $\sigma_c = 35$ and $\lambda_l = 0.5$. The history window length was set to 60 seconds. An example of detected gross body movement segments is given in Figure 2.6.

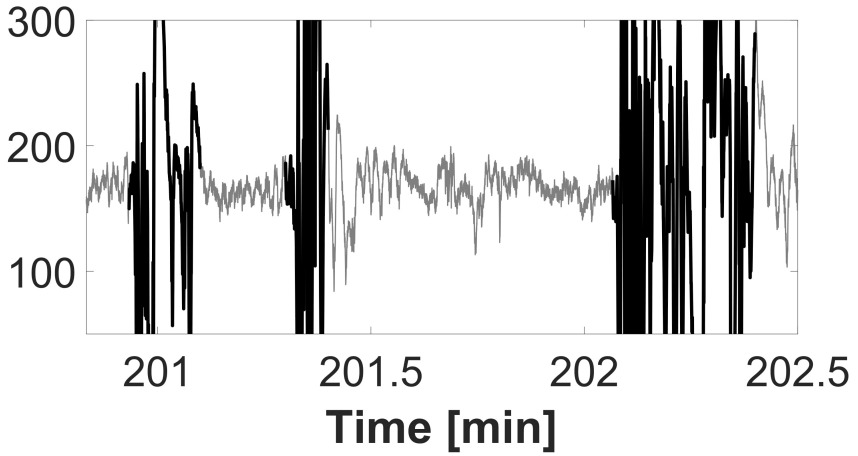
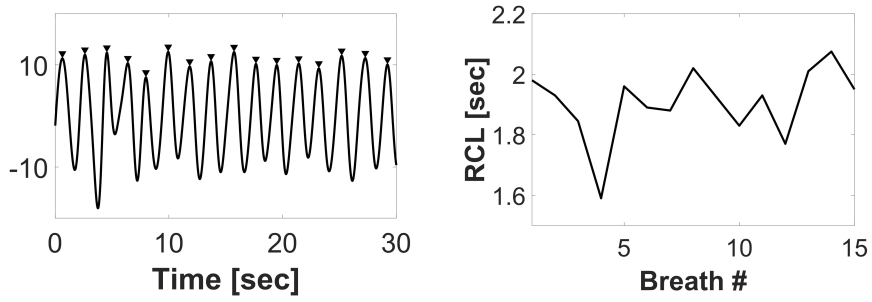


Figure 2.6: Detected gross body movement artefacts. Artefact free signal (gray) and gross body movements (black).

2.2.4 Identification of respiration cycle

RCL describes the evolution of breath duration as a function of time. Breath duration (expiration + inspiration) is simply the time between consecutive peaks in the artefact free segments of the LF signal. All the local peaks $P = \{p_1, ..., p_m\}$ can be found where the sign of the first derivative changes from positive to negative. Although this method finds every local maxima even those that are caused by noise. By applying restrictions based on infant respiration rates (see Table 1.1) to the peak prominence and inter peak distances the peak set can be reduced to cover only the meaningful ones.

From the peak series the RCL series is simply $RCL = \{p_2 - p_1, ..., p_{i+1} - p_i, ..., p_n - p_{n-1}\}$. In Figure 2.7 is illustrated an example of LF signal peaks and RCL series during NREM3.



(a) LF respiration signal **(b)** Respiration cycle length series

Figure 2.7: Respiration cycle length (RCL) series determination. **(a)** Low band pass filtered respiration signal where the peaks are detected (triangles). **(b)** Calculating the difference between the consecutive peaks yields RCL series.

2.2.5 Determination of heart beat intervals

Analysis of HBI series is the same as in respiration RCL series. In respiration, the inspiration - expiration transitions were the points of interest and after finding them the RCL series was calculated by determining the time difference between consecutive points. Now the points of interest are the heart beats and calculating the difference between them yields the HBI.

In this work the heart beats were detected from ECG using algorithm proposed by Pan J. and Tompkins W.J. [110]. It is divided into following steps

1. ECG signal is band pass filtered between 5-15 Hz to remove low trends and high frequency noise.
2. Differentiator filter is applied to enhance the QRS-complex.
3. Squaring the signal to make it positive and to enhance the R-peak.
4. Integrator filter to merge the QRS-complex.
5. Peak detection and locating true heart beats by applying adaptive thresholds and a set of rules to ignore noise peaks.

An example of ECG signal and HBI series is illustrated in Figure 2.8

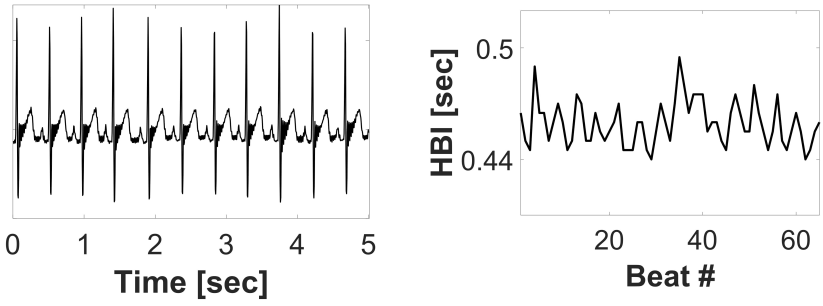
2.3 Feature extraction

Features were extracted from epochs,that did not contain anomalies, flatlines or infant absences. The purpose of feature extraction is to try to embed the time series into space where the sleep classes could be separated.

Several features were calculated from HF and LF frequency bands of the mattress signal and also from RCL and HBI series. Also the presence of gross body movements affected on feature calculation. Here the features are divided into four distinct categories; waveform, spectral, non-linear dynamics and miscellaneous, which are covered in Sections 2.3.1- 2.3.4. All the features calculated in this work are shown in Table 2.1.

Table 2.1: Collection of all the features and signals where they were calculated. LF - low and HF - high frequency band signals, where F - free and P - presence of gross body movements indicate segment types. RCL - respiration cycle length series and HBI - heart beat interval series. Oth. - other source than previos ones. Features 2 and 27 were calculated from the raw signal, whereas 23, 40 and 41 high pass filtered signal. Ref. - references to previous usage in sleep literature. Abb. - abbreviation. N. dynamics - nonlinear dynamics.

			LF		HF					
	Name	Abb.	F	P	F	P	RCL	HBI	Oth.	Ref.
Waveform	1 Mean	M					x	x		[111]
	2 Median	MED							x	
	3 Variance	VAR	x	x	x	x				[111, 112]
	4 Coefficient of variation	COFV					x			
	5 Standard deviation	STD					x	x		[112]
	6 Skewness	SKEW	x	x	x	x				
	7 Kurtosis	KURT	x	x	x	x				
	8 Hjorth mobility	MOB	x	x	x	x				[112]
	9 Hjorth complexity	COMP	x	x	x	x				[112]
	10 Peak median	PM	x							[111]
	11 Peak standard deviation	PSTD	x							[111]
	12 Trough median	TM	x							[111]
	13 Trough standard deviation	TSTD	x							[111]
	14 Mean of differences of local extrema	DMA	x		x					[112]
	15 Variance of differences of local extrema	DVA	x		x					[112]
	16 Mean of time difference of local extrema	DMT	x		x					[112]
	17 Variance of time difference of local extrema	DVT	x		x					[112]
	18 Zero crossings	ZRCA	x	x	x	x				[112]
	19 Zero crossings of first derivative	ZRCB	x	x	x	x				[112]
	20 Zero crossings of second derivative	ZRCC	x	x	x	x				[112]
	21 Envelope standard deviation	ESTD	x	x	x	x				
	22 Root mean square envelope crossings	RMSEC			x					
	23 Power	POW							x	
Spectral	24 First spectral moment	M1	x	x	x	x				
	25 Second spectral moment	M2	x	x	x	x				
	26 Spectral entropy	SE	x	x	x	x				
	27 Spectral energy in 3-6Hz	H3Z6							x	[112]
	28 Low and high freq. ratio	LF_h/HF_h						x		[113]
	29 Low and very low freq. ratio	LF_h/VLF_h						x		[113]
	30 High and very low freq. ratio	HF_h/VLF_h						x		[113]
	31 Modulus of high freq.	$MODHF_h$						x		[113]
	32 Phase of high freq.	PHF_h						x		[113]
N. dynamics	33 Approximate entropy	ApE	x		x					
	34 Sample entropy	SampE	x		x					
	35 Fuzzy entropy	FuzE	x		x					
	36 Permutation entropy	PermE	x		x					
	37 Hurst exponent	HE	x		x					
	38 Lyapunov exponent	LE	x		x					
Misc.	39 Increases respiratory resistance instances	IRRI			x					[114]
	40 Number of spectral change points	CPF							x	
	41 Number of movement initiations	CPM							x	
	42 Age	AGE							x	
	43 Total segment duration	TD	x							



(a) ECG **(b)** HBI series

Figure 2.8: Starting point of heart rate variability features. **(a)** Example of ECG signal where R-peaks have the highest amplitude. **(b)** Heart beat interval series from 30 sec epoch derived from the R-peak intervals. Sinus arrhythmia is clearly detectable causing the main oscillatory component.

2.3.1 Waveform

Waveform features contain statistical and generic waveform descriptors. These include statistical moments that are calculated for signals themselves or their envelopes or local extrema. Hjorth parameters and zero crossing values are also included.

Basic statistical moments, namely mean, variance, skewness and kurtosis, in addition to median, standard deviation and coefficient of variation were calculated for the signals themselves (M, MED, VAR, SKEW, KURT, COEFV, STD), signal envelopes (ESTD) or signal's local extrema (PM, PSTD, TM, TSD, DMA, DVA,DMT, DVT)[111, 112].

The mth central moment of discrete X with distribution $P(X)$ is defined as

$$\mu_m = \mathrm{E}[(X-a)^m] = \int_{\infty}^{\infty} (x-a)^m P(x)\mathrm{d}x, \tag{2.2}$$

where a is the first moment. With first moment the point is set to a=0 and for higher moments $a = \mu_1$. Here sample approximations were used for the moments. Let now $X = \{x_1, x_2, ..., x_T\}$ denote discrete series of length T. The first moments gives a characteristic typical value of the data set and it is simply calculated as a sample mean in this work

$$M = \hat{\mu}_1 = \frac{\sum_{i=1}^{T} x_i}{T}. \tag{2.3}$$

Mean calculated from the RCL and HBI series, M_{RCL} and M_{HBI}, are substitutes for respiration and heart rates (inverse of rates).

Mean is replaced with median generally in cases when heavy outliers are expected. Median is the middle observation of the ordered data set and it describes better a typical value in these cases. Here, the median calculated from the raw signal (MED) is used as descriptor of the total constant weight applied to the mattress (overlaying foam mattress + baby weight).

The second moment, variance, can be calculated as

$$VAR = \hat{\mu}_2 = \frac{\sum_{i=1}^{T}(x_i - \hat{\mu}_1)^2}{T-1}. \tag{2.4}$$

Standard deviation is simply the square root of variance. The coefficient of variation is COFV=$\frac{\sqrt{\hat{\mu}_2}}{\hat{\mu}_1}$. The signal variance is related to power and can be thought to describe power that is normalized by the length. Also the direct signal power was calculated as $POW(X) = \sum_i x_i^2$. Although the variance was calculated from LF and HF bands of both segments the power was calculated from the high band pass signal ($> 0.4Hz$).

The skewness is the third moment and it describes the symmetry of the data distribution. If skewness deviates from zero there is a difference between mean and median. In this work an unbiased estimate is calculated with the following formula;

$$SKEW = \frac{\sqrt{T(T-1)}}{T-2} \times \frac{\frac{1}{T}\sum_{i=1}^{T}(x_i - \hat{\mu}_1)^3}{\left(\frac{1}{T}\sum_{i=1}^{T}(x_i - \hat{\mu}_1)^2\right)^3}. \tag{2.5}$$

As in the all other sample moments above, bias-corrected formulation for fourth moment (kurtosis) was used;

$$KURT = \frac{T-1}{(T-2)(T-3)}\left((T+1)\frac{\frac{1}{T}\sum_{i=1}^{T}(x_i - \hat{\mu}_1)^4}{\left(\frac{1}{T}\sum_{i=1}^{T}(x_i - \hat{\mu}_1)^2\right)^2} - 3(T-1) + 3\right). \tag{2.6}$$

Kurtosis describes how heavily tailed the data distribution is, meaning how likely it is to observe extreme values.

Hjorth proposed three parameters to describe EEG signal: Activity, Mobility and Complexity [115]. The activity is same as the signal's variance. To calculate the mobility and complexity the first and second derivative of the signal is needed, $X' = \{x_{i+1} - x_i\}_{i=2}^{T}$ and $X'' = \{x_{i+1} - 2x_i + x_{i-1}\}_{i=2}^{T-1}$, respectively. The mobility is defined as

$$MOB = \sqrt{\frac{VAR_{X'}}{VAR_X}}, \tag{2.7}$$

where $VAR_{X'}$ is the variance of the first derivative and VAR_X of the original time series. The complexity is

$$COMP = \sqrt{\frac{VAR_{X''}VAR_X}{VAR_{X'}^2}}. \tag{2.8}$$

It is required to determine the local extremes, namely peaks and troughs, for several waveform features (features 10-17 in Table 2.1). The peak detection method discussed earlier in Section 2.2.4 with RCL series is directly applicable. For detecting troughs T the same procedure is applied to the negated signal.

Ideally for the LF band, signal peak values should indicate the transitions points from inspiration to expiration. Peak and trough medians and standard deviation (PM,TM,PSTD,TSTD) in addition to the mean and variance of differences of consecutive local extrema (DMA, DVA) describe basic breath amplitude characteristics. To extend the analysis from the amplitude to the time dimension the mean difference and variance of time difference of local extrema (DMT,DVT) were calculated. DMA, DVA, DMT and DVT were determined also for HF band.

The number of zero crossing indicate how many times the time series (ZRCA) or its first and second derivatives (ZRCB,ZRCC) cross the zero line [112].

The envelope of time series can be calculated using Hilbert transformation, splines (piecewise polynomials) to fit a curve to peaks and troughs or calculating the root mean square (RMS) value on a moving window. Each of these yield different envelope. Envelope standard deviation (ESTD) was determined on splines envelope. Root mean square envelope crossings (RMSEC) feature describes how many times the signal crosses its RMS envelope.

2.3.2 Spectral

These features analyse the frequency content of the signal by transforming the time series into the spectral domain (Table 2.1, features 23-31).

The spectrum of a signal can be characterized with a power spectral density (PSD), which describes how much power each frequency contains. In this work the PSD is estimated with the simplest method, the periodogram, which is simply the modulus -squared of the discrete Fourier transform (DFT), $Y(f_i)$[116]

$$PSD(f_i) = \frac{1}{N}|Y(f_i)|^2. \tag{2.9}$$

N is the sample size. Here, the DFT is calculated using the Fast Fourier Transform algorithm. PSD needs to be further normalized when it is used to estimate the first and second spectral moments accompanied with the spectral entropy. The normalization

$$PSD_n(f_i) = \frac{PSD(f_i)}{\sum_{f_i} PSD(f_i)} \tag{2.10}$$

sets the PSD to resemble a probability distribution. Now spectral moments can be derived from Equation 2.2

$$M1 = \sum_{f_i} f_i PSD_n(f_i) \quad \& \quad M2 = \sum_{f_i} (f_i - M1)^2 PSD_n(f_i). \tag{2.11}$$

The spectral entropy can be derived applying Shannon's entropy

$$SE = -\sum_{f_i} PSD_n(f_i) \log_2(PSD_n(f_i)). \tag{2.12}$$

The change in 3-6 Hz power has been assumed to reflect the change in cardiac activity [112]. This is calculated as

$$H3Z6 = 10 \log \sum_{f_i=3Hz}^{6Hz} PSD(f_i). \tag{2.13}$$

HBI series has been a great interest in sleep literature and spectral features derived from it have been applied to adult sleep classification [113, 117–119]. Here the HBI spectrum was estimated using Lomb-Scargle periodogram because the HBI locations vary in time resulting in unevenly spaced time series. The main principle of the method is to find the best least squares fit for sine waves to data [120]. Illustration of the HBI spectrum calculated on Figure 2.8b series is shown in Figure 2.9.

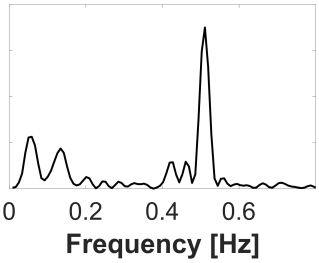


Figure 2.9: Lomb-Scargle periodogram of heart beat interval series from Figure 2.8b. Note that the x-axis in the HBI figure was replaced by time locations for intervals when calculating the periodogram.

Different frequency band powers were calculated by integrating the periodogram as in [113, 117–119]. The frequency range for total power (TP_h) was in $0.003-0.5$ Hz. For very low (VLF_h), low (LF_h) and high frequency (HF_h) the ranges were 0.003-0.04 Hz, 0.04-0.15 Hz and 0.15-0.5 Hz, respectively. VLF_h, LF_h and HF_h were normalized by TP_h and log-transformed. Then the following ratios were determined LF_h/HF_h, LF_h/VLF_h and HF_h/VLF_h. Also modulus and phase of HF_h pole were calculated.

2.3.3 Non-linear dynamics

Non-linear dynamics describe the chaotic qualities of a system. Several entropies calculated here work as regularity measures, the Hurst exponent as a descriptor for fractality and the maximum Lyapunov exponent as a chaoticity measure (Table 2.1 features 31- 37).

The Approximate entropy measure (ApE) was developed by Pincus [121]. The basic idea is that the sequence is regular if its subsequences and their expansions are similar. Let $X = \{x_1, x_2, ..., x_T\}$ again be a time series. It is divided into overlapping subsequences $\{X_1^m, X_2^m, ..., X_{T-m+1}^m\}$ of length m, where $X_i^m = \{x_i, ..., x_{i+m-1}\}$. For each subsequence i the proportion of close proximity subsequences are determined

$$C_i^m = \frac{\#\{\text{pairs } (i,j) \mid d(X_i^m, X_j^m) < r \ \ \& \ \ j \in \{1, ..., T-m+1\}\}}{T-m+1}, \quad (2.14)$$

where $d(X_i^m, X_j^m)$ is the maximum Chebyshev distance (max distance between elements) and r is a distance threshold. Next define

$$\phi^m = \frac{1}{T-m+1} \sum_{i=1}^{T-m+1} \ln C_i^m. \quad (2.15)$$

Then the same procedure is repeated for subsequences of length $m+1$. The approximate entropy is defined as

$$ApE = \phi^m - \phi^{m+1}. \quad (2.16)$$

Approximate entropy is biased towards regularity because it calculates self-mathces and also it calculates different amount of matches between two subsequence lengths m and $m+1$. An alternative is the Sample entropy where these

problems are fixed [122]. Thus the restriction $j \neq i$ is the imposed in nominator of Equation 2.14 when calculating the number of close subsequences. Also the dominator is set to $T-m-1$ for both m and $m+1$. The logarithm is removed from Equation 2.15 and it is updated to

$$\phi^m = \frac{1}{T-m} \sum_{i=1}^{T-m} C_i^{m^*}, \tag{2.17}$$

where $m^* = m$ or $m^* = m+1$ depending on the subsequence length. Sample entropy is calculated as

$$SampE = \ln \phi^m - \ln \phi^{m+1}. \tag{2.18}$$

Yet another alternative for the entropy measure was introduced by Chen *et. al.*, namely Fuzzy entropy[123]. The previous entropies are sensitive to the parameter r because they use the Heaviside membership function, which is replaced in Fuzzy entropy. This method uses subsequences where the subsequence mean $\hat{\mu}_1^{(i)}$ is substracted from the original sequence; $\bar{X}_i^m = \{x_i - \hat{mu}_1^{(i)}, ..., x_{i+m-1} - \mu_1^{(i)}\}$. A membership matrix is calculated from subsequences as

$$D_{i,j}^m = u(d(\bar{X}_i^m, \bar{X}_j^m), n, r), \tag{2.19}$$

where $u(\cdot)$ is a fuzzy membership function. Practically any membership function where the continuity and convexity hold can be applied [124]. In this work the exponential function proposed in the original article [123] was used

$$u(d(\bar{X}_i^m, \bar{X}_j^m), n, r) = \exp\{-(d(\bar{X}_i^m, \bar{X}_j^m)/r)^n\}, \tag{2.20}$$

where $d(\cdot)$ is the maximum Chebyshev distance as in Approximate and Sample entropy. It has been recommended to set the weight parameter n as a small integer [124] and thus here n=2.

$$\phi^m = \frac{1}{T-m} \sum_{i=1}^{T-m} \sum_{j=1, j\neq i}^{T-m} \frac{D_{i,j}^{m^*}}{T-m-1}. \tag{2.21}$$

Fuzzy entropy is now defined similarly as Sample entropy.

In permutation entropy the approach is bit different compared to the previous three measures [125], although the starting point is quite similar. The time series is divided into subsequences of m items. Then the values within each subsequence are ordered into ascending (or descending) order and representing marker Ω is assigned to a subsequence based on its permutation. E.g. let the subsequences be $X_1 = \{1,2,3\}$ and $X_2 = \{6,5,4\}$. If they are ordered the former can be assigned with a marker $\Omega(1) = 012 = \pi_1$ and the latter with $\Omega(2) = 210 = \pi_6$. Note that in this simple example there are total $m! = 3! = 6$ possible permutations and thus $\Omega \in \{\pi_1, ..., \pi_6\}$. The relative frequency of each permutation is calculated as

$$p(\pi_k) = \frac{\#\{X_i | i \in \{1, .., T-m+1\} \ \& \ \Omega(i) = \pi_k\}}{T-m+1} \quad \forall k \in \{1, ..., m!\} \tag{2.22}$$

and then the entropy is defined as

$$PermE(m) = -\sum_{T=1}^{m!} p(\pi_T) \log p(\pi_T). \tag{2.23}$$

The Hurst exponent describes fractality and long-range power law correlations within a time series and it is calculated here using rescaled range analysis [126]. Hurst exponent H describes the power law relation between the rescaled range $(R(n)/S(n))$ and the length of time series n

$$\mathrm{E}\left[\frac{R(n)}{S(n)}\right] \propto n^H \text{ as } n \to \infty, \tag{2.24}$$

where $R(n)$ is the range and $S(n)$ is the standard deviation. H can be estimated by calculating rescaled range for logarithmically decreasing time series lengths $n \in \{T, T/2, T/4, ...\}$, where T is the length of original time series $X = \{x_1, ..., x_T\}$. For each length $n = T/m$ the original time series partitioned into a set of subsequences $X^n = \{X_1^n, ..., X_m^n\} = \{\{x_1, ..., x_n\}, ..., \{x_{T-n}, ..., x_T\}\}$ and the cumulative series is determined for each zero mean normalized subsequence $j \in \{1, .., m\}$

$$Z_{j,c}^n = \sum_{i=n*(j-1)+1}^{c+n*(j-1)} \{x_i - \hat{\mu}_1^{(j)}\} \quad \forall c \in \{1, ..., n\}. \tag{2.25}$$

The range is calculated from the cumulative series for each subsequence $R_j(n) = \max_c Z_{j,c}^n - \min_c Z_{j,c}^n$ as well as the corresponding subsequence standard deviation $S_j(n)$ directly from X_j^n. The rescaled range estimate for n is just average over all subsequences. The Hurst exponent is the slope of log-log plot of $R(n)/S(n)$ and n.

The Lyapunov exponent λ describes the unpredictability and chaoticity of the system. It is the rate of separation (or convergence) of neighboring trajectories x and $x + \delta$ in phase space and can be formulated as how a small initial perturbation δ_0 increases (decreases) as a function of time [127]

$$||\delta(t)|| \sim ||\delta_o|| e^{\lambda t}. \tag{2.26}$$

In a multidimensional system there are Lyapunov exponents for each dimension and the maximum of them is typically of interest, because it is sufficient to determine the chaoticity of system. Note that $\lambda > 0$ indicates chaoticity because then the system is extremely sensitive to small perturbations in initial values (exponential growth of perturbation). With $\lambda < 0$ the system is highly predictable because the perturbation diminishes quickly and trajectory of system is stable.

Although the Lyapunov exponent is intuitive and well defined concept for an explicitly formulated system (differential equation system), there are ways to define it for time series. Phase space can be reconstructed from one dimensional time series with time delay coordinate embedding [128, 129]. The method for calculating the largest Lyapunov exponent proposed by Rosenstein M. T. et. al. was applied in this work [130].

1. m-dimensional phase space is constructed via time delay embedding from time series $\{x_1, ..., x_N\}$. $\mathbf{X} = [\mathbf{X}_1, ..., \mathbf{X}_M]^T$, where $\mathbf{X}_i j = [x_j, x_{j+\tau}, x_{j+2\tau}, ..., x_{j+(m-1)\tau}]$ is embedded phase-space vector with lag τ and $M = N - (m-1)\tau$. The lag parameter τ was chosen visually from autocorrelation function [131] and the embedding dimension m via false nearest neighbor method [132].

2. The evolution of perturbation is estimated for each X_j by finding the nearest neighbor $X_{\hat{j}}$ and determining the distance

$$d_j(0) = \min_{\mathbf{X}_{\hat{j}}} ||\mathbf{X}_j - \mathbf{X}_{\hat{j}}||. \tag{2.27}$$

Temporal distance need to be longer than the mean period to avoid temporally consecutive vectors.

3. Next the distance between the trajectories are tracked for $i = 1, ..., p$ time steps for each vector $\mathbf{X_j} \mid j < M - p$

$$d_j(i) = ||\mathbf{X}_{j+i} - \mathbf{X}_{\hat{j}+i}|| \tag{2.28}$$

and the average of logarithm is taken $\ln \hat{d}(i) = \langle \ln d(i)_j \rangle_{j \in \{1, M-p\}}$

4. The Lyapunov exponent is estimated by finding the slope of the least squares fit to $\ln \hat{d} = \lambda \cdot t$, where $t = \{1 \cdot \Delta t, ..., p \cdot \Delta t\}$.

2.3.4 Miscellaneous

Few extra features were also added. These include the increased respiratory events (IRRI), signal change points (CPF,CPM) in addition to the patient age and duration of artefact free signal (DT).

Increased respiratory resistance causes $\sim$ 0.4 sec duration high amplitude spiking in 6-16 Hz band. To detect spiking the algorithm proposed by Alametsä al. was implemented [114]. The outline of the algorithm is the following;

1. Fast Fourier Transform is applied to moving Hanning windows of 0.5 sec with step size of 0.2 sec on 6-16 Hz band to derive the amplitude spectrum $S_t[f_{\text{band}}]$. Before the FFT the window is zero padded to 512 points. Note that with 200 Hz sampling frequency 0.5 sec window contains 100 points.

2. Mean value from 6-16 Hz band is calculated for each window; $\bar{S}_t[f_{\text{band}}]$.

3. The windows are centered at $c \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ sec. From these windows the maximum value for each second k is calculated; $A[k] = \max_c \bar{S}_t[f_{\text{band}}]$, where $t = k + c$.

4. Local base line is determined by taking median form 60 sec history window $B[k] = \text{median}(A[k - 60 : k - 1])$.

5. The change relative to base line is calculated as $A_c[k] = \frac{A[k]-B[k]}{B[k]} \times 100\%$.

6. Step is to compare this to a threshold λ to determine whether or not a spiking event take place during that particular second;

$$\text{spiking at } k \;\mid\; A_c[k] > \lambda. \tag{2.29}$$

In this work the threshold was set to $\lambda = 25\%$ as in the original article. For each 30 sec epoch the number of spikes detected was used as a feature in upcoming classification.

The cross body movement detection algorithm introduced in Section 2.2.3 finds the change points when the signal spectrum changes significantly. The

number of these change points in an epoch was used as a feature. Note in the later steps the amplitude based threshold is applied to merge neighboring segments together. After this step another value for change points can be calculated, which indicates how many separate movement session are within an epoch. The total duration of artefact-free signal indicates how still the baby stays during the epoch.

2.3.5 Postprocessing of features

To improve the performance of the classifier, post-processing steps were added to handle the missing values, impact of different scales and imbalance between classes. Namely four steps were included; 1) log-modulus scaling, 2) Z-score normalization 3) missing value handling and 4) resampling of minority classes to obtain balance between sleep stages.

Because some feature distributions might span a wide power range it might be necessary to perform logarithmic scaling. Logarithmic scaling is useful when the features hold positive values ($x > 0$) but fails when they are negative because logarithm is undefined for them. To circumvent the problem log-modulus transformation [133]

$$x_{ij} := \text{sign}(x_{ij}) \cdot \log(|x_{ij}| + 1). \tag{2.30}$$

was applied in feature dimensions j, whose 90 percentile range spanned over 4 powers.

In general different features span different range of values, which need to be standardized to ensure that each feature could have equal impact when teaching the machine classifier. Otherwise those features with minuscule range tend to have zero impact compared to features with wider span. Z-score transformation is a typical normalization method, where the feature distribution mean is set to zero and the standard deviation to one

$$x_{ij} = \frac{x_{ij} - \hat{\mu}_{1j}}{\sqrt{\hat{\mu}_{2j}}}, \tag{2.31}$$

where $\hat{\mu}_{1j}$ and $\sqrt{\hat{\mu}_{2j}}$ are the sample mean and standard deviation of jth feature, respectively.

Several features were derived for both segments of LF and HF signals. For example if an epoch did not contain any movements there were no corresponding artefact segments, which in turn led to undefined values for segment features. Missing values were replaced by feature means (0 after Z-score transformation).

There were approximately twice as many awake epochs compared to other sleep stages (Figure 3.7a). Because there is significant imbalance between the number of epochs in each sleep stage, this might lead to an improper classifier that chooses the majority class no matter what the feature values an epoch has. This is due the fact that the SVM binary classifier handles each class symmetrically on behalf of the penalties applied during training phase. There are two major approaches to tackle the issue. One is to apply class-specific weighting. Then the minority class misclassification during training phase is more severely penalized. Another approach is to apply down or up sampling schema to balance the class sizes. Here a synthetic minority over-sampling technique called Adaptive Synthetic sampling (ADASYN) was applied [134].

The basic idea of ADASYN is to detect the region of hyperspace where there is significant overlap between the two classes and resample new minority instances on that neighborhood. The following steps are in ADASYN

1. Estimation of local density function for minority class instances as $r_i = \delta_i/K$, where $i = 1, ..., m_s$ (m_s number of minority samples), δ_i number of majority class instances among K nearest neighbors of ith minority class instance.
2. Normalization into density function $\hat{r}_i = r_i / \sum_{i=1}^{m_s} r_i$
3. Number of synthetic samples that need to be generated for each minority sample $g_i = \hat{r}_i \times (m_l - m_s)$, where m_l is the size of majority class.
4. Generate for each minority sample x_i g_i new instances by randomly choosing minority neighbor x_{zi} each time and generating sample as $s_i = x_i + (x_{zi} - x_i) \times \lambda$, where sample randomly $\lambda \in [0, 1]$.

In NREM3 vs. all other stages classification NREM3 class were minority sampled. In other two classification schema all the other classes were samples against awake stage.

2.4 Support vector machine

After describing epochs with a set of features, the next step is to optimally separate them into feature classes that correspond to sleep state. There are several different methods to choose the separating hyperplane. Support Vector Machines attempt to find the optimal hyperplane by maximizing the margin between two classes [135]. SVM has been a popular machine learning approach as it 1) directly solves the problem without taking any complex route (e.g. without estimating probability functions), 2) incorporates kernel functions to include more complex decision surfaces and 3) uses a convex optimization function (global optimum) [136].

The goal of this work was to perform one binary and two multiclass classifications. SVM was originally formulated solely for binary classification, but it can be extended to multiclass case by directly reformulating the basic optimization problem [137] or alternatively decomposing the multiclass classification into a set of two-class problems using error-coding output codes [138]. The latter approach was applied in this work.

2.4.1 Margin optimization

Learning the SVM to distinguish two classes is a margin maximization problem where the goal is to find a linear discriminant between classes. The following derivation follows closely the one that is represented in the book by Alpaydin [136].

Let $\mathbf{x}_i = [x_{i1}, x_{i2}, ..., x_{iD}]^T$ denote D dimensional feature vector of the ith epoch. Similarly let $y_i \in \{-1, 1\}$ indicate the class. The linear discriminant can be fitted directly in feature space or additional non-linear embedding $\mathbf{x} \mapsto \varrho(\mathbf{x}_i)$ can be applied and the discrimination can be done in the new space. Without any loss of generality the formulation here is derived using basis function $\varrho(\cdot)$

notation (simply $\varrho(\mathbf{x}) = [1 \quad \mathbf{x}^T]^T$ for original feature space). The goal is to find parameters $\mathbf{w} = [w_1, w_2, ..., w_P]$ for discriminant function $g(\mathbf{x}_i) = \mathbf{w}\varrho(\mathbf{x}_i)$ so that

$$y_i \mathbf{w} \varrho(\mathbf{x}_i) \geq 1, \tag{2.32}$$

which just tells that classes are on different side of the margin. For optimal margin the distance γ from sample to discriminant should to be at least

$$\frac{y_i \mathbf{w} \varrho(\mathbf{x_i})}{||\mathbf{w}||} \geq \gamma \quad \forall i = 1, .., n, \tag{2.33}$$

where n is the number of samples. For distance γ there does not exist unique solution and by enforcing $\gamma||\mathbf{w}|| = 1$ the problem can be reformulated as minimization task

$$\begin{aligned} \min \quad & \frac{1}{2}||\mathbf{w}||^2 && (2.34) \\ Subject \ \ to \quad & 1 - y_i \mathbf{w} \varrho(\mathbf{x}_i) \leq 0 \quad \forall i = 1, .., n && (2.35) \end{aligned}$$

In practice the features and their annotation contains noise so that the samples are not ideally separable and thus there are no feasible solution for the problem. The idea of soft margin SVM is to introduce slack variables ξ to allow some samples to be within a margin or on the wrong side of the hyperplane. The optimization goal is updated to

$$\begin{aligned} \min \quad & \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n} \xi_i && (2.36) \\ Subject \ \ to \quad & 1 - y_i \mathbf{w} \varrho(\mathbf{x}_i) \leq \xi_i \quad \forall i = 1, .., n \\ & \xi \geq 0, \end{aligned}$$

where $C > 0$ is a box constraint that balances between maximizing the margin and amount of slack needed. The constraints can be included into optimization by reformulating it as a Lagrangian dual problem, where the objective function is first optimized by original parameters. Then the optimal value is plugged back to formulation and the function is again optimized, but this time on Lagrangian penalty parameters. Lagrangian primal is

$$\min \quad L_p = \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n} \xi_i + \sum_{i=1}^{n} \alpha_i (1 - \xi_i - y_i \mathbf{w} \varrho(\mathbf{x}_i)) - \sum_{i=1}^{n} \mu_i \xi_i, \tag{2.37}$$

where α_i is ith sample's penalty associated by violating the first constraint in 2.36 and similarly μ_i for violating the second constaint. By taking the gradients with respect to $\mathbf{w}$ and ξ and setting them to zero (optimal point of convex problem)

$$\nabla_{\mathbf{w}} L_p = \mathbf{w} - \sum_{i}^{n} \alpha_i y_i \varrho(\mathbf{x}_i)^T = 0 \quad \Longrightarrow \quad \mathbf{w} = \sum_{i}^{n} \alpha_i y_i \varrho(\mathbf{x}_i)^T \tag{2.38}$$

$$\frac{\partial L_p}{\partial \xi_i} = C - \alpha_i - \mu_i \quad \Longrightarrow \quad \mu_i = C - \alpha_i \ \ \forall i \in [1, ..., n] \tag{2.39}$$

When these are inserted back to the primal it yields dual

$$\max_{\alpha} L_D = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_k y_i y_j \varrho(\mathbf{x}_i)^T \varrho(\mathbf{x}_j) \tag{2.40}$$

with constraints $\sum_{i=1}^{n} y_i \alpha_i = 0$ and $0 \leq \alpha_i \leq C$. The inner product of embeddings $\varrho(\mathbf{x}_i)^T \varrho(\mathbf{x}_j)$ in the dual can be thought as a distance measure of feature vectors in embedded space. Kernelization means that this is replaced with corresponding distance measure, a kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$, which replaces the necessity of embedding and operates directly on feature vectors. Using this approach it is straight-forward to find a non-linear discriminant in the original feature space without the need for basis function mappings. In this work the radial basis function kernel is applied

$$k(\mathbf{x}_i, \mathbf{x}_k) = \exp\left(\frac{-||\mathbf{x}_i - \mathbf{x}_j||^2}{s^2}\right), \tag{2.41}$$

where s is called kernel scale.

The dual is a quadratic programming problem, which was solved here with the Sequential Minimal Optimization (SMO) algorithm [139]. This yields for correctly classified samples $\alpha_i = 0$ and for support vectors $\alpha_i > 0$. Now Equation 2.38 gives weight parameters $\mathbf{w}$. Support vectors that lay on the margin have $\alpha_i < C$ and $\xi_i = 0$ and for instances that are within the margin or misclassified $\alpha_i = C$. The predicted class for a new unseen instance $\mathbf{x}_{new}$ can be now determined as

$$g(\mathbf{x}_{new}) = \mathbf{w}\varrho(\mathbf{x}_{new}) = \sum_{i=1}^{n} \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}_{new}) \tag{2.42}$$

and class +1 is assigned for $g(\mathbf{x}_{new}) > 0$ and if $g(\mathbf{x}_{new}) < 0$ then class is -1. Because $\alpha_i \neq 0$ only for support vectors it is necessary to store alone these training samples for later classification.

2.4.2 Multiclass SVM

Error correcting output coding is a method to decompose the K-class problem ($K > 2$) into a set of L binary classification problems [138]. Using different coding design matrices the problem can be coded into e.g. one-versus-all or one-versus-one, where in the former $L = K$ classifiers are determined by comparing one class to pooled set of other classes and in the latter $L = K(K-1)/2$ pairwise classifiers are created. In this work the one-versus-one coding scheme was chosen. This can be represented with a coding design matrix and for example in awake - REM - NREM case

$$M = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & -1 \end{bmatrix}, \tag{2.43}$$

where column indicates the binary classifier and row class. For example in the first classifier awake (the first row) is labelled as 1 and REM as -1 whereas NREM3 with a value of 0 is totally ignored. Now the class is determined by

taking average loss-score values $f_a(M, k, g_l(\mathbf{x}_{new}))$ over all binary classifiers for each class and the class that has the minimum average loss-score is chosen

$$\hat{k} = \text{argmin}_k f_a(M, k, g_l(\mathbf{x}_{new})) = \text{argmin}_k \frac{\sum_{l=1}^{L} |m_{kl}| f(m_{kl}, g_l(\mathbf{x}_{new}))}{\sum_{l=1}^{L} |m_{kl}|}, \quad (2.44)$$

where m_{kl} is a coding design matrix element and $f(\cdot)$ is a loss function, which in this work was chosen to be hinge loss

$$f(m_{kl}, g_l(\mathbf{x}_{new})) = \max(0, 1 - m_{kl} g_l(\mathbf{x}_{new})). \quad (2.45)$$

The function $g_l(\mathbf{x}_{new})$ is discriminant for binary classifier l according to Equation 2.42. Further away from the margin the class $g_l(\mathbf{x}_{new})$ greater the loss when it is assigned to m_{lk} with opposite sign whereas hinge loss is zero when assigned to the class with same sign.

2.4.3 Hyperparameter optimization

Machine learning approaches have a tendency to overfit; that is, they can almost perfectly classify training data whereas the performance on new unseen data remains poor. One way to overcome this problem is proper hyperparameter selection.

When choosing the hyperparameters, the goal is to maximize the classifier performance on unseen data. This can be evaluated using using cross-validation. In this approach, data is partitioned into K subsets of testing and training data, where there is no overlap between training and testing datasets. Because the goal was to create a classifier that works well on individuals, leave-one-patient-out (LOPO) cross-validation was applied. In this scheme on our dataset, a classifier is fitted 48 times using each time different set of 47 patients and the performance is calculated on the left-out patient. The average performance over validation rounds gives an estimate of expected performance.

There are two hyperparameters that need to be optimized in SVM, box constraint C and kernel scale s. Naive grid search was implemented where the performance was evaluated using LOPO cross-validation with grid values $[2^{-5}, 2^{-3}, 2^1, 2^2, 2^3, 2^6, 2^8]$ for both hyperparameters.

2.5 Postprocessing

The proposed SVM classifier is blind to the longer trend in the signal as decisions are independently made for each 30s epoch. This does not duplicate the human expert. To incorporate a wider scope to sleep stage prediction, a median filter based method was applied. The median filter was applied to the average loss-score value $f_a(M, k, g_l(\mathbf{x}_{new}))$ and the epoch was assigned to class k with the lowest value (Equation 2.44). To ensure comparability the manual sleep annotation was also filtered using the same median filter.

Chapter 3

Evaluation

Two different performance measures, Area Under Curve (AUC) and accuracy (ACC) were chosen in this work to evaluate the classifier performances, in addition to the confusion matrix (CM), which provides a way to identify which classes are easily mixed together, if classification problems arise. Here the CM matrix was calculated by pooling the epochs of test subjects together over all cross-validation rounds.

AUC is derived from receiving operating characteristic (ROC). The ROC is defined for a binary classifier by plotting the true positive rate (TPR) as a function of false positive rate (FPR) when the classifier decision threshold is varied. An optimal classifier can separate the classes perfectly, that is $TPR = 1$, $FPR = 0$ and $AUC = 1$. For a random classifier $AUC = 0.5$. In this work the AUC value in multiclass case was defined by taking an average over class AUCs. The accuracy is defined as

$$ACC = \frac{\#\text{ correctly classified instances}}{\#\text{ all instances}}. \tag{3.1}$$

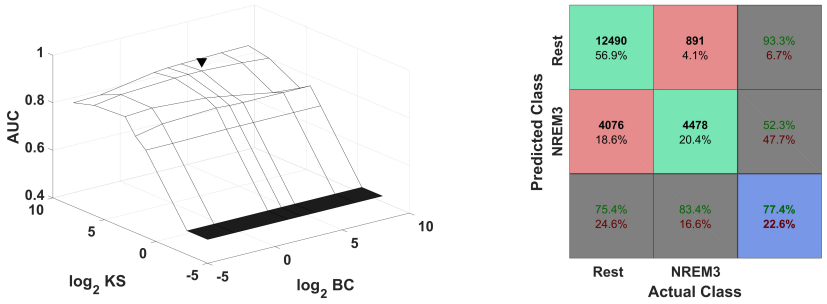
In this chapter the leave-one-patient-out cross validation (LOPO-CV) results for each of the proposed classifiers is evaluated; NREM3 vs rest, awake vs NREM vs REM and awake vs REM vs NREM1/2 vs NREM3. The postprocessing step is then applied for the NREM3 vs rest classifier to see whether or not it increases performance.

3.1 NREM3 - Rest

The LOPO-CV with grid search yielded an optimal box constraint with a value of $BC = 8$ and kernel scale $KS = 64$. This is visualized in Figure 3.1a. The median AUC over all 48 LOPO-CV rounds has a value of $AUC = 0.920$ and $IQR = [0.852,\ 0.942]$. Similarly the median accuracy has a value of $ACC = 0.780$ and with interquartile range of $IQR = [0.726,\ 0.852]$. Table 3.1 provides full summary including the class specific median accuracies. See Figure 3.1b for the pooled confusion matrix. The AUC and ACC values for each patient is shown Figure 3.2.

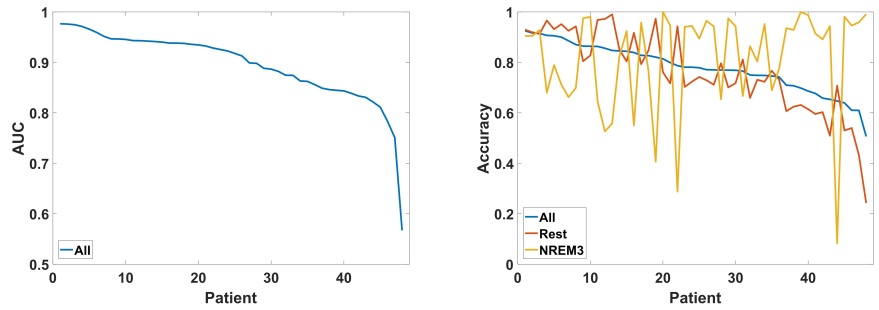
Table 3.1: Median (inter-quartile range) performances in NREM3 - detection.

	Total	NREM3	Rest
AUC	0.920 [0.852,0.943]	-	-
ACC	0.780 [0.726,0.852]	0.910 [0.694,0.949]	0.752 [0.680,0.918]



(a) Grid search for hyperparameters **(b)** Confusion matrix.

Figure 3.1: Hyperparameter optimization and pooled confusion matrix. **(a)** Hyperparameters, Kernel scale (KS) and Box constraint (BC), were chosen via leave-one-patient-out cross validation (LOPO-CV) applied to a grid of points. KS and BC with the highest area under curve (AUC) value, indicated as triangle, were chosen. **(b)** Pooled confusion matrix over 48 different LOPO-CV rounds. Class specific accuracies and predictive values can be seen in gray colored rows and columns, respectively.



(a) Patient AUC **(b)** Patient accuracy

Figure 3.2: Patient AUC values and accuracies in descending performance order. **(a)** AUC values from the best to worst and **(b)** similarly for accuracies.

It is clearly visible from the confusion matrix 3.1b that the classifier is biased towards the NREM3 class, because the number of false positives is quite high. As all the epochs are pooled together from different individuals in the CM, those that have very low performance with a longer recording session have a larger contribution to the result. From Figure 3.2b can be deduced that the performance has a high variability between patients. For some patients the NREM3 detection rate drops below 0.60. Also for the last infant, the amount of

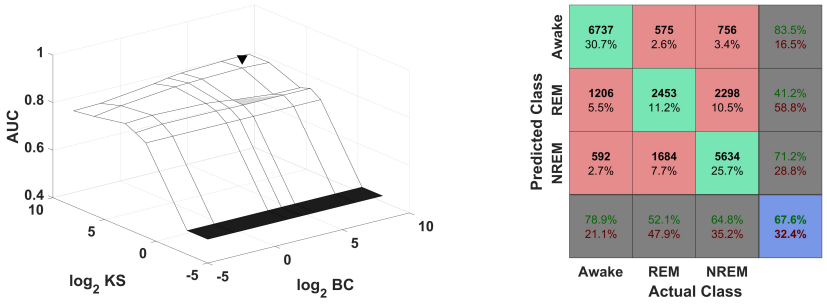
NREM3 is clearly overestimated because it has very low accuracy for detecting other stages. Although overall the performance is reasonable.

3.2 Awake - NREM - REM

For the 3-class case the grid search yielded the hyperparameter values BC=KS=64, which in shown in Figure 3.3 along side with the confusion matrix. Median accuracy dropped to value of $ACC = 0.680$ with $IQR = [0.611,\ 0.722]$ and $AUC = 0.873$ with $[0.831,\ 0.896]$. The more detailed breakdown of the class performances is given in Table 3.2 and the variability between patient performances can be noticed from Figure 3.4.

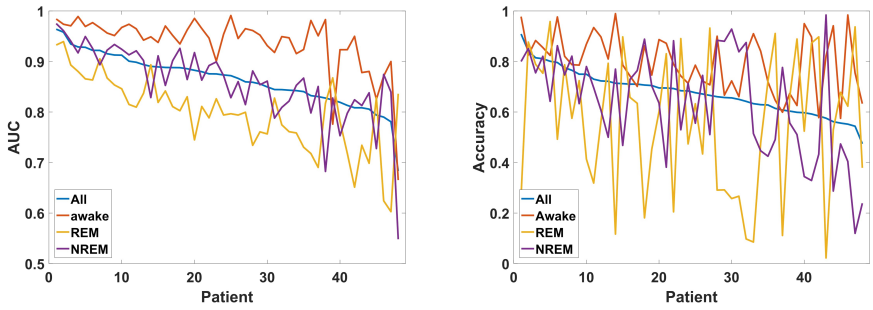
Table 3.2: Awake - REM - NREM median (inter-quartile range) performances.

	Total	Awake	REM	NREM
AUC	0.873 [0.831,0.896]	0.951 [0.924,0.968]	0.810 [0.758,0.844]	0.864 [0.826,0.915]
ACC	0.680 [0.611,0.723]	0.799 [0.717,0.885]	0.588 [0.305,0.791]	0.669 [0.482,0.810]



(a) Grid search for hyperparameters **(b)** Confusion matrix.

Figure 3.3: Awake-REM-NREM hyperparameter optimization and pooled confusion matrix. **(a)** Hyperparameters, KS and BC with the highest AUC value, indicated as triangle, were chosen. **(b)** Confusion matrix of pooled epochs over 48 different LOPO-CV rounds.



(a) Patient AUC **(b)** Patient accuracy

Figure 3.4: Awake vs REM vs NREM patient AUC values and accuracies in descending performance order. **(a)** Patient AUC values from the best to worst and **(b)** similarly for accuracies.

Based on the confusion matrix 3.3b it can be seen that the REM stage is poorly classified. Large periods of awake and NREM are misclassified as REM. Also several REM epochs are thought to be NREM. From the patient accuracies (Fig. 3.4b) it can be deduced that the awake state seems to be relatively easy to distinguish and does not have high variability between subjects.

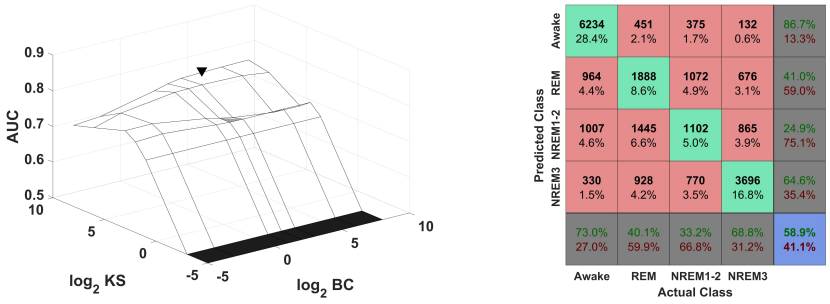
3.3 Awake - REM - NREM1-2 - NREM3

The hyperparameter estimation yielded kernel scale value of KS=64 and the optimal box constraint now is BC=8. Median accuracy is $ACC = 0.584$ with $IQR = [0.538,\ 0.650]$ and $AUC = 0.831$ with $IQR = [0.799,\ 0.861]$. The class specific performances are shown in Table 3.3 and Figure 3.5 illustrates the grid search of hyperparameters and confusion matrix.

By assessing the confusion matrix and the performances of each patient (Figure 3.5) and it can be deduced that the awake and NREM3 stages more separable than REM and NREM1-2. Similarly to other classifiers there exists significant variability between patients. It can be seen that detecting NREM1-2 and REM is extremely difficult in some patients. Awake and NREM3 are the most identifiable states with relative few infants with poor performance.

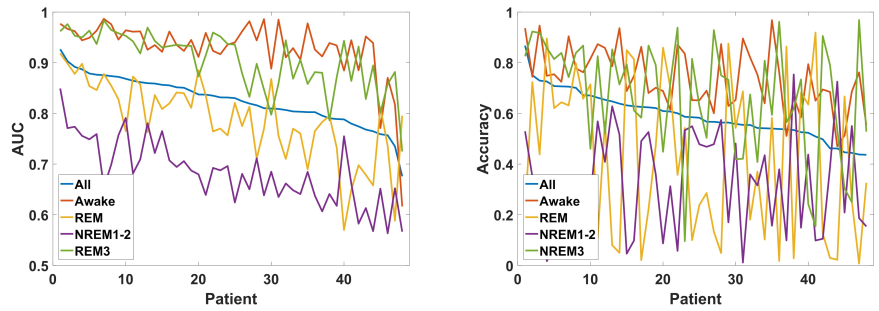
Table 3.3: Awake - REM - NREM1-2 - NREM3 - classifier median (IQR) performances.

	Total	Awake	REM	NREM1-2	NREM3
AUC	0.831 [0.799,0.861]	0.940 [0.919,0.962]	0.795 [0.756,0.844]	0.683 [0.640,0.717]	0.931 [0.870,0.947]
ACC	0.584 [0.538,0.650]	0.745 [0.653,0.828]	0.456 [0.117,0.684]	0.3423 [0.137,0.486]	0.733 [0.524,0.839]



(a) Grid search for hyperparameters **(b)** Confusion matrix.

Figure 3.5: Awake - REM - NREM1-2 - NREM3 hyperparameter optimization and pooled confusion matrix. **(a)** Hyperparameters, KS and BC with the highest AUC value, indicated as triangle, were chosen. **(b)** Confusion matrix of pooled epochs over 48 different LOPO-CV rounds.



(a) Patient AUC **(b)** Patient accuracy

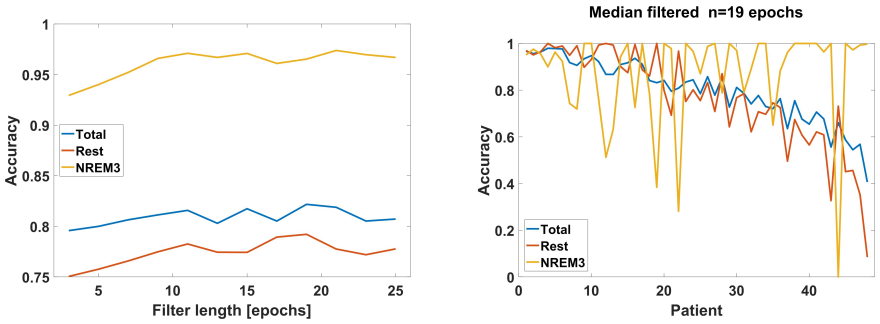
Figure 3.6: Awake - REM - NREM1-2 - NREM3 classification patient *AUC* and *ACC* values in descending performance order. **(a)** Patient *AUC* values and **(b)**accuracies from the best to worst.

3.4 Postprocessed NREM3 vs Rest

The implementation of the SVM classifies the sleep state on an epoch by epoch basis. It does not take into account the classification of nearby epochs which may contain important temporal information. We adapt the NREM3 vs the rest classifier to take into account any temporal information using an additional postprocessing stage, In postprocessing a median filter is applied to the SVM output, the loss-score value. Twelve median filters with different epoch lengths were tested. Because an aim of postprocessing is to move from the precise sleep scoring to a long-term detection of sleep trend. The manual hypnogram was also filtered. The results are shown in Figure 3.7a, where it can be seen that the length of 17 epochs yields the best median accuracy, $ACC = 0.822$ with $IQR = [0.724, 0.914]$ (Table 3.4). Because the original epoch length was set to 30 seconds this corresponds to 9.5 min median filtering.

Table 3.4: Median (inter-quartile range) performances in NREM3 - detection after postprocessing.

	Total	NREM3	Rest
ACC	0.822 [0.724,0.914]	0.965 [0.789,1.000]	0.792 [0.658,0.958]

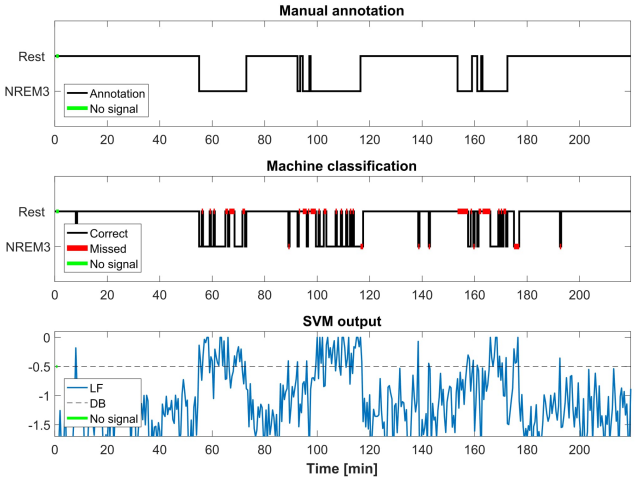


(a) Accuracy with different filter lengths **(b)** Postprocessed patient accuracy

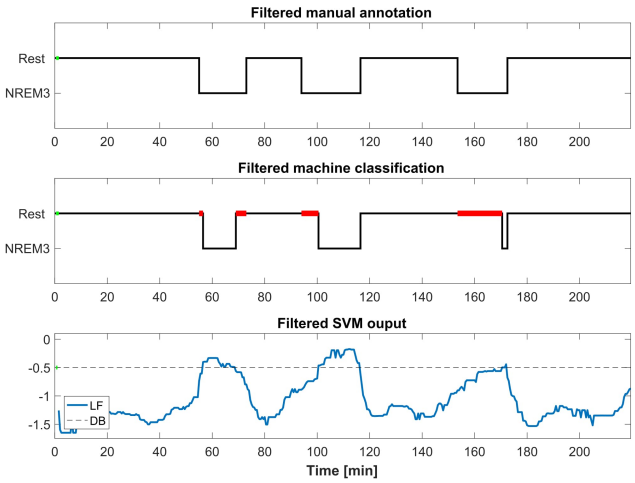
Figure 3.7: Median accuracy with different median filter lengths. **(b)** The best accuracy is achieved with the median filter length of n=19 epochs. **(a)** Slight improvement on patient accuracy after applying the filter to SVM output.

The accuracies for each patient can be found in Appendix Table B.1. For visualization five patients (id=$\{19, 18, 33, 39, 21\}$) were chosen based on the performance before preprocessing; the best, the quartiles (Q_1,Q_2,Q_3) and the worst, respectively. In this Chapter the patients #18, #39 and #21 are presented and the remaining two can be seen in Appendix B.

The effect of postprocessing is clearly visible in patient #18 shown in Figure 3.8. The original manual scoring is depicted in the top plot of Figure 3.8a, where three NREM3 periods can be recognized. Two latter periods are not fully continuous and are interrupted with short state transitions. The middle plot shows SVM based classification and bottom is the SVM loss-function value (LF), which is scaled to interval $[-\infty, 0]$. The dotted line indicates the decision boundary (DB) on which the classification is based. With LF$\geq$DB the epoch is classified as NREM3. Because the SVM does not consider information from the preceding or proceeding epochs, the results are unstable. The median filtering removes instability and produces the long time scale sleep trend (the middle plot in Figure 3.8b), where the three NREM3 periods are distinguishable. There still remains error in state transitions. The last NREM3 period is shortened, although the LF is close to the decision boundary. Note that, this patient has the 4th worst NREM3 accuracy, $ACC = 0.546$ (Figure B.1b), but still manages to detect all NREM3 periods.



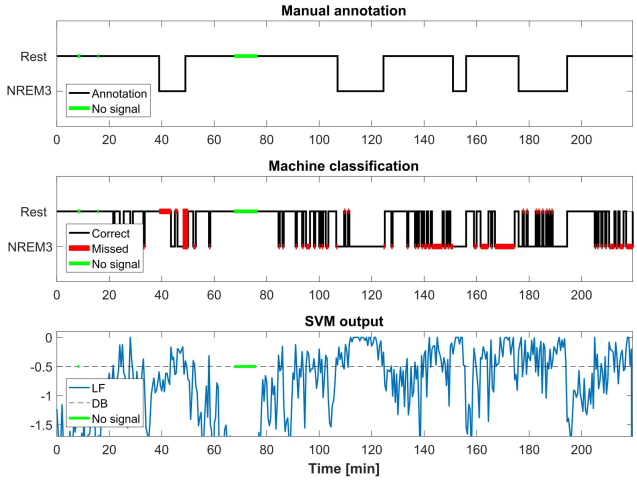
(a) Without postprocessing



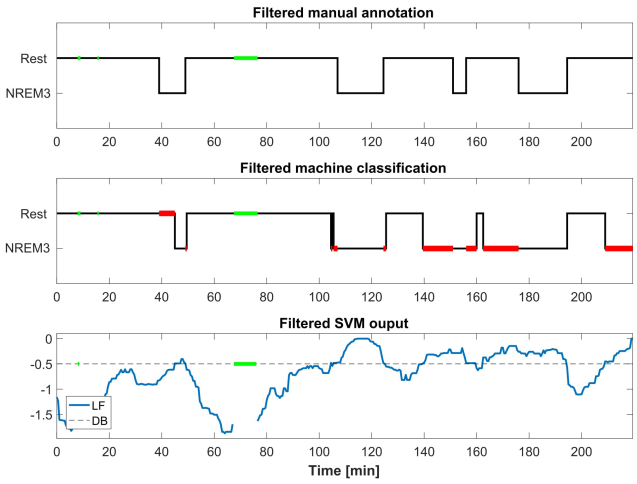
(b) Postprocessed results

Figure 3.8: The effect of postprocessing in the patient #18 that was ranked 12th in performance. **(a)** Because the SVM lacks temporal knowledge the classification results are quite unstable (the middle plot), **(b)** which can be corrected using the median filter as a post processing step. The loss-function (LF, blue) that depicts the distance from the class decision boundary (DB, dotted line) determines the class. If LF$\geq$BD the class is set to NREM3.

In Figure 3.9 is shown the patient #39. The original SVM result is temporally unstable but settles after filtering. The overall pattern is close to the filtered manual annotation but the timing of the state transitions are imprecise. Also the classifier wrongly assumes that there is a NREM3 period at the end of the recording.



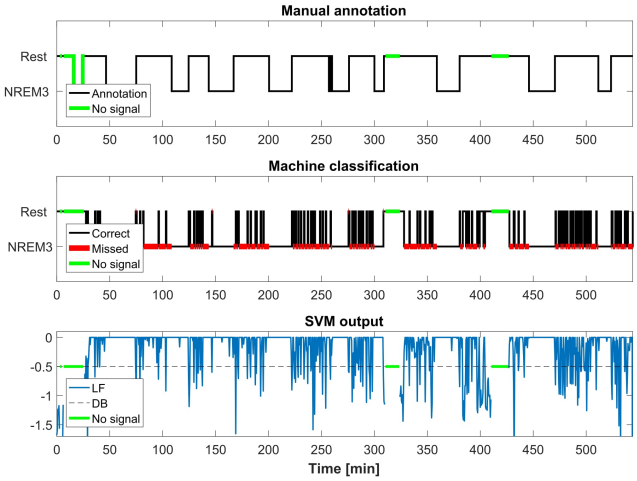
(a) Without postprocessing



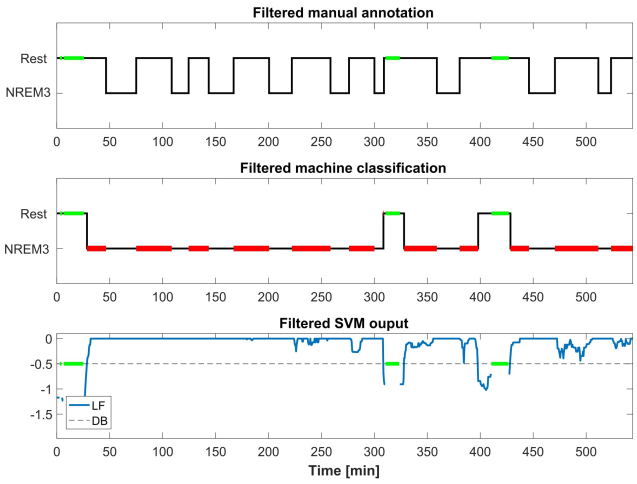
(b) Postprocessed results

Figure 3.9: The effect of postprocessing on the machine classification in the patient #39 ranked 36th. **(a)** Without postprocessing there are several unnecessary state transitions and **(b)** the postprocessing reveals the continuous NREM3 periods with a moderate accuracy.

The last subject #21 presented here illustrates the worst case scenario. See Figure 3.10. From the original SVM plot we can see that the LF is strongly biased toward NREM3. The rest-class detection is very unstable, but from the silhouette of the SVM classification it can be seen that these rest periods match with the manual annotation. Postprocessing the filtering has a negative effect and sets almost whole recording to NREM3.



(a) Without postprocessing



(b) Postprocessed results

Figure 3.10: The worst ranked patient #21 and the effect of the postprocessing. **(a)** The SVM classification results show that the LF is heavily biased towards NREM3 although the silhouette of machine hypnorgram matches with manual scoring to some extend. **(b)** In this case postprocessing does more harm and sets almost whole recording period to NREM3.

Finally it was evaluated would the performance be dependent on age or diagnosis, see Figure 3.11. Because of the lack of clear separable pattern there is no indication that the diagnosis or age would affect on the performance. Also Mann-Whitney U test for the equality of medians between the total accuracy of patients with breathing problem ($n = 28$) and healthy ($n = 20$) does not distinguish the groups with a p-value of 0.9084. Similarly when the patients were divided based on age; < 10 weeks old with $n = 35$ and ≥ 10 with $n = 13$, the test yielded insignificant results with a p-value of 0.8710.

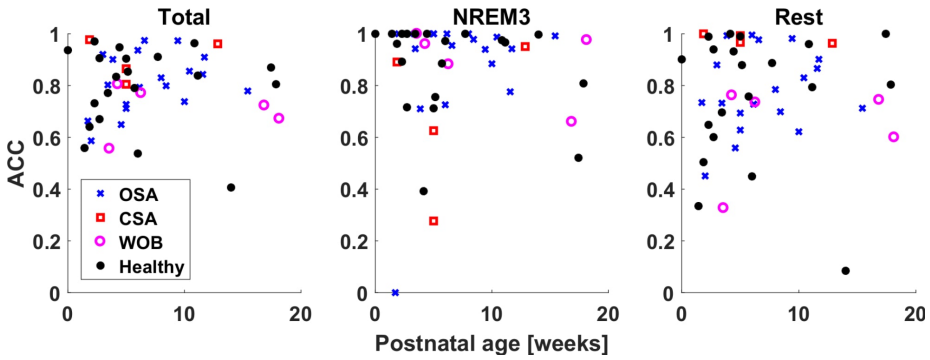


Figure 3.11: Performance scatter plots as a function of age for each diagnosis. Left plot indicates total, middle NREM3 and right Rest accuracies. There are no clear indication that the performance would be dependent on age or diagnosis.

Chapter 4

Discussion

While unable to discern all sleeps states, the classifier is able to detect deep sleep with relatively high accuracy. This accuracy is comparable to other non-PSG methods. The detection of the NREM3 stage permits the estimation of vigilance state cycling in infants.

The differentiation of other states such as NREM1 and REM is far more difficult with the sleep mattress signal and ECG. These difficulties are apparent as infants lack adult characteristic muscle atonia during REM resulting in extensive gross-movements which are typically associated with awake and NREM1. Also the individual performances vary considerably, which are not affected by the age or diagnosis as examined in the previous chapter. This calls for more thorough investigation for methods to achieve patient invariance.

4.1 Comparison to previous work

As summarized in Table 1.4, infant sleep classifiers have been designed for active sleep, quiet sleep and awake discrimination [94, 97, 98, 101] and active vs quiet detection [96, 102, 105, 106]. Also several sleep vs awake classifiers have been proposed [78, 99, 100, 104].

The classifiers evaluated by Thoman et al. [94] and Hukka [98] applied BMS. Thoman et al. reported the total accuracy of 81% (active, quiet, transition, awake) for their respiration and cross-body movement based classifier, which was in conjunction with their earlier manual BMS sleep staging [95]. However, they only assessed their algorithm with a small patient set ($n = 10$). Similarly Hukka analysed small group of only 15 infants and she reported 59% accuracy, which is lower than achieved in this work (68%) and could indicate that including heart based features into analysis can improve classification results. Proposing manual scoring criteria, Kirjavainen et al. introduced a novel methodology to include paroxysmal phasic activation events and achieved 85% agreement with PSG. Although the REM detection rate was only 61%, which could be due the fact that intermediate state (epoch with both REM and NREM characteristics) was set to NREM in their evaluation.

Previously reported cardiorespiratory classifiers have claimed to be able to separate active sleep (REM) from quiet sleep (NREM) [94, 96, 97, 102, 105, 106]. One reason for the lower performance in our methods could be due to the different manual annotation approach. Applied scoring criteria here do not have indeterminate class for difficult epochs that fall between active and quiet stage

characteristic. Haddad et al. [96] and Harper et al. [97] excluded indeterminate epochs from their analysis. Also the recurrence plot analysis considered only long continuous epochs [102]. Isler et al. did not state explicitly how they handled indeterminate epochs [105]. Werth et al. [106] used Prechtl scoring criteria [58] that does not have indeterminate state, but they discarded 21% of the epochs due to signal quality and inter-scorer disagreement. It is challenging to compare classifiers between different authors due to the different data sets and methods.

The results from the awake vs REM vs NREM and awake vs REM vs NREM1/2 vs NREM3 classifiers show that the awake state can be separated from other classes. This is due to the fact that infants have many gross movements during awake periods, which is easily seen in the BMS signal by high amplitude periods. The results are in concordance with previous infant classifiers (Table 1.4). In adults, movement based classifiers have difficulties due to quiet awake that contain only little movement.

The proposed NREM3 classifier showed difficulty in detecting sleep state transitions. Determining the exact location of state transition is a well-known problem. The transitions from different NREM states form a continuum and there is also disagreement between manual annotators [140]. Another aspect that lowers the accuracy in state transitions is the observed time delay between cortical phenomena and phenomena originating from the brain stem and basal forebrain. The EEG which measures cortical phenomena is especially important in PSG scoring, whereas the heart and respiration related phenomena, manifestations of autonomic nervous system, are the basis of classification here. It has been reported that the autonomous nervous system changes precede the EEG changes by 1-3 min [92].

4.2 Inter-observer agreement

Sleep scoring is not unanimous between human annotators. For infant sleep scoring the inter-observer agreement have been reported to be 80.6% [141] and this has been reported to be similar in adults too, 83.6% [140]. This raises a significant problem when designing an automatic sleep classifier. Because there is disagreement even with human observers, high classification performance approaching close to total agreement, typically pursued in the machine learning scene, is not a realistic goal in sleep classification.

A more realistic approach for assessing performance of a sleep classifier could be based on the temporal sleep state sequence alignment. Assessing performance based on some edit distance measure [142] would diminish the problems caused by determining the exact locations for state transitions although excessive transitions that are not observed in the manual hypnogram could be heavily penalized. In this work it was shown that without postprocessing the output sleep sequence contained an unrealistic amount of state transitions (e.g. see Figure 3.8). Although after postprocessing the performance values were not radically improved, however, the outcome was more sensible based on visual inspection. In this sense performance evaluation based on edit distance measures would capture the temporal sensibility better than epoch based accuracies and AUC values.

4.3 Future directions

The proposed classifier uses a large amount of features. There are, however, some interesting phenomena that are not covered in this work. The phase synchronization between heart and respiration rate has aroused curiosity in the pediatric community. It has been reported that cardio-respiratory phase locking is more common in quiet than active sleep [143]. The disadvantage of having large amount of features is the inflated complexity of the classifier which may result in excessive compute times. Feature redundancy was not assessed in this work and was left for a future endeavour.

Another open question is the optimal epoch length. There is increasing evidence that 30-second window, which is standard in manual PSG scoring, is too short for a cardiorespiratory and gross-body movement based classifier. Without postprocessing step the results of our classifier were unstable with several false state transitions. Thoman et al. also smoothed their 30-second epoch classifier result [94]. An alternative approach is to include several preceding epochs as input for the classifier [99–101] or use the Prechtl scoring cirteria [58] that requires a state to persist for a minimum duration of 3 min [105, 106]. Other way is to calculate the features from longer epochs, from 1 to 5 minutes [96–98, 102, 106]. Werth et al. reported that their AUC improved from 0.71 to 0.87 when they changed the epoch length from 30 sec to 5 min.

The ultimate goal is to monitor transitions in infant sleep states unobtrusively. The ECG requires the attachment of electrodes to an infant. It was investigated if it would be possible to extract infant heart beats from BCG, but these were too weak to be detected robustly. We, therefore, used ECG for the purpose of adding heart related feature vectors. In future work, the ECG would be excluded. The results of previously reported bed mattress sensor classifiers indicate that this could be viable venture.

Despite these challenges, the proposed method permits the estimation of vigilance state cycling in infants. If integrated into monitors in neonatal intensive care units the method would allow the estimation of individual sleep trend patterns. These in turn would help clinical staff to set personalized nursing schedules, adapt alarm sounds and monitor sleep architecture for normal neurological development [144]. Besides using BSM for sleep classification, algorithms could be developed further to detect sleep related breathing disorders.

The method could be applied in longitudinal development studies with large sample sizes to provide sleep trend estimates. Large studies are impractical with traditional manual PSG scoring. These studies could provide information how sleep affects the infant's cognitive, psychomotor, temperament and developmental outcomes.

A BSM-based classifier could be also introduced into a home setting, alike to the consumer products targeted for the adults [107, 145].

Chapter 5

Conclusions

Infants spend majority of their time asleep. Although there have been many studies on the effects of sleep deprivation, sleep onset time, duration and fragmentation on the infant's cognitive, psychomotor, temperament and developmental outcomes, the causal links remain elusive. One reason for contradictory results is the difficulty of long term, reliable sleep assessment. Infant sleep questionnaires are too vague to assess exact structure of sleep and, actigraphy cannot be used to discriminate the sleep stages. The gold standard, polysomnography, is very obtrusive and requires attachment of several electrodes and cables. and requires manual annotation.

Besides basic research purposes, unobtrusive and automatic sleep monitoring could have a clinical value in preterm/neonatal intensive care units, where it could be for example used to plan a nursing schedule so that it hinders minimally infant sleep. Unobtrusive sleep trackers designed for adults have become available in the consumer market, whereas there are no equivalent for infants.

In this work, a signal processing and machine learning pipeline for infant sleep stage prediction from the EMFi bed mattress sensor and ECG was introduced. The proposed method is suitable for long term sleep monitoring due to the minimal use of sensors in contact with the infant. Data from a clinical study was used to train the pipeline and provided authentic challenges.

This work evaluated three radial basis function support vector machine classifiers; 1) NREM3 vs rest 2) awake vs REM vs NREM and 3) awake vs REM vs NREM1/2 vs NREM3. The latter classifiers had great difficulties distinguishing REM from NREM1/2 (or NREM). The NREM3 vs rest had a promising performance and after temporal smoothing it was shown that this could be used for monitor infant vigilance state oscillations. The classifier had median accuracy of 0.822 with interquartile range of [0.724,0.914]. Although several challenges remain, the work provides a proof of concept for infant NREM3 stage monitoring based on the EMFi mattress and ECG.

Bibliography

[1] C. Cirelli and G. Tononi, "Is sleep essential?" *PLOS Biology*, vol. 6, no. 8, pp. 1–7, 08 2008. [Online]. Available: https://doi.org/10.1371/journal.pbio.0060216

[2] M. F. Bear, B. W. Connors, and M. A. Paradiso, *Neroscience: Exploring the Brain*, 3rd ed. Lippincott Williams and Wilkins, 2006.

[3] A. C. Skeldon, G. Derks, and D.-J. Dijk, "Modelling changes in sleep timing and duration across the lifespan: changes in circadian rhythmicity or sleep homeostasis?" *Sleep medicine reviews*, vol. 28, pp. 96–107, 2016.

[4] J. A. Hobson, "Sleep is of the brain, by the brain and for the brain," *Nature*, vol. 437, no. 7063, pp. 1254–1256, 2005.

[5] B. Faraut, K. Z. Boudjeltia, L. Vanhamme, and M. Kerkhofs, "Immune, inflammatory and cardiovascular consequences of sleep restriction and recovery," *Sleep medicine reviews*, vol. 16, no. 2, pp. 137–149, 2012.

[6] O. M. Buxton, S. W. Cain, S. P. O'Connor, J. H. Porter, J. F. Duffy, W. Wang, C. A. Czeisler, and S. A. Shea, "Adverse metabolic consequences in humans of prolonged sleep restriction combined with circadian disruption," *Science translational medicine*, vol. 4, no. 129, pp. 129ra43–129ra43, 2012.

[7] M. Dattilo, H. K. M. Antunes, A. Medeiros, M. Mônico-neto, H. d. S. Souza, K. S. Lee, S. Tufik, and M. T. de Mello, "Paradoxical sleep deprivation induces muscle atrophy," *Muscle & nerve*, vol. 45, no. 3, pp. 431–433, 2012.

[8] R. C. Anafi, R. Pellegrino, K. R. Shockley, M. Romer, S. Tufik, and A. I. Pack, "Sleep is not just for the brain: transcriptional responses to sleep in peripheral tissues," *BMC genomics*, vol. 14, no. 1, p. 362, 2013.

[9] H. Wang, Y. Liu, M. Briesemann, and J. Yan, "Computational analysis of gene regulation in animal sleep deprivation," *Physiological genomics*, vol. 42, no. 3, pp. 427–436, 2010.

[10] M. Geyfman, V. Kumar, Q. Liu, R. Ruiz, W. Gordon, F. Espitia, E. Cam, S. E. Millar, P. Smyth, A. Ihler *et al.*, "Brain and muscle arnt-like protein-1 (bmal1) controls circadian cell proliferation and susceptibility to uvb-induced dna damage in the epidermis," *Proceedings of the National Academy of Sciences*, vol. 109, no. 29, pp. 11 758–11 763, 2012.

[11] C. Stringari, H. Wang, M. Geyfman, V. Crosignani, V. Kumar, J. S. Takahashi, B. Andersen, and E. Gratton, "In vivo single-cell detection of metabolic oscillations in stem cells," *Cell reports*, vol. 10, no. 1, pp. 1–7, 2015.

[12] R. Ohkubo and D. Chen, "Aging: rewiring the circadian clock." *Nature structural & molecular biology*, vol. 24, no. 9, p. 687, 2017.

[13] E. Aserinsky and N. Kleitman, "Regularly occurring periods of eye motility, and concomitant phenomena, during sleep," *Science*, vol. 118, no. 3062, pp. 273–274, 1953.

[14] O. G. Jenni, A. A. Borbély, and P. Achermann, "Development of the nocturnal sleep electroencephalogram in human infants," *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, vol. 286, no. 3, pp. R528–R538, 2004. [Online]. Available: http://ajpregu.physiology.org/content/286/3/R528

[15] A. W. de Weerd and R. A. van den Bossche, "The development of sleep during the first months of life," *Sleep Medicine Reviews*, vol. 7, no. 2, pp. 179–191, 2003.

[16] E. Van Cauter, K. Spiegel, E. Tasali, and R. Leproult, "Metabolic consequences of sleep and sleep loss," *Sleep medicine*, vol. 9, pp. S23–S28, 2008.

[17] T. Abel, R. Havekes, J. M. Saletin, and M. P. Walker, "Sleep, plasticity and memory from molecules to whole-brain networks," *Current biology*, vol. 23, no. 17, pp. R774–R788, 2013.

[18] R. Stickgold and M. P. Walker, "Sleep-dependent memory triage: evolving generalization through selective processing," *Nature neuroscience*, vol. 16, no. 2, pp. 139–145, 2013.

[19] S. Ackermann and B. Rasch, "Differential effects of non-rem and rem sleep on memory consolidation?" *Current neurology and neuroscience reports*, vol. 14, no. 2, p. 430, 2014.

[20] C. von Economo, "Sleep as a problem of localization," *Journal of Nervous and Mental Disease*, vol. 71, no. 3, pp. 249–259, 1930.

[21] M. D. Schwartz and T. S. Kilduff, "The neurobiology of sleep and wakefulness," *The Psychiatric clinics of North America*, vol. 38, no. 4, p. 615, 2015.

[22] R. A. España and T. E. Scammell, "Sleep neurobiology from a clinical perspective," *Sleep*, vol. 34, no. 7, pp. 845–858, 2011.

[23] J. J. Fraigne, Z. A. Torontali, M. B. Snow, and J. H. Peever, "Rem sleep at its core–circuits, neurotransmitters, and pathophysiology," *Frontiers in neurology*, vol. 6, 2015.

[24] E. Bathory and S. Tomopoulos, "Sleep regulation, physiology and development, sleep duration and patterns, and sleep hygiene in infants, toddlers, and preschool-age children," *Current problems in pediatric and adolescent health care*, vol. 47, no. 2, pp. 29–42, 2017.

[25] D. K. Welsh, J. S. Takahashi, and S. A. Kay, "Suprachiasmatic nucleus: cell autonomy and network properties," *Annual review of physiology*, vol. 72, pp. 551–577, 2010.

[26] G. J. Tortora and B. Derrickson, *Principles of Anatomy & Physiology Maintenance and Vontinuity of the Human Body*, 13th ed. John Wiley & Sons Inc., 2011.

[27] O. Dergacheva, L. A. Weigand, J. Dyavanapalli, J. Mares, X. Wang, and D. Mendelowitz, "Function and modulation of premotor brainstem parasympathetic cardiac neurons that control heart rate by hypoxia-, sleep-, and sleep-related diseases including obstructive sleep apnea." *Progress in brain research*, vol. 212, pp. 39–58, 2013.

[28] N. M. Mellen and M. Thoby-Brisson, "Respiratory circuits: development, function and models," *Current opinion in neurobiology*, vol. 22, no. 4, pp. 676–685, 2012.

[29] K. Heimann, N. Heussen, P. Vaessen, C. Wallmeier, T. Orlikowsky, and T. G. Wenzl, "Basic values for heart and respiratory rates during different sleep stages in healthy infants," *Biomedizinische Technik/Biomedical Engineering*, vol. 58, no. 1, pp. 27–34, 2013.

[30] F. Chouchou and M. Desseilles, "Heart rate variability: a tool to explore the sleeping brain?" *Frontiers in neuroscience*, vol. 8, p. 402, 2014.

[31] F. Yasuma and J.-i. Hayano, "Respiratory sinus arrhythmia: why does the heartbeat synchronize with respiratory rhythm?" *Chest Journal*, vol. 125, no. 2, pp. 683–690, 2004.

[32] N. De Beer, P. Andriessen, R. Berendsen, S. Oei, P. Wijn, and S. B. Oetomo, "Customized spectral band analysis compared with conventional fourier analysis of heart rate variability in neonates," *Physiological measurement*, vol. 25, no. 6, p. 1385, 2004.

[33] K. R. Ross and C. L. Rosen, "Sleep and respiratory physiology in children," *Clinics in chest medicine*, vol. 35, no. 3, pp. 457–467, 2014.

[34] P. Peirano, C. Algarin, and R. Uauy, "Sleep-wake states and their regulatory mechanisms throughout early human development," *The Journal of pediatrics*, vol. 143, no. 4, pp. 70–79, 2003.

[35] R. C. Knickmeyer, S. Gouttard, C. Kang, D. Evans, K. Wilber, J. K. Smith, R. M. Hamer, W. Lin, G. Gerig, and J. H. Gilmore, "A structural MRI study of human brain development from birth to 2 years," *Journal of Neuroscience*, vol. 28, no. 47, pp. 12 176–12 182, 2008.

[36] M.-s. Choe, S. Ortiz-Mantilla, N. Makris, M. Gregas, J. Bacic, D. Haehn, D. Kennedy, R. Pienaar, V. S. Caviness Jr, A. A. Benasich *et al.*, "Regional infant brain development: an MRI-based morphometric analysis in 3 to 13 month olds," *Cerebral Cortex*, vol. 23, no. 9, pp. 2100–2117, 2012.

[37] M. Ednick, A. P. Cohen, G. L. McPhail, D. Beebe, N. Simakajornboon, and R. S. Amin, "A review of the effects of sleep during the first year of life on cognitive, psychomotor, and temperament development," *Sleep*, vol. 32, no. 11, pp. 1449–1458, 2009.

[38] P. S. Douglas and P. S. Hill, "Behavioral sleep interventions in the first six months of life do not improve outcomes for mothers or infants: a systematic review," *Journal of Developmental & Behavioral Pediatrics*, vol. 34, no. 7, pp. 497–507, 2013.

[39] T. Field, "Infant sleep problems and interventions: A review," *Infant Behavior and Development*, vol. 47, no. 11, pp. 1449–1458, 2009.

[40] E. R. Hager, C. J. Calamaro, L. M. Bentley, K. M. Hurley, Y. Wang, and M. M. Black, "Nighttime sleep duration and sleep behaviors among toddlers from low-income families: Associations with obesogenic behaviors and obesity and the role of parenting," *Childhood Obesity*, vol. 12, no. 5, pp. 392–400, 2016.

[41] J. F. Bell and F. J. Zimmerman, "Shortened nighttime sleep duration in early life and subsequent childhood obesity," *Archives of Pediatrics & Adolescent Medicine*, vol. 164, no. 9, pp. 840–845, 2010.

[42] Y. Wu, L. Zhai, and D. Zhang, "Sleep duration and obesity among adults: a meta-analysis of prospective studies," *Sleep Medicine*, vol. 15, no. 12, pp. 1456–1462, 2014.

[43] V. Mavanji, C. J. Billington, C. M. Kotz, and J. A. Teske, "Sleep and obesity: a focus on animal models," *Neuroscience & Biobehavioral Reviews*, vol. 36, no. 3, pp. 1015–1029, 2012.

[44] E. Van Cauter, K. Spiegel, E. Tasali, and R. Leproult, "Metabolic consequences of sleep and sleep loss," *Sleep medicine*, vol. 9, pp. S23–S28, 2008.

[45] M. Hysing, B. Sivertsen, S. Garthus-Niegel, and M. Eberhard-Gran, "Pediatric sleep problems and social-emotional problems. a population-based study," *Infant Behavior and Development*, vol. 42, pp. 111–118, 2016.

[46] A. Sadeh, G. De Marcas, Y. Guri, A. Berger, L. Tikotzky, and Y. Bar-Haim, "Infant sleep predicts attention regulation and behavior problems at 3–4 years of age," *Developmental Neuropsychology*, vol. 40, no. 3, pp. 122–137, 2015.

[47] B. G. Cooper, J. E. White, L. A. Ashworth, K. G. M. Alberti, and G. J. Gibson, "Hormonal and metabolic profiles in subjects with obstructive sleep apnea syndrome and the acute effects of nasal continuous positive airway pressure (CPAP) treatment," *Sleep*, vol. 18, no. 3, pp. 172–179, 1995.

[48] I. F. Wagner, "Curves of sleep depth in newborn infants," *The Pedagogical Seminary and Journal of Genetic Psychology*, vol. 55, no. 1, pp. 121–135, 1939.

[49] E. Aserinsky and N. Kleitman, "Two types of ocular motility occurring in sleep," *Journal of applied physiology*, vol. 8, no. 1, pp. 1–10, 1955.

[50] ——, "A motility cycle in sleeping infants as manifested by ocular and gross bodily activity," *Journal of Applied Physiology*, vol. 8, no. 1, pp. 11–18, 1955.

[51] 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, 1957.

[52] C. Dreyfus-Brisac and N. Monod, "Sleep of premature and full-term neonates–a polygraphic study." *Proceedings of the Royal Society of Medicine*, vol. 58, pp. 6–7, 1965.

[53] A. H. Parmelee, F. J. Schulte, Y. Akiyama, W. H. Wenner, M. A. Schultz, and E. Stern, "Maturation of eeg activity during sleep in premature infants," *Electroencephalography and Clinical Neurophysiology*, vol. 24, no. 4, pp. 319–329, 1968.

[54] A. Parmelee, Y. Akiyama, E. Stern, and M. Harris, "A periodic cerebral rhythm in newborn infants," *Experimental Neurology*, vol. 25, no. 4, pp. 575–584, 1969.

[55] S. E. Beck and C. L. Marcus, "Pediatric polysomnygraphy," *Sleep Medicine Clinics*, vol. 4, no. 3, pp. 393–406, 2009.

[56] A. Rechtschaffen and A. Kales, *A manual of standardized technology, techniques, and scoring system for sleep stages of human subjects*, UCLA Brain Information Service/ Brain Research Institute, Los Angeles, 1968.

[57] T. F. Anders, R. N. Emde, and A. H. Parmelee, *A Manual of Standardized Terminology, Techniques and Criteria for Scoring of Sleep and Wakefullness in Newborn Infants*, UCLA Brain Information Service/ BRI Publications Office, Los Angeles, 1971.

[58] H. F. Prechtl, "The behavioural states of the newborn infant (a review)," *Brain research*, vol. 76, no. 2, pp. 185–212, 1974.

[59] T. Hoppenbrouwers, *Sleep in infants*, Raven Press, New York, 1987.

[60] C. Quilleminault and M. Souquet, *Sleep states and related pathology*, Spectrum ed., New York, 1968.

[61] D. H. Crowell, L. J. Brooks, T. Colton, M. J. Corwin, T. T. Hoppenbrouwers, C. E. Hunt, L. E. Kapuniai, G. Lister, M. R. Neuman, M. Peucker *et al.*, "Infant polysomnography: reliability," *Sleep*, vol. 20, no. 7, pp. 553–560, 1997.

[62] S. Scholle and T. Schafer, "Atlas of states of sleep and wakefulness in infants and children," *Somnologie*, vol. 3, no. 3, pp. 163–165, 1999.

[63] M. Grigg-Damberger, D. Gozal, C. L. Marcus, S. F. Quan, C. L. Rosen, R. D. Chervin, M. Wise, D. L. Picchietti, S. H. Sheldon, and C. Iber, "The visual scoring of sleep and arousal in infants and children," *Journal of Clinical Sleep Medicine*, vol. 3, no. 2, pp. 201–240, 2007.

[64] C. Brüser, C. H. Antink, T. Wartzek, M. Walter, and S. Leonhardt, "Ambient and unobtrusice cardiorespiratory monitoring techniques," *IEEE Reviews in Biomedical Engineering*, vol. 8, pp. 30–43, 2015.

[65] J. Werth, L. Atallah, P. Andriessen, X. Long, E. Zwartkruis-Pelgrim, and R. M. Aarts, "Unobtrusive sleep state measurements in preterm infants - a review," *Sleep Medicine Reviews*, vol. 67, no. 4, pp. 379–387, 1987.

[66] A. Sadeh, P. J. Hauri, D. F. Kripke, and P. Lavie, "The role of actigraphy in the evaluation of sleep disorders," *Sleep*, vol. 18, no. 4, pp. 288–302, 1995.

[67] A. Sadeh and C. Acebo, "The role of actigraphy in sleep medicine," *Sleep Medicine Reviews*, vol. 6, no. 2, pp. 113–124, 2002.

[68] I. Starr, A. Rawson, H. Schroeder, and N. Joseph, "The estimation of cardiac ouptut in man, and of abnormalities in cardiac function, from the heart's recoil and the blood's impacts; the ballistocardiogram," *American Journal of Physiology - Legacy Content*, vol. 127, no. 1, pp. 1–28, 1939.

[69] J. Alihanka, K. Vaahtoranta, and I. Saarikivi, "A new method for long-term monitoring of the ballistocardiogram, heart rate, and respiration," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 240, no. 5, pp. R384–R392, 1981.

[70] F. Wang, M. Tanaka, and S. Chonan, "Development of a PVDF piezopolymer sensor for unconstrained in-sleep cardiorespiratory monitoring," *Journal of Intelligent Material Systems and Structures*, vol. 14, no. 3, pp. 185–190, 2003.

[71] H. Raisanen, "Electromechanical transducer and method for manufacturing an electromechanical transducer," Patent US 6 916 979B2, 7 12, 2005.

[72] E. B. Thoman and W. D. Tynan, "Sleep states and wakefulness in human infants: Profiles from motility monitoring," *Physiology & Behavior*, vol. 23, no. 3, pp. 519–525, 1979.

[73] M. Erkinjuntti, P. Kero, J.-P. Halonen, H. Mikola, and K. Sainio, "Scsb method compared to EEG-based polygraphy in sleep state scoring of newborn infants," *Acta Paediatrica*, vol. 79, no. 3, pp. 274–279, 1990.

[74] T. Kirjavainen, D. Cooper, O. Polo, and C. E. Sullivan, "Respiratory and body movements as indicators of sleep stage and wakefulness in infants and young children," *Journal of sleep research*, vol. 5, no. 3, pp. 186–194, 1996.

[75] D. P. Redmond and F. W. Hegge, "Observations on the design and specification of a wrist-worn human activity monitoring system," *Behavior Research Methods, Instruments, & Computers*, vol. 17, no. 6, pp. 659–669, 1985.

[76] R. J. Cole, D. F. Kripke, W. Gruen, D. J. Mullaney, and J. C. Gillin, "Automatic sleep/wake identification from wrist activity," *Sleep*, vol. 15, no. 5, pp. 461–469, 1992.

[77] A. Sadeh, C. Acebo, R. Seifer, S. Aytur, and M. A. Carskadon, "Activity-based assessment of sleep-wake patterns during the 1st year of life," *Infant Behavior and Development*, vol. 18, no. 3, pp. 329–337, 1995.

[78] B. C. Galland, G. J. Kennedy, E. A. Mitchell, and B. J. Taylor, "Algorithms for using an activity-based accelerometer for identification of infant sleep–wake states during nap studies," *Sleep Medicine*, vol. 13, no. 6, pp. 743–751, 2012.

[79] B. H. Jansen and K. Shankar, "Sleep staging with movement-related signals," *International journal of bio-medical computing*, vol. 32, no. 3, pp. 289–297, 1993.

[80] S. J. Redmond, P. de 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, 2007.

[81] W. Karlen, C. Mattiussi, and D. Floreano, "Improving actigraph sleep/wake classification with cardio-respiratory signals," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*. IEEE, 2008, pp. 5262–5265.

[82] M. Mendez, M. Matteucci, S. Cerutti, A. Bianchi, and J. M. Kortelainen, "Automatic detection of sleep macrostructure based on bed sensors," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE, 2009, pp. 5555–5558.

[83] D. C. Mack, J. T. Patrie, R. A. Felder, P. M. Suratt, and M. Alwan, "Sleep assessment using a passive ballistocardiography-based system: Preliminary validation," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE, 2009, pp. 4319–4322.

[84] M. Mendez, M. Matteucci, S. Cerutti, F. Aletti, and A. Bianchi, "Sleep staging classification based on hrv: Time-variant analysis," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE, 2009, pp. 9–12.

[85] M. Migliorini, A. M. Bianchi, D. Nisticò, J. Kortelainen, E. Arce-Santana, S. Cerutti, and M. O. Mendez, "Automatic sleep staging based on ballistocardiographic signals recorded through bed sensors," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. IEEE, 2010, pp. 3273–3276.

[86] M. O. Mendez, M. Migliorini, J. M. Kortelainen, D. Nistico, E. Arce-Santana, S. Cerutti, and A. M. Bianchi, "Evaluation of the sleep quality based on bed sensor signals: Time-variant analysis," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE.* IEEE, 2010, pp. 3994–3997.

[87] J. M. Kortelainen, M. O. 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.

[88] G. S. Chung, J. S. Lee, S. H. Hwang, Y. K. Lim, D.-U. Jeong, and K. S. Park, "Wakefulness estimation only using ballistocardiogram: Nonintrusive method for sleep monitoring," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE.* IEEE, 2010, pp. 2459–2462.

[89] S. M. Isa, I. Wasito, A. M. Arymurthy *et al.*, "Kernel dimensionality reduction on sleep stage classification using ECG signal," *International Journal of Computer Science Issues*, 2011.

[90] D. Austin, Z. T. Beattie, T. Riley, A. M. Adami, C. C. Hagen, and T. L. Hayes, "Unobtrusive classification of sleep and wakefulness using load cells under the bed," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE.* IEEE, 2012, pp. 5254–5257.

[91] G. Guerrero-Mora, P. Elvia, A. Bianchi, J. Kortelainen, M. Tenhunen, S. Himanen, M. Mendez, E. Arce-Santana, and O. Gutierrez-Navarro, "Sleep-wake detection based on respiratory signal acquired through a pressure bed sensor," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE.* IEEE, 2012, pp. 3452–3455.

[92] X. Long, "On the analysis and classification of sleep stages from cardiorespiratory activity," Ph.D. dissertation, Eindhoven: Technische Universiteit Eindhoven, 2015.

[93] H. Kaji, H. Iizuka, and M. Hayashi, "Sleep stage prediction using respiration and body-movement based on probabilistic classifier," in *International Conference on Neural Information Processing.* Springer, 2016, pp. 491–500.

[94] E. B. Thoman and R. C. Glazier, "Computer scoring of motility patterns for states of sleep and wakefulness: Human infants," *Sleep*, vol. 10, no. 2, pp. 122–129, 1987.

[95] E. B. Thoman, "Sleep and wake behaviors in neonates: consistencies and consequences," *Merrill-Palmer Quarterly of Behavior and Development*, vol. 21, no. 4, pp. 295–314, 1975.

[96] G. G. Haddad, H. J. Jeng, T. L. Lai, and R. B. Melling, "Determination of sleep state in infants using respiratory variability," *Pediatric Research*, vol. 21, no. 6, pp. 556–562, 1987.

[97] R. M. Harper, V. L. Schechtman, and K. A. Kluge, "Machine classification of infant sleep state using cardiorespiratory measures," *Electroencephalography and clinical neurophysiology*, vol. 67, no. 4, pp. 379–387, 1987.

[98] S. Hukka, "Automatic SCSB analysis of motor and autonomic nervous activity in infants compared to eeg-based sleep stages," Master's thesis, Department of Psychology, University School of Jyväskylä, Jyväskylä, Finland, 2002, http://urn.fi/URN:NBN:fi:jyu-2002893375.

[99] A. Lewicke, E. Sazonov, and S. Schuckers, "Sleep-wake identification in infants: Heart rate variability compared to actigraphy," in *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE*, vol. 1. IEEE, 2004, pp. 442–445.

[100] A. Lewicke, E. Sazonov, M. J. Corwin, M. Neuman, and S. Schuckers, "Sleep versus wake classification from heart rate variability using computational intelligence: Consideration of rejection in classification models," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 1, pp. 108–118, 2008.

[101] N. A. Sazonova, E. E. Sazonov, B. Tan, S. A. Schuckers, C. S. Group *et al.*, "Sleep state scoring in infants from respiratory and activity measurements," in *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*. IEEE, 2006, pp. 2462–2465.

[102] P. I. Terrill, S. J. Wilson, S. Suresh, D. M. Cooper, and C. Dakin, "Attractor structure discriminates sleep states: Recurrence plot analysis applied to infant breathing patterns," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 5, pp. 1108–1116, 2010.

[103] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Physics Reports*, vol. 438, no. 5-6, pp. 237–329, 2007.

[104] G. Cohen and P. De Chazal, "A multi-modal approach to sleep-wake classification in infants using minimally invasive sensors," in *Computing in Cardiology Conference (CinC), 2014*. IEEE, 2014, pp. 149–152.

[105] J. R. Isler, T. Thai, M. M. Myers, and W. P. Fifer, "An automated method for coding sleep states in human infants based on respiratory rate variability," *Developmental Psychobiology*, vol. 58, no. 8, pp. 1108–1115, 2016.

[106] J. Werth, X. Long, E. Zwartkruis-Pelgrim, H. Niemarkt, W. Chen, R. M. Aarts, and P. Andriessen, "Unobtrusive assessment of neonatal sleep state based on heart rate variability retrieved from electrocardiography used for regular patient monitoring," *Early human development*, vol. 113, pp. 104–113, 2017.

[107] E. Ltd. (2018) Emfit - sensing what matters. [Online]. Available: https://www.emfit.com

[108] M. Paajanen, J. Lekkala, and K. Kirjavainen, "Electromechanical film (EMFi) - a new multipurpose electret material," *Sensors and Actuators A: Physical*, vol. 84, no. 1-2, pp. 95–102, 2000.

[109] V. Schetinin and L. Jakaite, "Extraction of features from sleep EEG for Bayesian assessment of brain development," *PloS one*, vol. 12, no. 3, p. e0174027, 2017.

[110] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, p. 1985, 230-236.

[111] X. Long, J. Foussier, P. Fonseca, R. Haakma, and R. M. Aarts, "Analyzing respiratory effort amplitude for automated sleep stage classification," *Biomedical Signal Processing and Control*, vol. 14, pp. 197–205, 2014.

[112] B. H. Jansen and K. Shankar, "Sleep staging with movement-related signals," *International Journal of Bio-medical Computing*, vol. 32, no. 3-4, pp. 289–297, 1993.

[113] M. Mendez, M. Matteucci, S. Cerutti, A. Bianchi, and J. M. Kortelainen, "Automatic detection of sleep macrostructure based on bed sensors," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE.* IEEE, 2009, pp. 5555–5558.

[114] J. Alametsä, E. Rauhala, E. Huupponen, A. Saastamoinen, A. Värri, A. Joutsen, J. Hasan, and S.-L. Himanen, "Automatic detection of spiking events in EMFi sheet during sleep," *Medical engineering & physics*, vol. 28, no. 3, pp. 267–275, 2006.

[115] B. Hjorth, "EEG analysis based on time domain properties analyse eeg basee sur les series temporelles," *Electroencephalography and Clinical Neurophysiology*, vol. 29, no. 3, 1970.

[116] M. H. Hayes, *Statistical Digital Signal Processing and Modeling.* John Wiley & Sons, INC, 1996.

[117] M. Mendez, M. Matteucci, S. Cerutti, F. Aletti, and A. Bianchi, "Sleep staging classification based on hrv: Time-variant analysis," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE.* IEEE, 2009, pp. 9–12.

[118] M. Migliorini, A. M. Bianchi, D. Nisticò, J. Kortelainen, E. Arce-Santana, S. Cerutti, and M. O. Mendez, "Automatic sleep staging based on ballistocardiographic signals recorded through bed sensors," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE.* IEEE, 2010, pp. 3273–3276.

[119] J. M. Kortelainen, M. O. 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.

[120] N. R. Lomb, "Least-squares frequency analysis of unequally spaced data," *Astrophysics and space science*, vol. 39, no. 2, pp. 447–462, 1976.

[121] S. M. Pincus, "Approximate entropy as a measure of system complexity." *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, 1991.

[122] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology - Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.

[123] W. Chen, Z. Wang, H. Xie, and W. Yu, "Characterization of surface EMG signal based on fuzzy entropy," *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 15, no. 2, pp. 266–272, 2007.

[124] W. Chen, J. Zhuang, W. Yu, and Z. Wang, "Measuring complexity using FuzzyeEn, ApEn, and SampEn," *Medical Engineering & Physics*, vol. 31, no. 1, pp. 61–68, 2009.

[125] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Physical Review Letters*, vol. 88, no. 17, 2002.

[126] B. B. Mandelbrot and J. R. Wallis, "Noah, Joseph, and operational hydrology," *Water resources research*, vol. 4, no. 5, pp. 909–918, 1968.

[127] S. H. Strogatz, *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering.* Westview press, 1994.

[128] F. Takens *et al.*, "Detecting strange attractors in turbulence," *Lecture notes in mathematics*, vol. 898, no. 1, pp. 366–381, 1981.

[129] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a time series," *Physical review letters*, vol. 45, no. 9, p. 712, 1980.

[130] M. T. Rosenstein, J. J. Collins, and C. J. De Luca, "A practical method for calculating largest lyapunov exponents from small data sets," *Physica D: Nonlinear Phenomena*, vol. 65, no. 1-2, pp. 117–134, 1993.

[131] E. Bradley and H. Kantz, "Nonlinear time-series analysis revisited," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 25, no. 9, p. 097610, 2015.

[132] M. B. Kennel, R. Brown, and H. D. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Physical review A*, vol. 45, no. 6, p. 3403, 1992.

[133] J. A. John and N. R. Draper, "An alternative family of transformations," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 29, no. 2, 1980.

[134] H. He, Y. Bai, E. A. Garcia, and S. Li, "Adasyn: Adaptive synthetic sampling approach for imbalanced learning," in *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on.* IEEE, 2008, pp. 1322–1328.

[135] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[136] E. Alpaydin, *Introduction to Machine Learning*, 2nd ed. The MIT Press, 2010.

[137] J. Weston and C. Watkins, "Multi-class support vector machines," Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London, May, Tech. Rep., 1998.

[138] T. G. Dietterich and G. Bakiri, "Solving multiclass learning problems via error-correcting output codes," *Journal of artificial intelligence research*, vol. 2, pp. 263–286, 1995.

[139] R.-E. Fan, P. H. Chen, and C.-J. Lin, "Working set selection using second order information for training support vector machines," *Journal of Machine Learning Research*, vol. 6, no. Dec., pp. 1889–1918, 2005.

[140] R. S. Rosenberg and S. Van Hout, "The American Academy of Sleep Medicine inter-scorer reliability program: sleep stage scoring," *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine*, vol. 9, no. 1, p. 81, 2013.

[141] A.-L. Satomaa, O. Saarenpää-Heikkilä, E. J. Paavonen, and S.-L. Himanen, "The adapted American Academy of Sleep Medicine sleep scoring criteria in one month old infants: A means to improve comparability?" *Clinical Neurophysiology*, vol. 127, no. 2, pp. 1410–1418, 2016.

[142] D. Gusfield, *Algorithms on strings, trees and sequences: computer science and computational biology*. Cambridge university press, 1997.

[143] M. Lucchini, N. Pini, W. P. Fifer, N. Burtchen, and M. G. Signorini, "Cardio-respiratory phase locking in newborn and one month infants as a function of sleep state," in *EMBEC & NBC 2017*. Springer, 2017, pp. 791–794.

[144] J. Werth, L. Atallah, P. Andriessen, X. Long, E. Zwartkruis-Pelgrim, and R. M. Aarts, "Unobtrusive sleep state measurements in preterm infants–a review," *Sleep medicine reviews*, vol. 32, pp. 109–122, 2017.

[145] Beddit. (2018) Beddit - sleep monitor. [Online]. Available: https://www.beddit.com/

[146] M. Kubat, D. Flotzinger, and G. Pfurtscheller, "Towards automated sleep classification in infants using symbolic and subsymbolic approaches," *Biomedizinische Technik/Biomedical Engineering*, vol. 38, no. 4, pp. 73–80, 1993.

[147] I. Koprinska, G. Pfurtscheller, and D. Flotzinger, "Sleep classification in infants by decision tree-based neural networks," *Artificial intelligence in Medicine*, vol. 8, no. 4, pp. 387–401, 1996.

[148] M. S. Scher, S. G. Dokianakis, M. Sun, D. A. Steppe, R. D. Guthrie, and R. J. Sclabassi, "Computer classification of sleep in preterm and full-term neonates at similar postconceptional term ages," *Sleep*, vol. 19, no. 1, pp. 18–25, 1996.

[149] C. Holzmann, C. Perez, C. Held, M. San Martín, F. Pizarro, J. Pérez, M. Garrido, and P. Peirano, "Expert-system classification of sleep/waking states in infants," *Medical & biological engineering & computing*, vol. 37, no. 4, pp. 466–476, 1999.

[150] J. Heiss, C. Held, P. Estevez, C. Perez, C. Holzmann, and J. Perez, "Classification of sleep stages in infants: a neuro fuzzy approach," *IEEE Engineering in Medicine and Biology Magazine*, vol. 21, no. 5, pp. 147–151, 2002.

[151] P. Estévez, C. Held, C. Holzmann, C. Perez, J. Pérez, J. Heiss, M. Garrido, and P. Peirano, "Polysomnographic pattern recognition for automated classification of sleep-waking states in infants," *Medical and Biological Engineering and Computing*, vol. 40, no. 1, pp. 105–113, 2002.

[152] C. M. Held, J. E. Heiss, P. A. Estévez, C. A. Perez, M. Garrido, C. Algarín, and P. Peirano, "Extracting fuzzy rules from polysomnographic recordings for infant sleep classification," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 10, pp. 1954–1962, 2006.

[153] A. Piryatinska, G. Terdik, W. A. Woyczynski, K. A. Loparo, M. S. Scher, and A. Zlotnik, "Automated detection of neonate EEG sleep stages," *Computer Methods and Programs in Biomedicine*, vol. 95, no. 1, pp. 31–46, 2009.

[154] V. Krajča, S. Petránek, J. Mohylová, K. Paul, V. Gerla, and L. Lhotská, "Modeling the microstructure of neonatal EEG sleep stages by temporal profiles," in *13th International Conference on Biomedical Engineering*. Springer, 2009, pp. 133–137.

[155] V. Gerla, K. Paul, L. Lhotska, and V. Krajca, "Multivariate analysis of full-term neonatal polysomnographic data," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 1, pp. 104–110, 2009.

[156] A. M. Mora, C. M. Fernandes, L. J. Herrera, P. A. Castillo, J. Merelo, F. Rojas, and A. C. Rosa, "Sleeping with ants, SVMs, multilayer perceptrons and SOMs," in *Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on*. IEEE, 2010, pp. 126–131.

[157] L. Fraiwan, K. Lweesy, N. Khasawneh, M. Fraiwan, H. Wenz, and H. Dickhaus, "Time frequency analysis for automated sleep stage identification in fullterm and preterm neonates," *Journal of medical systems*, vol. 35, no. 4, pp. 693–702, 2011.

[158] L. Fraiwan and K. Lweesy, "Newborn sleep stage identification using multiscale entropy," in *Biomedical Engineering (MECBME), 2014 Middle East Conference on*. IEEE, 2014, pp. 361–364.

[159] A. Dereymaeker, K. Pillay, J. Vervisch, S. Van Huffel, G. Naulaers, K. Jansen, and M. De Vos, "An automated quiet sleep detection approach in preterm infants as a gateway to assess brain maturation," *International journal of neural systems*, vol. 27, no. 06, p. 1750023, 2017.

[160] N. Koolen, L. Oberdorfer, Z. Rona, V. Giordano, T. Werther, K. Klebermass-Schrehof, N. Stevenson, and S. Vanhatalo, "Automated classification of neonatal sleep states using EEG," *Clinical Neurophysiology*, vol. 128, no. 6, pp. 1100–1108, 2017.

Appendix A

PSG and EEG infant sleep classifiers

Table A.1: Collection of infant polysomnography/ electroencephalography based automatic sleep classifiers.

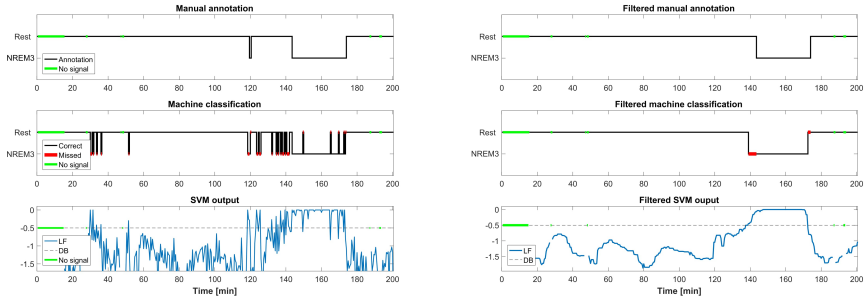
Year	Author	Signals	Method	N	Classes	Perf
1993	Kubat et al. [146]	EEG, EOG, ECG, EMG, resp., acto	LVQ+decision trees	N/A	awake, REM, N1, N2, N3/4, M, A	79-82%
1996	Koprinska et al. [147]	EEG,EOG, ECG,EMG, resp,acto	Tree-based NN	8	awake, REM, N1, N2, N3/4, M, A	75.3%
1996	Scher et al. [148]	EEG,EOG, ECG,EMG,resp	Discriminant analysis	54	awake,AS,QS	93.3%
1999	Holzmann et al. [149]	EEG,EOG,EMG	Expert system	N/A	awake,REM,N1, N2,N3/4	84.9-96.4%
2001	Heiss et al. [150]	EEG,EOG,EMG, limb movement	NeuroFuzzy Classifier	8	awake,REM,N1, N2,N3/4	83.9%
2002	Estevez et al. [151]	EEG,EOG,EMG, limb movement	Threshold	11	awake,REM,N1, N2,N3/4	87.7%
2006	Held et al. [152]	EEG,EOG,EMG, limb movement	Neuro-Fuzzy Classifier	14	awake,REM,N1, N2,N3/4	83.9%
2009	Piryatinska et al. [153]	EEG	K-means and heurestic rules	37	AS/QS	80-90%
2009	Krajca et al. [154]	EEG	K-means and heurestic rules	40	AS/QS	82.5%
2009	Gerla et al. [155]	EEG,EOG	Hidden Markov model	12	awake,AS,QS	82.5%
2010	Mora et al. [156]	EEG,EOG,EMG	Several	9	awake,REM,N1, N2,N3/4	70%
2011	Fraiwan et al. [157]	EEG	Neural network	29	awake,AS,QS	90-70%
2014	Fraiwan et al. [158]	EEG	Neural network, random forest, regression	27	awake,AS,QS	81.3-86.4%
2017	Dereymaeker et al.[159]	EEG	Cluster-based adaptive sleep staging	89	QS	N/A
2017	Koolen et al. [160]	EEG	SVM	67	AS/QS	83-85%

Appendix B

Postprocessed performances and hypnogram examples

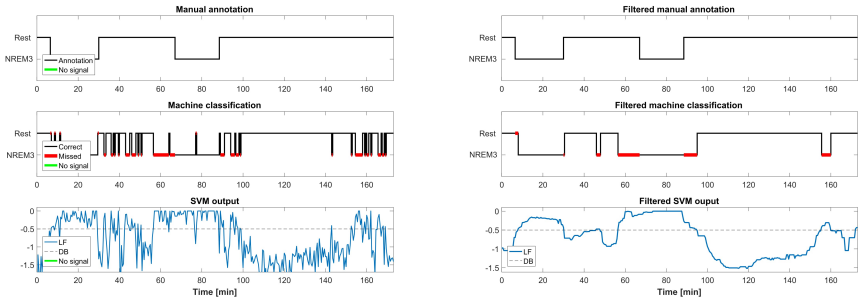
Table B.1: Postprocessed NREM3 vs rest classification results for each patient. **ACC** is total accuracy, whereas **ACC REM3** and **ACC Rest** denote class specific accuracies. The infant postnatal age (days) and the clinical observation are included. OSA - obstructive sleep apnea, CSA - central sleep apnea, hypov. - hypoventilation, WOB - work of breathing, inc. - increased and F - frequency.

ID	Age (d)	Diagnosis	ACC	ACC REM3	ACC Rest
1	56	Mild OSA	0.842	1.000	0.799
2	35	Mild CSA	0.809	0.281	0.967
3	73	OSA	0.858	0.987	0.833
4	29	Normal	0.832	0.384	1.000
5	125	Normal	0.787	0.790	0.785
6	13	Normal	0.635	0.962	0.496
7	42	Normal	0.545	0.973	0.456
8	35	Mild OSA	0.706	1.000	0.621
9	31	Normal	0.948	1.000	0.932
10	46	Mild OSA	0.963	0.956	0.965
11	10	Normal	0.568	0.993	0.352
12	12	Mild OSA	0.660	0.000	0.731
13	40	Normal	0.785	0.870	0.755
14	13	CSA+hypoventilation	0.979	0.900	1.000
15	70	Mild OSA	0.741	0.891	0.622
16	19	Normal	0.918	0.743	0.949
17	108	OSA	0.777	1.000	0.708
18	122	Normal	0.868	0.513	1.000
19	90	CSA	0.967	0.951	0.970
20	76	Normal	0.956	0.975	0.951
21	98	Normal	0.406	0.997	0.085
22	35	Severe CSA + mild hypoventilation	0.867	0.632	0.993
23	35	Normal	0.906	0.719	0.990
24	30	Normal - only very slightly increased WOB	0.812	0.970	0.768
25	25	Mildly increased WOB	0.556	1.000	0.327
26	35	OSA	0.730	1.000	0.697
27	118	Breathing F increased - and WOB mildly incerased	0.721	0.650	0.746
28	42	OSA	0.936	0.726	0.995
29	43	Mild OSA + increased WOB	0.779	1.000	0.709
30	81	OSA + increased WOB	0.841	0.784	0.709
31	127	Mildly increased WOB	0.677	0.964	0.609
32	16	Normal	0.727	1.000	0.643
33	78	Normal	0.844	0.967	0.802
34	82	Normal/mild OSA	0.910	0.942	0.901
35	27	OSA	0.922	0.760	0.993
36	66	OSA	0.978	0.963	0.982
37	10	Normal	0.934	1.000	0.898
38	54	Normal	0.911	1.000	0.887
39	44	Mildly increased WOB	0.764	0.882	0.726
40	24	Normal	0.755	1.000	0.674
41	19	Normal	0.676	1.000	0.608
42	16	Normal	0.977	0.923	0.989
43	14	OSA	0.587	1.000	0.451
44	24	Mild OSA	0.835	1.000	0.752
45	32	Mild OSA	0.654	1.000	0.565
46	59	OSA	0.795	0.978	0.693
47	21	Mild OSA	0.917	1.000	0.875
48	36	Normal	0.854	0.788	0.870



(a) Without postprocessing **(b)** Postprocessed results

Figure B.1: The effect of postprocessing on the patient #19. There is only one NREM3 session in this subject and the postprocessing fixed dubious epochs. Only some error happens in the state transitions.



(a) Without postprocessing **(b)** Postprocessed results

Figure B.2: The effect of postprocessing on the patient #33. There are two NREM3 periods, which are clearly detected. The postprocessing removes temporal instability. Two extra periods are wrongly classified as NREM3. Also there are error in the stage transition timing.