On the automated analysis of preterm infant sleep states from electrocardiography

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus prof.dr.ir. F.P.T. Baaijens, voor een commissie aangewezen door het College voor Promoties, in het openbaar te verdedigen op donderdag 7 november 2019 om 13:30 uur

door

Jan Valentin Sebastian Wendelin Werth

geboren te Mainz, Duitsland

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	prof.dr.ir. P.G.M.Baltus
1e promotor:	prof.dr. R.M. Aarts
copromotoren:	dr.ir. X. Long
	dr. P. Andriessen
leden:	prof.dr.ir. S.J.A. Van Huffel (KU Leuven)
	prof.dr.ir. L.M.G. Feijs
	prof.dr. S. Overeem
adviseurs:	dr. J. Weda (Orikami)
Reserve:	prof.dr. M. Mischi

Het onderzoek of ontwerp dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

On the automated analysis of preterm infant sleep states from electrocardiography

Jan Werth

If we aid a preterm infant, we aid a lifetime

Für meine Familie

Summary

   Sleep in preterm infants plays a vital part in neural development. The preterm infant sleep can be separated in mainly active sleep, quiet sleep, and wake. During the different sleep states, various developmental tasks are executed. These tasks range from the formation of new neural connections, the definition of neural regions, and the launch of developmental processes in stimulus-deprived regions during the more active sleep state, to error correction, neural region redefinition, and resting during the more quiet sleep state. Thereby, each state activates different parts of the autonomous nervous system resulting in physiological changes such as heartbeat ac-/deceleration, blood pressure variations, alteration in respiration, and electrical activity of the brain.

   These physiological changes can be measured using, for example, electrocardiography (ECG), electroencephalography (EEG), and electromyography (EMG). If we can monitor the preterm infant sleep states continuously using these measures, we can also monitor the preterm infant neural development, as the change of the sleep state distribution over time is linked to the course of neural development. Thereby, continuous neural monitoring might help to identify miss-development early on. However, in preterm infants, the central- and autonomous nervous system, which control the circadian oscillator and steer physiological responses, are not yet fully developed, resulting in not yet entrained state patterns and clear sleep state borders. Therefore, classifying sleep states in preterm infants is a more challenging task compared with that in adults. Further, preterm infants spend most of their time asleep to create a developmental supporting environment by shielding themselves from excessive external stimuli. Thus, preterm infants are awake for only a small portion of the day, which further increases the challenge to classify all states, including wake. To date, mainly polysomnography is applied to gain insight into the sleeping patterns of a preterm infant. Polysomnography pools several measurements, such as the already mentioned EEG, ECG, and EMG, which requires to apply multiple sensors. They

often present an additional burden on the fragile preterm infants due to electrode application on the infirm skin, and the amount of weight added by connecting cables. Therefore, the goal of this thesis is to investigate the use of only ECG analysis for sleep state classification. ECG is chosen as it is often already implemented as a standard, continuous measurement, so no additional sensors have to be introduced. We further examine in Chapter 2 the possibilities to measure ECG unobtrusively and review other methods to measure vital signs in preterm infants unobtrusively for sleep state analysis.

With the rise of machine learning, especially with the public releases of the Lua library Torch in 2002, python library SciKit-learn in 2012, and Googles TensorFlow in 2015 among others, complex signal and feature relation analyses have become manageable for tasks that contain a multitude of different parameters influencing each other. We investigated in Chapter 3 the possibilities of machine learning techniques, such as support vector machine, to classify the primary sleep states in preterm infants. As results were reasonable, we further used those methods to investigate in Chapter 4 whether machine learning would have a similar outcome when used in combination with capacitive ECG. Capacitive ECG uses capacitive sensors applied, e.g., in a mattress, to measure the ECG contactless. Using capacitive ECG would be a big step forward for in-hospital use as it is a measurement technique constituting no harm or additional burden to the preterm infant.

With the latest advancements in machine learning towards deep learning, particularly in time series analysis and implementation of concepts such as long short-term memory (LSTM) units and gated recurrent units (GRUs), sleep classification received another strong analysis tool as patterns of complex interactions within sleep can now be more easily identified over long time periods. We examined the use of LSTM and GRU architectures for preterm infant sleep analysis in Chapter 5. The results were similar compared to the classification with classic machine learning techniques, even though deep learning has much greater potential compared with classic machine learning in understanding highly complex and nonlinear cross-correlations between

features, because of the general difficulty of classifying the ambiguous preterm infant sleep states in combination with the low amount of training data. Gathering patient data of the preterm infant patient group is particularly difficult and constrained us from expanding the amount of data. In the final part of this thesis in Chapter 6, next to the used methods and overall results, we discuss ethical questions regarding continuous monitoring for preterm infants. We also give an outlook into the future of preterm infant sleep monitoring and give suggestions for further research directions.

In summary, we investigate the classification of preterm infant sleep based on ECG and capacitive ECG using classic machine learning techniques and recent LSTM and GRU deep learning architectures.

Nederlandse samenvatting

Slapen bij premature kinderen speelt een vitale rol in de neurale ontwikkeling. De vroeggeboorteslaap kan worden gescheiden in voornamelijk actieve slaap, stille slaap en wakker. Tijdens de verschillende slaaptoestanden worden verschillende ontwikkelingstaken uitgevoerd. Deze taken strekken zich uit van de vorming van nieuwe neurale verbindingen, de definitie van neurale gebieden, en de het starten van ontwikkelingsprocessen in stimulus-arme gebieden tijdens de actievere slaaptoestand, tot foutcorrectie, herdefinitie van neurale gebieden en rust tijdens de rustiger slaaptoestand. Daarbij activeert elke staat verschillende delen van het autonome zenuwstelsel, wat resulteert in fysiologische veranderingen zoals gehoorde acceleratie en afremming, bloeddrukschommelingen, veranderingen in de ademhaling en elektrische activiteit van de hersenen.

Deze fysiologische veranderingen kunnen worden gemeten met behulp van bijvoorbeeld elektrocardiografie (ECG), elektro-encefalografie (EEG) en elektromyografie (EMG). Als we de premature kinderslaaptoestand continu kunnen monitoren met behulp van deze maatregelen, kunnen we ook de premature neurale ontwikkeling van het kind volgen, aangezien de verandering van de slaaptoestandsverdeling in de tijd gekoppeld is aan het verloop van de neurale ontwikkeling. Daarbij kan continue neurale monitoring helpen om mis-ontwikkeling in een vroeg stadium op te sporen. Bij premature kinderen is het centrale en autonome zenuwstelsel, dat de circadiane oscillator controleert en de fysiologische reacties stuurt, echter nog niet volledig ontwikkeld, wat resulteert in nog niet ingesloten toestandspatronen en duidelijke grenzen van de slaaptoestand. Daarom is het classificeren van slaaptoestanden bij premature zuigelingen een meer uitdagende taak in vergelijking met dat bij volwassenen. Verder brengen premature kinderen het grootste deel van hun tijd door in slaap om een ontwikkelingsondersteunende omgeving te creëren door zich te beschermen tegen overmatige externe prikkels. Zo zijn te vroeg geboren baby's slechts een klein deel van de dag wakker, wat de uitdaging

om alle staten, inclusief het wakker worden, verder vergroot. Tot op heden wordt voornamelijk polysomnografie toegepast om inzicht te krijgen in de slaappatronen van een prematuur kind. Polysomnografie bundelt verschillende metingen, zoals het reeds genoemde EEG, ECG en EMG, waarbij meerdere sensoren moeten worden toegepast. Ze vormen vaak een extra belasting voor de kwetsbare premature kinderen door het aanbrengen van elektroden op de zwakke huid en de hoeveelheid gewicht die wordt toegevoegd door het aansluiten van kabels. Het doel van deze dissertatie is dan ook om het gebruik van alleen ECG-analyse voor de classificatie van de slaaptoestand te onderzoeken. Er wordt gekozen voor ECG, omdat het vaak al als een standaard, continue meting wordt toegepast, zodat er geen extra sensoren moeten worden ingevoerd. Verder onderzoeken we in hoofdstuk 2 de mogelijkheden om ECG onopvallend te meten en bekijken we andere methoden om de vitale functies bij premature kinderen onopvallend te meten voor analyse van de slaaptoestand.

Met de opkomst van machinaal leren, vooral met de publieke releases van de Lua bibliotheek Torch in 2002, de pythonbibliotheek SciKit-learn in 2012 en Google's TensorFlow in 2015 zijn complexe signaal- en kenmerkende relatieanalyses beheersbaar geworden voor taken die een veelheid aan verschillende parameters bevatten die elkaar beïnvloeden. In hoofdstuk 3 hebben we de mogelijkheden onderzocht van technieken voor machinaal leren, zoals een ondersteunende vectormachine, om de primaire slaaptoestand bij premature kinderen te classificeren. Aangezien de resultaten redelijk waren, hebben we deze methoden verder gebruikt om in hoofdstuk 4 te onderzoeken of machinaal leren in combinatie met capacitieve ECG's een vergelijkbaar resultaat zou hebben. Capacitieve ECG's gebruiken capacitieve sensoren die bijvoorbeeld in een matras worden toegepast om het ECG contactloos te meten. Het gebruik van capacitieve ECG's zou een grote stap voorwaarts zijn voor gebruik in het ziekenhuis, omdat het een meettechniek is die geen schade of extra belasting voor de premature zuigeling oplevert.

Met de laatste ontwikkelingen in het machinaal leren in de richting van diepgaand leren, met name in de analyse van tijdreeksen en de implementatie van concepten

zoals long short term memory eenheden (LSTM) en gated recurrent units (GRU's), heeft de slaapclassificatie een ander sterk analyse-instrument gekregen, aangezien patronen van complexe interacties in de slaap nu gemakkelijker kunnen worden geïdentificeerd over lange perioden. We onderzochten het gebruik van LSTM en GRU architecturen voor premature zuigelingsslaapanalyse in hoofdstuk 5. De resultaten waren vergelijkbaar met de classificatie met klassieke technieken voor machinaal leren, hoewel diepgaand leren veel meer potentieel heeft in vergelijking met klassiek machinaal leren om zeer complexe en niet-lineaire cross-correlaties tussen kenmerken te begrijpen, vanwege de algemene moeilijkheid van het classificeren van de dubbelzinnige premature kinderslaaptoestanden in combinatie met de geringe hoeveelheid trainingsgegevens. Het verzamelen van patiëntgegevens van de groep premature zuigelingen is bijzonder moeilijk en beperkt ons ertoe de hoeveelheid gegevens uit te breiden. In het laatste deel van dit proefschrift in hoofdstuk 6 bespreken we, naast de gebruikte methoden en de algemene resultaten, ethische vragen met betrekking tot continue monitoring voor premature kinderen. We geven ook een vooruitblik op de toekomst van het onderzoek.

Samenvattend onderzoeken we de classificatie van premature zuigelingsslaap op basis van ECG en capacitieve ECG's aan de hand van klassieke technieken voor machinaal leren en de meest recente LSTM- en GRU-dieplerende architecturen.

List of Abbreviations

ADASYN	Adaptive synthetic sampling approach
Adam	Adaptive moment estimation
aEEG	Amplitude integrated electroencephalography
aLL	Mean line length
AM	Antedependence models
ANN	Artificial neural networks
ANS	Autonomic nervous system
Ar	Activity regulation
AS	Active sleep
AUC	Area under the curve
BCG	Ballistocardiogram
BP	Blood pressure
BPD	Bronchopulmonary dysplasia
BpE	Beats per epoch
BR	Breathing rate
BW	Birth weight
cECG	Capacitive Electrocardiogram
CFS	Correlation based feature selection
CHIME	Collaborative home infant monitoring evaluation
CNN	Convolutional neural network
CNS	Central nervous system
CPAP	Continuous positive airway pressure
CTW	Caretaking + wake
DAQ	Data acquisition system
DT	Decision tree
ECG	Electrocardiogram
EDR	ECG derived respiration
EEG	Electroencephalogram
EMG	Electromyogram
EOG	Electrooculogram
ERF	Extra tree random forest
FLOPS	Floating-point operations per second
FSMC	Minority based feature selection method
GA	Gestational age
GB	Gradient boosting
GDP	Gross domestic product
GRU	Gated recurrent unit
HF	High frequency
HR	Heart rate
HRV	Heart rate variability
IBI	Inter-burst intervals

ICN	Intensive care nursery
IQR	Inter quartile range
IR	Infra-red
IS	Intermediate/undetermined sleep
IVH	Intraventricular hemorrhage
Kr	Kernel regulation
LDA	Linear discriminant analysis
LF	Low frequency
LL	Line length
LOOCV	Leave one out cross validation
LSTM	Long short-term memory
LVQ	Learning vector quantization
LZ	Lempel-Ziv complexity measure
MLP	Multi-layer perceptron
M-NICU	Mother neonatal intensive care units
MTLE	Mesial temporal lobe epilepsy
NEC	Necrotizing enterocolitis
NICU	Neonatal intensive care unit
NIDCAP	Neonatal individualized developmental care and assessment program
NIPU	Newborn intensive parenting unit
NIR	Near infra-red
NMCU	Neonatal medium care unit
NMDAR	Anti-N-methyl-d-aspartate receptor
NN	Normal-to-normal beat
NREM	Non rapid eye movement (sleep)
PCA	Postconceptional age
pDec	Percentage of heart rate decelerations
PMA	Postmenstrual age
PNS	Parasympathetic nervous system
PPG	Photoplethysmogram
PSG	Polysomnogram
PTSD	Post-traumatic stress disorder
QS	Quiet sleep
QSE	Quadratic sample entropy
RBF	Radial basis function
ReLu	Rectified linear unit
REM	Rapid eye movement
RF	Random forest
RNN	Recurrent neural networks
ROC	Receiver operating characteristic
ROP	Retinopathy of prematurity
RR	Respiratory rate
S	Sleep
SDaLL	Standard derivation of mean line length
SDDec	Magnitude of HR deceleration

SDLL	Standard derivation of line length
SE	Sample entropy
SEAUC	Sample entropy area under the curve
SER	Sample entropy over parameter range
SIDS	Sudden infant death syndrome
SMOTE	Synthetic minority over-sampling technique
SNR	Signal to noise ratio
SNS	Sympathetic nervous system
SVM	Support vector machine
TST	Total sleep time
VLF	Very low frequency
W	Wake
WHO	World health organization

Table of contents

Thesis outline

The thesis is composed of six chapters. In the **first chapter** (General introduction), we give a general overview of preterm infants and preterm infant sleep, including the different sleep states and their functions in preterm infant neural development. Next to the motivation for this thesis, we also state the research question, which this thesis tries to answer.

In the **second chapter** (Methods to obtain ECG signals unobtrusively in preterm infants), we will summarize different measurement techniques for preterm infant sleep, with a focus on unobtrusive methods to measure ECG signals, minimizing harm and disturbance for the preterm infants. Several methods are compared and rated based on their unobtrusiveness and signal recording capability.

The **third chapter** (Machine learning on preterm infant sleep) aims to classify AS and QS by using a nonlinear kernel support vector machine. Here classic ECG signals are used recorded with standard adhesive electrodes.

The **fourth chapter** (Use of unobtrusive ECG measurement for the use with machine learning) combines the ideas of automated sleep classification and unobtrusive signal acquisition. An ECG signal was used for machine learning, which was recorded by capacitive sensors embedded in a NICU incubator mattress. For machine learning, different algorithms were compared based on their sleep classification performance. The all-state classification was performed with a focus on AS and QS classification.

The **fifth chapter** (Deep learning approach for sleep state classification in preterm infants) examines different deep learning architectures to classify all preterm infant sleep states with different recurrent neural network topologies. The used features were again based on conventional ECG recordings.

The **sixth chapter** (Conclusions and future perspective) discusses the work of this thesis and answers the main research question raised in the introduction. Further, it

explores possible progress in the field of preterm infant sleep to encourage and support research, suggesting further possible directions and use cases.

1 General introduction

1.1 Motivation

Every year, around 15 million babies are born before 37 weeks of pregnancy [1]. Those early births may face a lifetime of medical issues, general complications, daily handicaps, and worries for the parents. Every bit of aid for such a fragile patient group can create a positive impact on a whole human life and its interconnected persons. As technicians, we might not be able to aid a patient directly, but we can provide the professional caretakers with the latest tools at hand, enabling them to focus their full attention, time, and energy on the most pressing tasks. The base for our research was the connection between sleep development and neural development. On this base, we were once confronted with the question why sleep should be automatically monitored regarding neural development, as the overall state of a preterm infant can be determined by a capable neonatologist or nurse via observations. The simple answer would be "knowledge is the key to aid." Only with continuous monitoring do we have the chance to reveal new patterns, hidden connections, and nonlinear dependencies, which are concealed to the human eye. Automatic, continuous sleep monitoring might at first only seem as a caretaking aid, but as the sleep states changes over time are closely linked to the neural development, sleep monitoring can indicate miss-development at an early stage where interventions have the greatest effect. Further, continuous information on sleep and development can lead to new research questions regarding ex- and internal influences on sleep. New caretaking-, structural-, or regulatory concepts, and their effect on sleep and development can be investigated much easier in large scale. Further, different treatment plans for miss-development can be explored with an immediate observation of interaction and effectiveness. To summarize, automated, continuous sleep and neural development monitoring can become the gateway to next-level preterm infant development care.

1.2   Definitions of preterm infants and age determination

A baby is defined as a preterm infant when it is born alive before 37 weeks of pregnancy. Preterm infancy can be divided into three subcategories based on gestational age (GA). *Extremely* preterm are infants born with less than 28 weeks of pregnancy. *Very* preterm are infants born from 28 weeks to 32 weeks of pregnancy. *Moderate* and *late* preterm infants are born from 32 weeks to 37 weeks of pregnancy [1]. Term babies are infants born 37 weeks to 42 weeks of pregnancy.

In this thesis, age determinations are used as described by the American Academy of Pediatrics (Figure 1) [2].

The two main used definitions are as follows: first, the GA, which is the age from the last gestation to the time of birth; and second, the postmenstrual age (PMA) or postconceptional age (PCA), which are both the duration until the date of assessment. PMA starts with the last gestation, whereas PCA starts with conception and is calculated with two additional weeks.
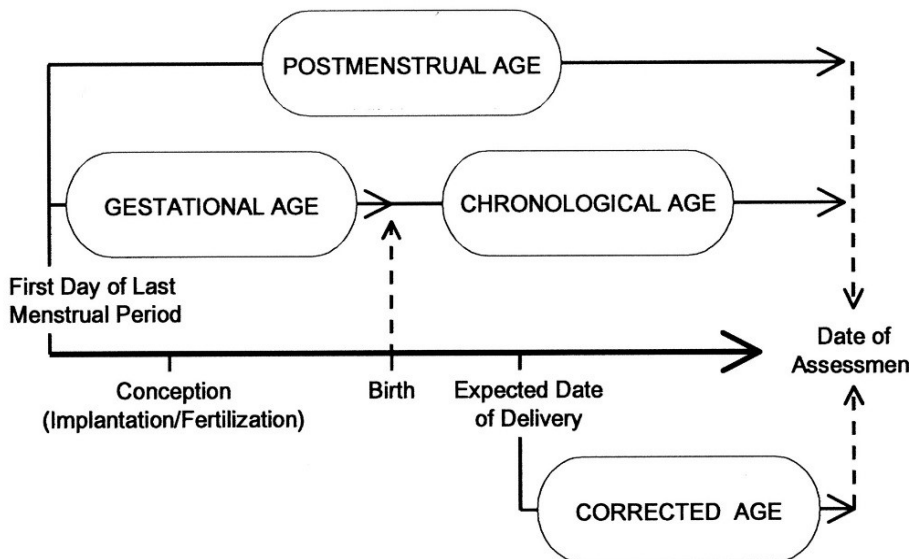
Figure 1 Age terminology during the perinatal period. From [2].

Another useful classification of preterm infants is by weight. Small for gestational age defines preterm infants below 2500 g (10th percentile for their GA). Large for gestational age are preterm infants heavier than 4500 g (90th percentile for their GA). Low birth weight is further categorized into very low birth weight and extremely low birth weight with less than 1500 and 1000 g, respectively [1], [3].

1.3   Neonatal intensive care unit

Preterm infants are stationed, if existing, at the neonatal intensive care unit (NICU) [4], intensive care nursery (ICN), neonatal medium care unit (NMCU), newborn intensive parenting unit (NIPU) [5], or, similar, mother neonatal intensive care units (M-NICU) [6] to mitigate or ideally prevent as many short- and long-term health risks in this particular fragile patient group as possible. The NICU was in the care of preterm infants with specific equipment and trained staff to the needs of preterm infant care. The preterm infants are placed in an incubator to shield them from the hostile environment outside the womb. The incubator can be regarded as a mimic of the womb to give the preterm infant the best possible environment to continue the disrupted development process. Incubators can range from a simple "box" to an elaborate shielded and monitored environment. The latest equipment for high-end incubators, to name only a few, includes built-in blue light therapy for jaundice, automatic temperature control, shielding airflow to prevent temperature loss with opened incubator doors, positive air pressure to prevent infection intrusion, video monitoring with streaming function to the parents, and detachable monitoring systems for easy and fast patient transfer without loss of vital sign monitoring. Those high-end incubators are mainly affordable in high-income countries. In addition to the latest equipment, modern NICUs also focus on developmental monitoring. Given an increased survival rate over the years, modern NICUs are changing their care from a mostly survival centric toward a survival and developmental centered care. The Neonatal Individualized Developmental Care and Assessment Program (NIDCAP) was introduced in 1986. High technological and medical standards, in combination

with developmental centered programs, result in different mortality and morbidity rates among low-, mid-, and high-income countries. The preterm mortality risk is positively linked to lower per capita gross domestic product (GDP) [7]–[9].

1.4   Statistics on preterm infants

Preterm infants are a very specific, highly critical, and small patient group in comparison with adult patients. Preterm care often receives lower to no budgets depending on the general economic situation of a country and the overall standard of care. As a result, preterm infant birth rates, mortality rates, and morbidity rates diverge among high-, mid-, and low-income countries, correlating positive with the GDP [7]–[9].

The World Health Organization (WHO) states that more than 60% of preterm births occur in Africa and South Asia [9]–[11]. Those individual countries with over 250,000 preterm infant births per year (measured in 2010) are India, China, Nigeria, Pakistan, Indonesia, USA, Bangladesh, the Philippines, Dem. Rep. of Congo, and Brazil. Among those, the countries where over 15% of births are preterm are mainly found in the continent Africa and the countries Pakistan and Indonesia [11]. The differences become more prominent when the survival rate is considered. In high-income countries, almost 95% of preterm born between 28 and 32 weeks GA survive. Around 90% of those preterm infants survive without impairment. By contrast, only 30% of the same preterm infants survive in low-income countries where almost all preterm infants born below 28 weeks GA die in the first few days after birth. Almost all those born at <28 weeks do not survive in the first few days of life [9]. Besides differences among countries, socioeconomic factors of patients also play a role in preterm infant birth rates. For example, in 2009, the USA reported up to 17.5% of preterm births among black citizens compared with 10.9% in white citizens [12].

Nevertheless, all preterm infants around the globe have to fight the same struggles and are confronted with the same health problems. In the next section, we give an overview of preterm infant health and long-term outcome.

1.5   Health and outcome in preterm infants

The development of the lungs is not finished in the last trimester of pregnancy and continues into the postnatal period. In particular, the surfactant system and the development of the true alveoli are incomplete. As preterm infants are born too early, respiratory system problems and diseases are at 73.8% most common in this patient group [13]. Respiratory system diseases can be respiratory distress syndrome, pneumothorax, apnea, bronchopulmonary dysplasia (BPD), and pneumonia (even though pneumonia is caused by an infection). They can lead to respiratory symptoms during infancy. About 51%–97% of preterm infants with a respiratory system disease report coughs, and 39%–45% report wheezes [14]–[16]. Re-hospitalization in the first two years was required in 14%–38% of the cases, whereas long-term inhalation therapy was necessary in 13% of the cases [16]–[18]. As the immune system is also developing during the last trimester of pregnancy, preterm infants are highly vulnerable to infections [19]. With 39.4%, infectious diseases are the second common neonatal disease. Late-onset infections are quite common, occurring in about 20% of very low birth weight infants [20]. The early-onset of systemic infections in very low birth weight infants has a mortality of close to 40% in high-income countries, and the same group without early-onset infections has only a 13.3% mortality rate. In addition to an increased mortality rate, early-onset infections can lead to neurodevelopmental impairments such as cerebral palsy, visual- and hearing impairment [20]. The third most common diseases in preterm infants are nervous system diseases (38.3%) such as peri-/intraventricular hemorrhages and periventricular leukomalacia [13]. Both are highly likely in preterm infants compared with full-term babies. However, (well)-developing neural areas make them less likely to occur [21]. The autonomic nervous system is generally very immature, resulting in overall physiological complications, for

example, desaturation, bradycardia, and apnea. A recent study by Joshi et al. [22] analyzed a total of 8,522 patient days to find the distribution of physiological complications displayed in Table 1.

Table 1 Distribution of physiological complications in preterm infants [22].

Desaturation	66.5%
Bradycardia	21.8%
Apnea	4.5%
Asystole	2.0%
High arterial blood pressure	1.8%
Ventricular fibrillation/tachycardia	1.5%
Low arterial blood pressure	1.4%
Tachycardia	0.4%

They later found that part of the alarms, marking the physiological complications, was tied to the enteral feeding routine of the trial unit and resulted in advice to feed infants in a lateral position by the gravity feeding method [23].

Despite all the technological advances in the field of pediatrics, all these complications can lead to a high mortality rate for extremely premature infants. The mortality rate in developed countries lies at 50% for preterm infants born at 23–24 weeks GA and 34 weeks GA in low- and middle-income countries. The chance of survival drastically rises with GA. With 28 weeks GA, over 90% of preterm infants survived [24]–[26] (see Figure 2). In survivors of such early preterm infants, the risk of morbidity lies at 20%–50%. The morbidity is closely connected with the GA of the preterm infant. Stoll et al. [26] reported the rates for surviving without morbidity, as shown in Figure 3. The survival rate per GA has been increased due to modern technology, especially for extreme preterm infants around 24 weeks GA. The number of acute and chronic morbidities also increases as more early preterm infants survive. Glass et al. [27] pooled data from different trials presenting an overview of acute and chronic morbidities (presented in Table 2).
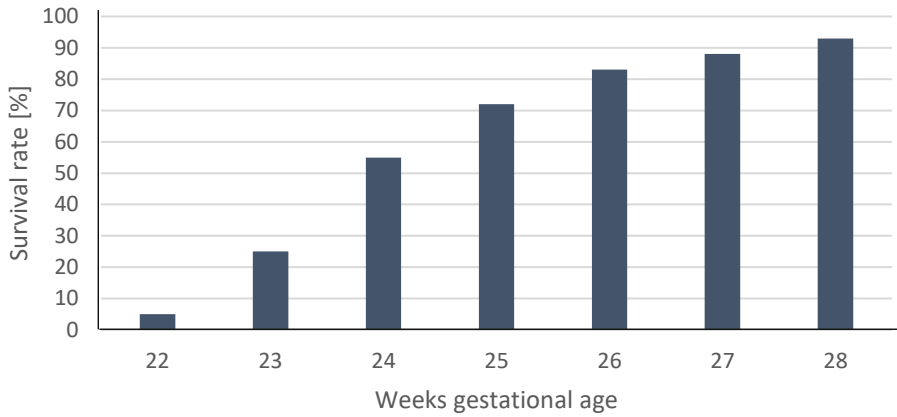
Figure 2 Rate of survival in percent over gestational age in developed countries. Adapted from [26].
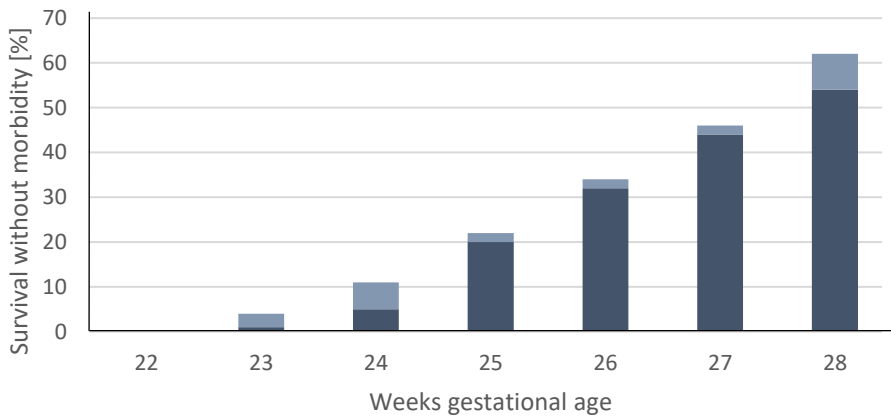
Figure 3 Mean percentage of survival without morbidity over gestational age (pos. std. light grey). Adapted from [26].

Table 2 Acute and chronic morbidity Risks. From [27].

Acute/NICU	Percent
Any retinopathy of prematurity (ROP)	63.7
Severe ROP	12.3
Intraventricular hemorrhage (IVH)	14.1
Surgical necrotizing enterocolitis (NEC)	10.1
Chronic Problems	
Bronchopulmonary dysplasia (BPD)	42.2
Blindness	0.8
Hearing loss	3.1
Cerebral Palsy	6.1
Cognitive Delay	5.2

1.5.1 Chronic respiratory system disease - bronchopulmonary dysplasia

As respiratory system diseases are the number one occurrence in preterm infancy, long-term morbidities are most commonly lung related. In particular, BPD is a frequent issue. The frequency of BPD is related to the continuous development of the respiratory system during the last trimester. Injuries or infections during that period often initiate BPD. Injuries and infections are often caused by ventilation and/or intubation during preterm infancy, which is necessary as the respiratory system is not yet fully developed. BPD results in airway inflammation, fibrosis with severe epithelial injury, smooth muscle hyperplasia, abnormal vascular growth, reduced alveoli count, and fibroproliferation. Unfortunately, a preterm infant with severe BPD survives the first three years of life only in 5% of the cases [27].

1.5.2 Neurological development, complications, and long-term outcome

This thesis focuses on preterm infant sleep classification methods, which could be used for continuous sleep and ensuring neural development monitoring. To stress the importance of neural development monitoring and to understand the fragile construct and the high potential for aberrations and defects, the following sections describe the neural development in the last trimester of pregnancy.

In the last trimester of pregnancy, four main neural developmental phases can be identified:

- neural migration,
- dendrite and axon sprouting,
- synaptogenesis,
- apoptosis.

After millions of neurons and glial cells are produced (neural proliferation), the neurons migrate to their designated destination during the neural migration phase. The neurons are pushed from the core to the outer hemisphere by newly generated neurons in the core, or they migrate along the glial cells that function during that phase as a guiding system. If something goes wrong during the migration phase, indications will appear immediately after birth. The most articulated signs of migration disorder are seizures. A more rare condition is a semi- or non-developed corpus callosum [28].

After the neurons reach their destination, they start to sprout dendrites and axons to build up a neural network. For this network, the axons need a connector, which is the synapse. The first synapses can be found at 8 weeks GA in the spinal cord. After 28 weeks GA, synapse generation starts to boom in an episode called synaptogenesis, which lasts until around three months post-term. Synaptogenesis and synapse elimination are the bases for lifelong brain plasticity [19], [28], [29].

Apoptosis refers to programmed cell death. This cell death is initiated by the cells and is part of cell metabolism. In the time of brain development before birth, apoptosis is important to ensure the right connection between the developing brain regions and the single nerve cells. In the earlier processes of neuronal proliferation and migration, a large number of neurons flood the developing brain, giving it high plasticity. About half of the neurons are eliminated before maturation to remove the overhead. The process of apoptosis reduces brain plasticity but also stabilizes it [29]. Failure in cell death can have a significant impact on brain development. Unintended

surviving cells can cause many different defects, such as glioma, cranial defects, and encephalocele [30]. Cell death peaks in the last trimester in the globus pallidus between weeks 26 and 33 weeks GA [28].

During the last trimester, the brain volume increases by two-thirds of its volume at term age (see Figure 4). At around 32 weeks GA, 50% of the volume is reached [31]. The myelinated white matter increases its volume by 500%, the cortical grey matter increases up to 50%, and the cerebellar volume increases its volume by 25% [32]. During this time, the brain is at high risk of injuries and interrupted development. Apart from the former stated direct neurological impairments linked to errors during certain development phases, preterm infants with reduced total brain volume show reduced naming vocabulary, word reading, and overall poor language skills; reduced picture memory; lowered pattern construction; and generally low Bracken school readiness assessment in later life [32], [33]. Moreover, reduced hippocampal volume is linked to a low score on the Bayley Scales of Infant Development [32], which
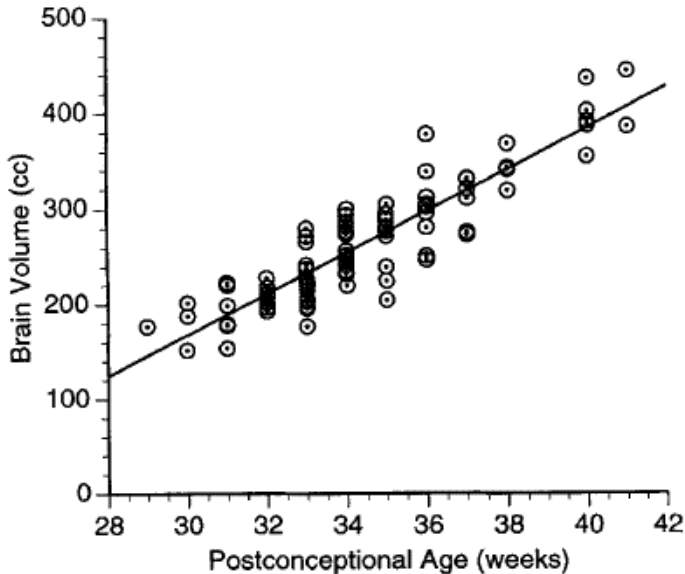
Figure 4 Brain volume growth over gestational age. With permission from [31].

quantifies developmental outcome during the first three years of life [34]. Disturbed connectivity in the neural network linked to cognitive deficiency can lead to anxiety, hyperactivity, and general social and behavioral problems [35]. Volpe et al. [29] stated that injury or neural misdevelopment during preterm infancy constitutes neuropsychological abnormalities, which can last into late childhood, adolescence, and adulthood. Volpe's statement is recognized by the WHO in the "born to soon" report [9]–[11] by stating that preterm babies born at 34–36 weeks GA have the most significant public health impact due to their large number.

Given that neural development plays such a significant role in the long-term outcome of a preterm infant, we want to examine the connection between sleep and neural development.

1.6   Sleep and neural development

Sleep is one of the most critical factors for the neural development of preterm infants [36]. In preterm infants, four different states are commonly distinguished: active sleep (AS), quiet sleep (QS), intermediate/transitional sleep (IS), arousal, and wake (W) [37], [38].

During the first month after birth, preterm infants spend up to 70% of their time sleeping, whereas term babies spend around 60% of their time asleep [39]. The change in total sleep time (TST) with neonatal development is accompanied by a change in sleep organization. As a result, the length of AS decreases while QS increases [39]– [45]. This change in sleep state organization is a strong representation of the maturation and consolidation process. Differentiation of the single sleep states and determination of their share in the TST indicate whether the developmental process follows the anticipated path or not [45]. These changes in sleep states over time are shown in Figure 5.

Interestingly, some work of Curzi Dascalova et al. showed slightly contrary results to the generally accepted trends with increasing AS and QS in an early study [40], and
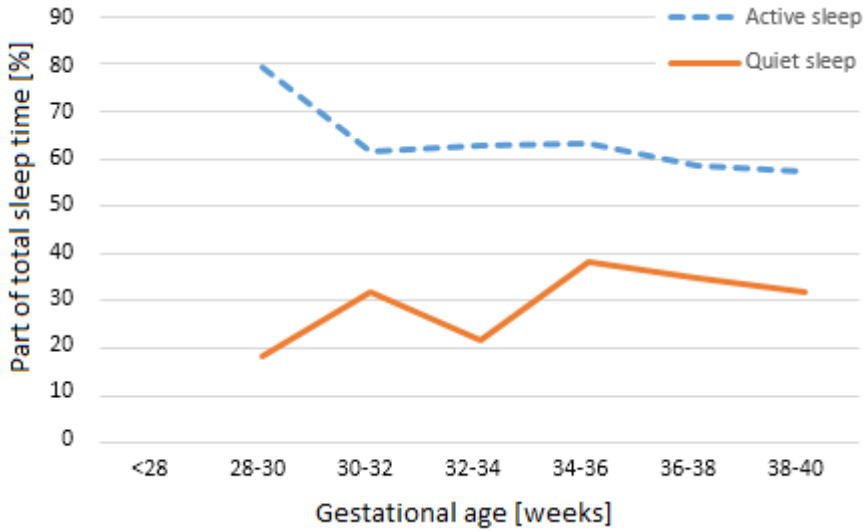
Figure 5 Sleep state development over time.
The graphic shows active sleep and quiet sleep over gestational age. The active sleep is decreasing while quiet sleep is increasing. The data are compiled from [37], [42]–[44], [46]–[49].

increasing AS and decreasing QS in a later study [41]. The diverse results may result from different annotation methodologies, definitions, and/or signal qualities, which they acknowledge in their publication [40]. Moreover, the separation of sleep states at a very early age of 27 weeks GA is known to be extremely difficult, which is reflected in a high standard derivation of their results in the later study [41].

The IS state is used for the transitions between states. During IS the sleep state cannot be identified explicitly as AS, QS, or wake, as the boundaries are blurred. Even though it is a blend of features from AS and QS, it is seen as an independent mode of the CNS functioning [50]. The IS state is stable and well organized in healthy preterm infants. The duration differs between AS-QS and QS-AS transitions with a longer duration of AS-QS (~5 min) than QS-AS (~2 min) [50]. Given that the differentiation of AS and QS improves with maturation, the time spent in IS rapidly decreases from 34 weeks GA onwards [51], [52].
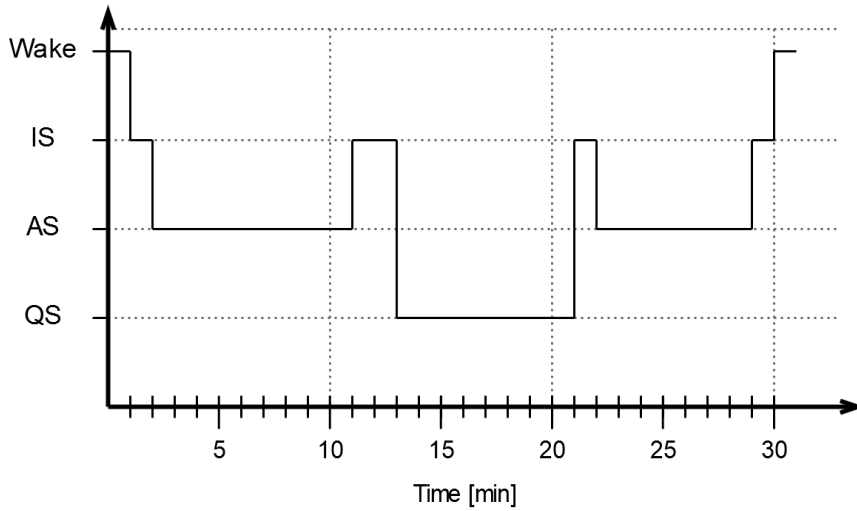
Figure 6 Example extract of one sleep cycle hypnogram of an early preterm infant (around 30 weeks GA) with a sleep cycle duration of 30 min. The total duration gradually increases to about 75 min until term age.

In the preterm infants, a sleep cycle is defined from either wake or QS to the same state reappearing. The duration of sleep cycles appears stable at 30–70 min. With development, the mean sleep cycle duration progressively increases to about 75 min [53]. Early preterm infants can show erratic and unstable sleep cycles [54], which we also observed in few recordings. From experience, each AS and QS state can last approximately 20 min in stable preterm infants of around >32 weeks GA. The preterm sleep cycle follows the pattern wake-AS-QS-AS-wake [38] (see Figure 6).

1.6.1   Importance of active sleep for brain development

Observations indicate that AS triggers the developmental process [55]. Given a lack of external impulses, the brain activates selectively several neural processes during AS, resulting in activity patterns in neural populations who need activation to develop [56]. AS functions as a brain development ignition [55]. In general, neural activity has been shown to increase during AS [52], [55]. In addition, the preterm REM phases during AS indicate central neural activity in the brain [45]. Those REM phases in AS

are mainly detected directly after transition to QS or from QS [50].

To further explore the impact of AS on brain development experimentally, several groups have tested the outcome of REM (AS) deprivation in neonatal and premature animals [56]–[60]. Lopez et al. [57] demonstrated the mentioned endogenous stimulation on mammals, and deprivation was found to lead to a destabilization of the neural circuits during development. This phenomenon also delays the development of synaptic plasticity [57] and long-term potentiation, which is linked to the refinement of neural circuits [57], [61]. Newborn REM-deprived mammals from the experiments of Mirmiran et al. [59], [60] showed altered patterns of fight-or-flight, social, and sexual behavior in later life. Furthermore, Marks et al. [56] showed that the operational capabilities of brain areas and their functions (e.g., the visual cortex) are less developed with the lack of REM sleep. Sleep itself is altered consequently from a REM deprivation by an increase in basal forebrain adenosine, thereby inhibiting the release of neurotransmitters, including serotonin [62].

1.6.2   Importance of quiet (NREM) sleep for brain development

Only a few studies have analyzed QS in detail [52], [56], [63]. This limitation might be due to the low share of QS compared with AS in the early life of a preterm infant or the minimized neural activities during QS [56] compared with AS, making AS more suited for the most common sleep analysis method using EEG. Nonetheless, QS is also essential for development as described by Peirano et al. [52], who pointed out that the increase in QS over time corresponds in animals with thalamocortical and intracortical innervation, as well as heightened synaptogenesis. Furthermore, QS is associated with pre- and postsynaptic region remodeling [52], which is critical for brain development and neural plasticity [64]. They added that visual cortex plasticity is dominant in animals during QS if monocular deprivation requires visual cortex adjustment. This phenomenon illustrates the capability of QS to adjust for failures during development [52]. Further, it has been shown that habituation to audio stimulation was found in both AS and QS states but was achieved faster during QS

[65]. Considering the known dominant parasympathetic nerve activity during QS with lowered blood pressure (BP) and heart rate (HR) [66], another task of QS is to consolidate and reenergize the system of the preterm infant, as the parasympathetic nervous system is known for its resting function by reducing sensory activation and response to external stimulation [67]. This hypothesis is underlined by the "sleep stress" phenomena, which is a change in the regular neonatal wake-AS-(IS)-QS sleep state sequence into a wake-QS pattern when faced with excessive stimulation to achieve a stimulus withdrawal. The preterm infant skips AS and changes directly into QS after wake as a response adaption to external stimuli to preserve energy by reducing responsiveness [65].

In the following sections, we demonstrate how many influencing factors are interconnected and have to be considered when we try to classify sleep. Therefore, we look into the interconnection between sleep and the cardio-respiratory system. All these systems are linked via the autonomous nervous system.

1.6.3   Adult and preterm sleep states

While focusing mainly on preterm infant sleep, a brief look into the comparability of adult and preterm infant sleep is of interest. In comparison to the four preterm infant states, including IS and wake, adults sleep is segmented into five states with three deep sleep states, one REM sleep state, and the wake state [68]. At first glance, the sleep states seem comparable despite the different sum of states. The impression of comparability is based on the seeming summation of the adult deep sleep states in the preterm infant QS state, and the adult REM sleep in the preterm infant AS state. Preterm infant AS even shows episodes of REM [69] and the muscle tone known in adults [46]. In addition, the naming convention is sometimes similar using REM and N-REM in preterm sleep description.

Nevertheless, there are some differences. The first difference is the time duration and cycle of sleep states. Generally, preterm infants spend about 70% in 24 h sleeping

while adults only 16-29%. The states themselves are also shifted, with adults only spending 20-25% of their sleep in REM [70], while preterm infants between 80% and 60% (Figure 5). However, more important, there are also small differences in the brain activity itself, represented in the different EEG signals. The waves and their frequency ranges are similar but with a larger spread in preterm infants and lower mean voltage for the lower borders [71]. Further, sleep spindles, which appear to be important for memory consolidation [72], are not apparent before 42 weeks GA [73]. Another difference is the unique pattern of delta brushes in preterm infants. Delta brushes are transient patterns characterized by a slow delta wave with superimposed fast activity. The occurrence, amplitude, frequency, and occurrence during specific sleep states change over the course of development. In preterm infants, delta brushes seem to be a response to audio-visual stimuli and stroking [74]. Delta brushes can appear in adults but are an indication of patients with either anti-N-methyl-d-aspartate receptor (NMDAR) encephalitis, intensive care unit patients with high mortality rate, or, in rare cases, mesial temporal lobe epilepsy (MTLE) [75]. The missing sleep spindles and unique delta brushes indicate that the function of sleep in preterm infants and adults are still different in some core functions.

Of course, sleep from preterm births is the rudimentary basis from which adult sleep develops. Therefore, many functions in preterm infant sleep constitute as preliminary stages of the later functions in adult sleep. This manifests in some operational similarities; for example, Diekelmann et al. [76] describe that during adult REM sleep, the brain plasticity is increased by a local upregulation of immediate-early-genes and reduced electroencephalographic coherence between brain regions. This seems to effectively support local synaptic consolidation. Within preterm sleep, synaptic consolidation takes place mainly during QS, where also the brain plasticity is heightened. Apparently, the functionality is very similar but shifted from QS in preterm infant sleep to REM sleep in adults. Further, during preterm infant AS, synaptic interconnections are created while in adults, those are consolidated, which could be seen as a morphed/developed functionality. Further described in the Nature

review [76], during adult deep sleep system consolidation promotes the reactivation and redistribution of selected memory traces for long-term storage. This can be compared to the region remodeling and adjustment during QS, posing similar core functionality [52]. Stronger comparability between QS and adult deep sleep can be seen in the role of error processing. As described in chapter 1.6.2, QS adjusts for failures during development [52] and change of sleep pattern in favor of QS to preserve energy after stressful events [65]. For adults, it has been shown that deep sleep may help in the recovery from brain injury. Increased deep sleep has shown beneficial effects in stroke, traumatic brain injury, and Alzheimer's disease patients [77].

A more philosophical approach to AS and memory consolidation is, to ask whether the neural activation in stimuli deprived brain areas [55] could be considered as artificial memories, or whether we factor them as real memories of ad hoc experiences of stimuli. We cannot answer this question, but maybe these represent the preterm infant versions of memories to prepare for the following memory consolidation task of adult REM sleep.

1.6.4   Interconnection between sleep and the cardio-respiratory system

Sleep in the form of AS and QS influences physiological events via the autonomic nervous system (ANS). The sleep itself is partly controlled via the central nervous system (CNS) circadian oscillator function, the suprachiasmatic hypothalamic nucleus (SCN). Both ANS and CNS are linked in cognitive and autonomic modulations with complex interactions summarized in a central autonomic network model [78], [79]. As the SCN circadian clock first develops in humans from 38 weeks GA [80], [81], and the signals we use for sleep classification are mainly influenced via the ANS, we will only summarize the ANS steered physiological interconnection of the cardio-respiratory system (Figure 14). As the interconnections are multiple, multi-dependent,
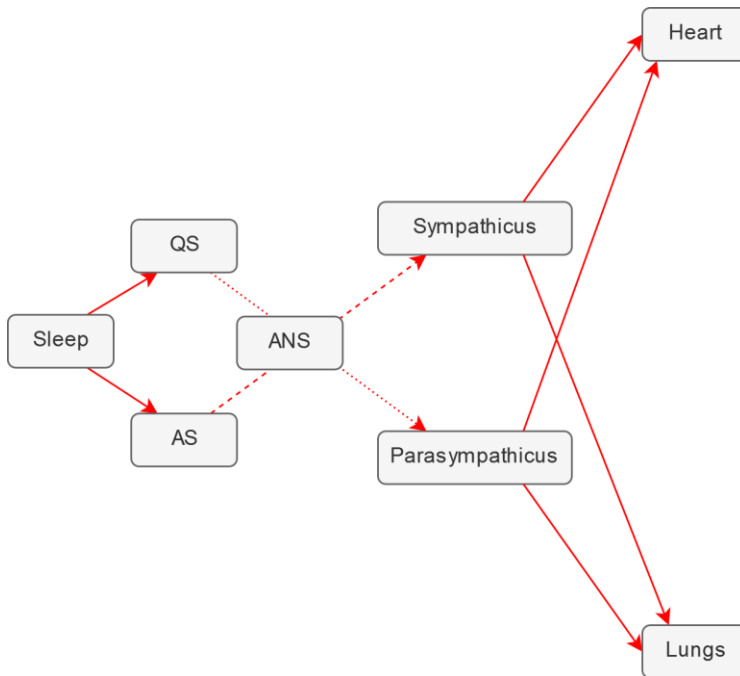
Figure 7 The autonomous nervous system has a direct influence on the heart and lung activities via the para- and sympathetic nervous system. The different dotted lines resemble the connection between QS and the PNS, and AS and the SNS.

and highly complex, we do not claim to provide a complete overview. This section is merely an overview, highlighting the many and non-trivial influencing factors on preterm infant sleep states or features that require equivalent complex methods and algorithms to classify the sleep states based on cardiac signals.

As a basic principle, the QS activates the parasympathetic nervous system (PNS), whereas the AS activates the sympathetic nervous system (SNS). The SNS is a quick response and mobilizing system, whereas the PNS is a more gradually activated dampening system. Both the SNS and PNS activate the lung functions, whereas the SNS enhances the lungs and dilates the bronchioles of the lung through circulating epinephrine, which allows for greater alveolar oxygen exchange. The PNS, in return, constricts the bronchiolar diameter when the need for oxygen has diminished [82]. Both systems have a direct effect on the heart and lungs (Figure 7). The SNS increases
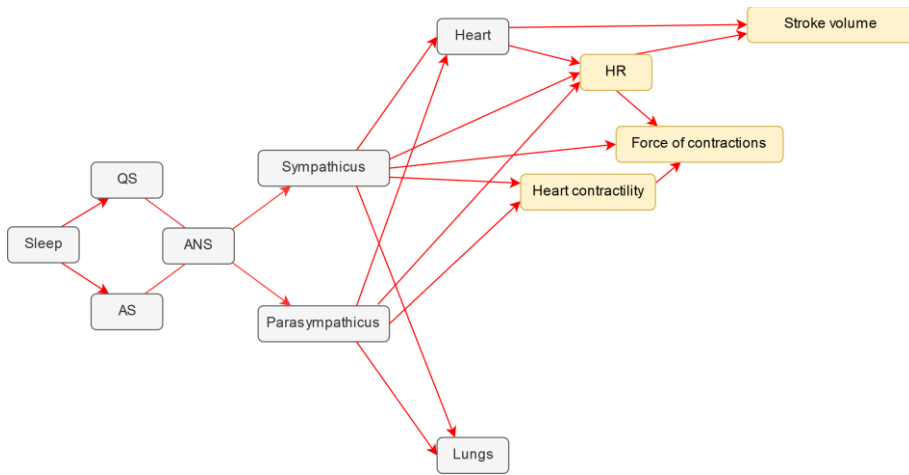
Figure 8 Stroke volume and force of contraction are influenced by cardiac contractility, heart rate, and the heart itself. These connections are marked yellow.

the HR and heart contractility of cardiac cells to enhance blood flow to the skeletal cells and provides vasodilation (widening of blood vessels) for the coronary vessels of the heart. Cardiac contractility changes the ability to produce the force during contraction [83]. The PNS increases the resting potential and decreases the rate of diastolic depolarization; under these circumstances, the HR slows down, and cardiac contractility decreases. The force of contraction, developed by the heart muscle, also depends on the frequency at which the muscle is stimulated (HR). As the stimulus frequency is increased, the force is increased as well until the maximum is reached; here, the force of contraction begins to decrease. An increase in the level of circulating epinephrine and norepinephrine from the SNS also increases the force of contraction [83]. Therefore, the stroke volume is directly related to the heart by the heart volume, and by the time the heart has to fill the chambers. With increased cardiac contractility, the possible force of contraction increases (Figure 8).
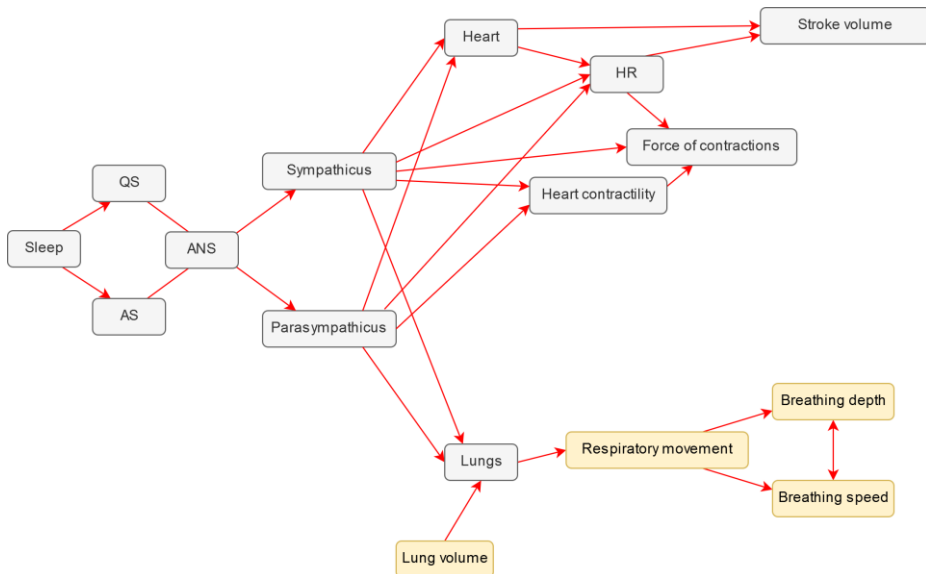
Figure 9 The lung regulates over the respiratory movement the depth and speed of breathing. With a larger volume, the respiratory movement is altered. These connections are marked yellow.

As mentioned before, the CNS influences breathing by enhancing the lungs and dilating the bronchioles, thereby affecting the breathing depth and breathing rate. With greater lung volume, breathing can be deeper and slower. Greater oxygen exchange by dilated bronchioles can decrease the breathing rate. The breathing rate and breathing depth also influence each other. Less depth increases the rate and vice versa to keep the oxygen saturation on the same level with no changed surrounding parameters (e.g., behavior). The overall lung volume changes with age and influences the maximum breathing depth and its rate (Figure 9).

The breathing depth influences the R peak amplitude (normal amplitude) of the HR due to the sinus arrhythmia. The force of contraction and the stroke volume affect the normal amplitude; the higher the force of contraction and the higher the stroke volume, the higher the normal amplitude. Although the force of contraction and the stroke volume depend on each other to influence the normal amplitude, as the force
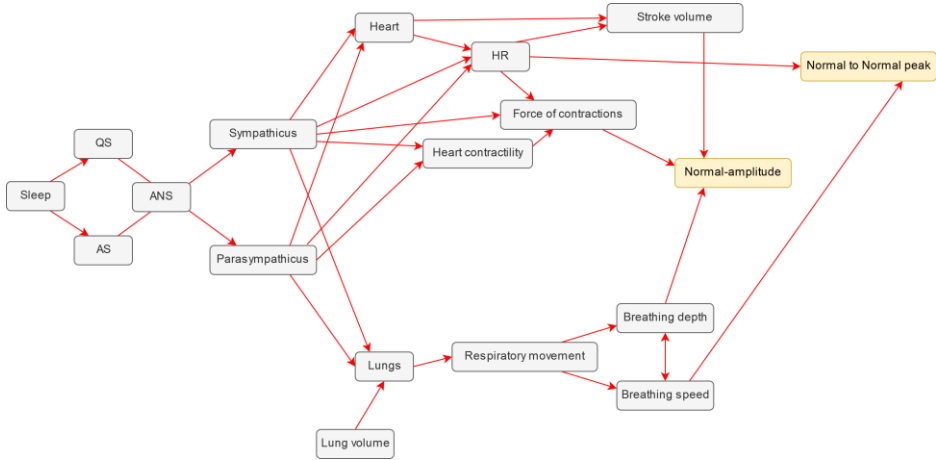
Figure 10 The Normal to Normal peak is influenced by the heart rate and speed of breathing. The Normal amplitude is tied to the force of contraction, stroke volume, and breathing depth. These connections are marked yellow.

of contraction and the stroke volume are small in preterm infants, those effects are less noticeable in this patient group. The distance of the R peaks (normal to normal peak) is set by the HR. However, the breathing speed influences the normal to normal peaks as faster breathing shortens the time between normal to normal peaks [84] (Figure 10).

The normal to normal peaks can be convolved into the HR variability (HRV). Respiratory events such as sighs, apnea, or periodic breathing influence the HR. Moreover, generally slow rhythmical fluctuations within a period of 30 heartbeats or around 10s can be observed in many preterm RR series. Sucking and thermoregulation may cause this 10s rhythm. Similarly, cardiac arrhythmia (premature beats) creates short HR deceleration (spikes) and missed R peak detection, changing the HRV signal [85]. Furthermore, the stroke volume is influenced by the in- and exhaling where stroke volume and cardiac output decrease with inspiration. This reduction is also transferred via changes in the intrathoracic pressure. The large negative intrathoracic pressure increases the pressure across the wall of the left ventricle. This pressure gradient causes an increase in afterload. This results in a
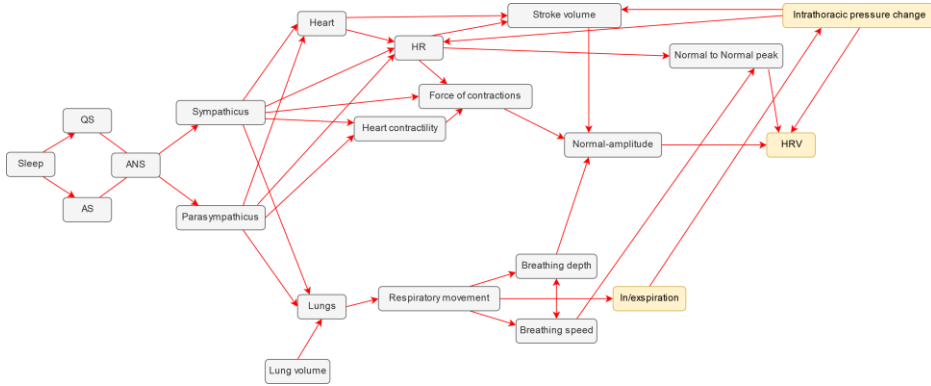
Figure 11 Heart rate variability is directly tied to the Normal to Normal peaks and amplitude. Inspiration/expiration influences the intrathoracic pressure change, which influences the heart rate variability, heart rate, and stroke volume. These connections are marked yellow.

decrease in stroke volume and an increased HR [86] (Figure 11). Inspiration/expiration via intrathoracic pressure changes does not only change the HR/HRV but also the BP [87]. Normal BP variations during respiration come from negative intrathoracic pressure during inspiration in contrast to expiration and is known as the pulsus paradoxus [88]. As described above, the large negative intra-thoracic pressure increases the pressure across the wall of the left ventricle. This pressure gradient causes an increase in afterload, resulting in a decrease in stroke volume and contributing to the decreased BP [89]. A change in the contraction force also changes the BP. With higher contraction force, the BP increases, and vice versa. The ANS itself also influences the BP. The BP fluctuates substantially with behavior (e.g., feeding and moving), but the 24 h BP is tightly regulated. The neural control of circulation is primarily designed to regulate blood volume and blood flow at the expense of BP [90]. How a set point for BP is encoded by the CNS, and the nature of the error signal has yet to be determined. The only well-identified neural (ANS) sensor of BP is the baroreceptor. The decrease in systolic BP leads to a faster HR due to the baroreceptor reflex, which stimulates sympathetic outflow to the heart [91].
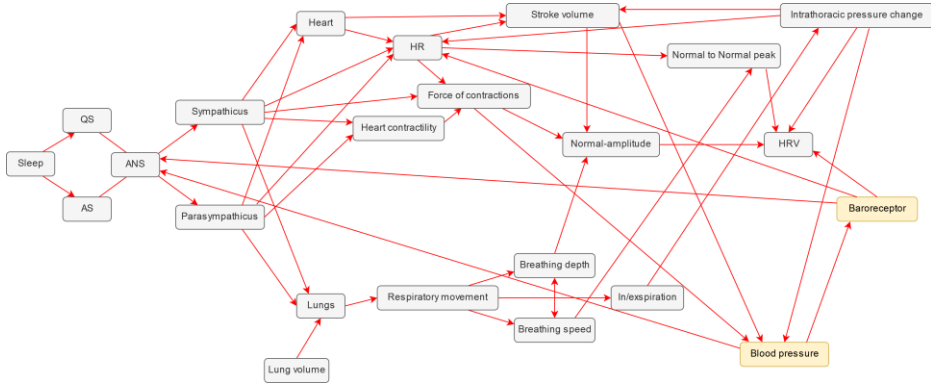
Figure 12 Blood pressure is tied to the intrathoracic pressure change, stroke volume of the heart, and the force of the heart contractions. The blood pressure has a direct influence on the ANS and an indirect influence via the baroreceptors. The baroreceptor also influences the heart rate and heart rate variability. These connections are marked yellow.

The cardiac baroreceptor reflexes become more functional with postnatal development feedback loop produces low-frequency periodicities of the HRV, and it is affected by both sympathetic and parasympathetic modulation of the heart [92]. The low frequency does not reflect cardiac sympathetic tone [93] (Figure 12).

The BP is also influenced by vascular resistance, which is influenced by the CNS and thermoregulation [90]. Thermoregulation has an impact on the HRV and vascular resistance. Very low-frequency power may reflect thermoregulation to ambient and are poorly developed until 34 weeks GA [94]. The baroreflex temperature changes [95]. Vascular resistance can change due to blood viscosity (low frequency) and the radius of the vessel, which can change more quickly due to environmental changes (incl. thermoregulation) or the CNS. Chemoreceptors create information flow from the HRV to the respiratory system. Thus, fluctuations in arterial pH and pCO_2 drives respiration via the central and peripheral chemoreceptors [96] (Figure 13).

The time delay and dynamics in this feedback system are modified by fluctuations in systemic and cerebral circulations, which are influenced by HRV [97]. The central and peripheral chemoreceptors are the sensors of the ANS
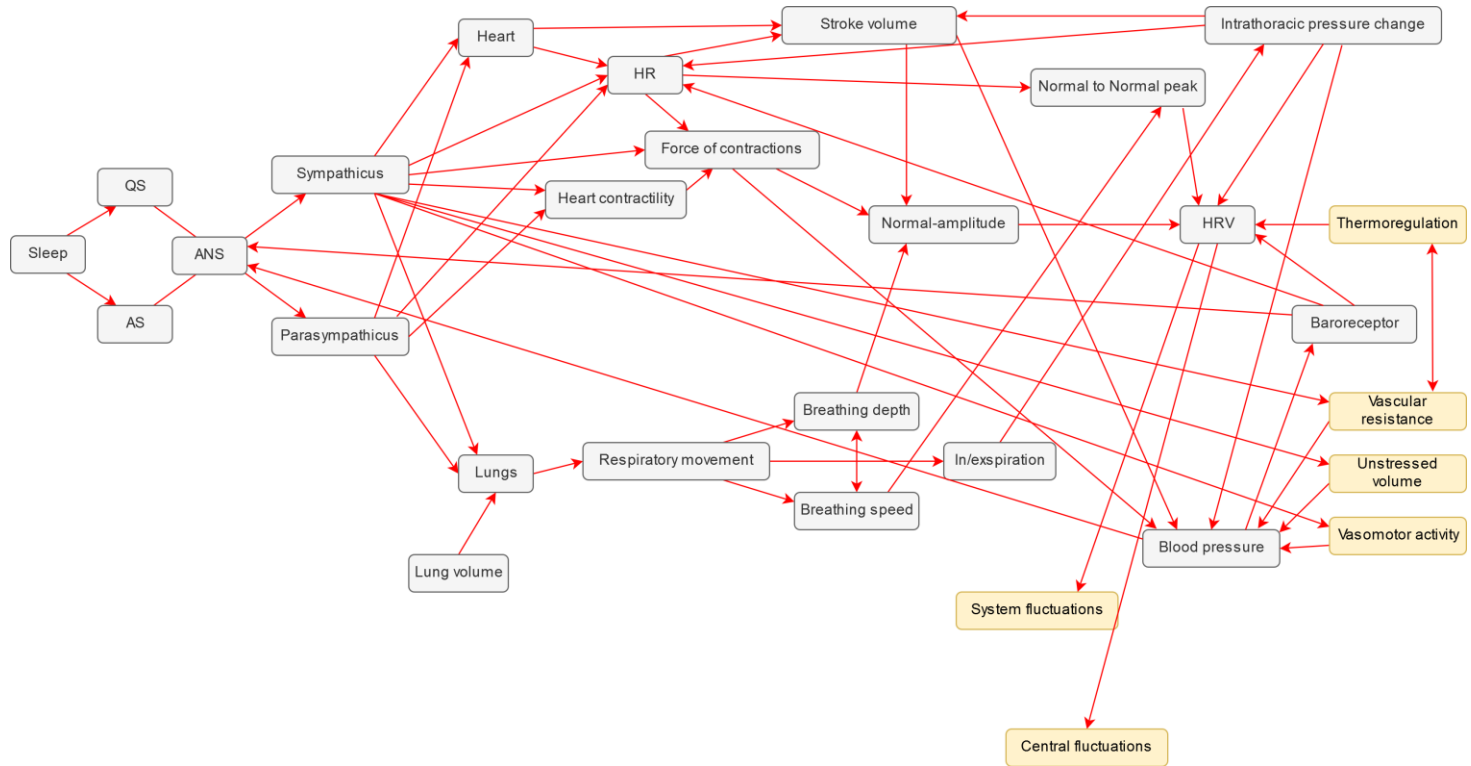
Figure 13 Blood pressure is further influenced by mechanical parameters of vascular resistance, unstressed volume, and vasomotor activity. Thermoregulation also influences vascular resistance. This has a direct impact on heart rate variability. The heart rate variability accounts for central and system fluctuations. These connections are marked yellow.
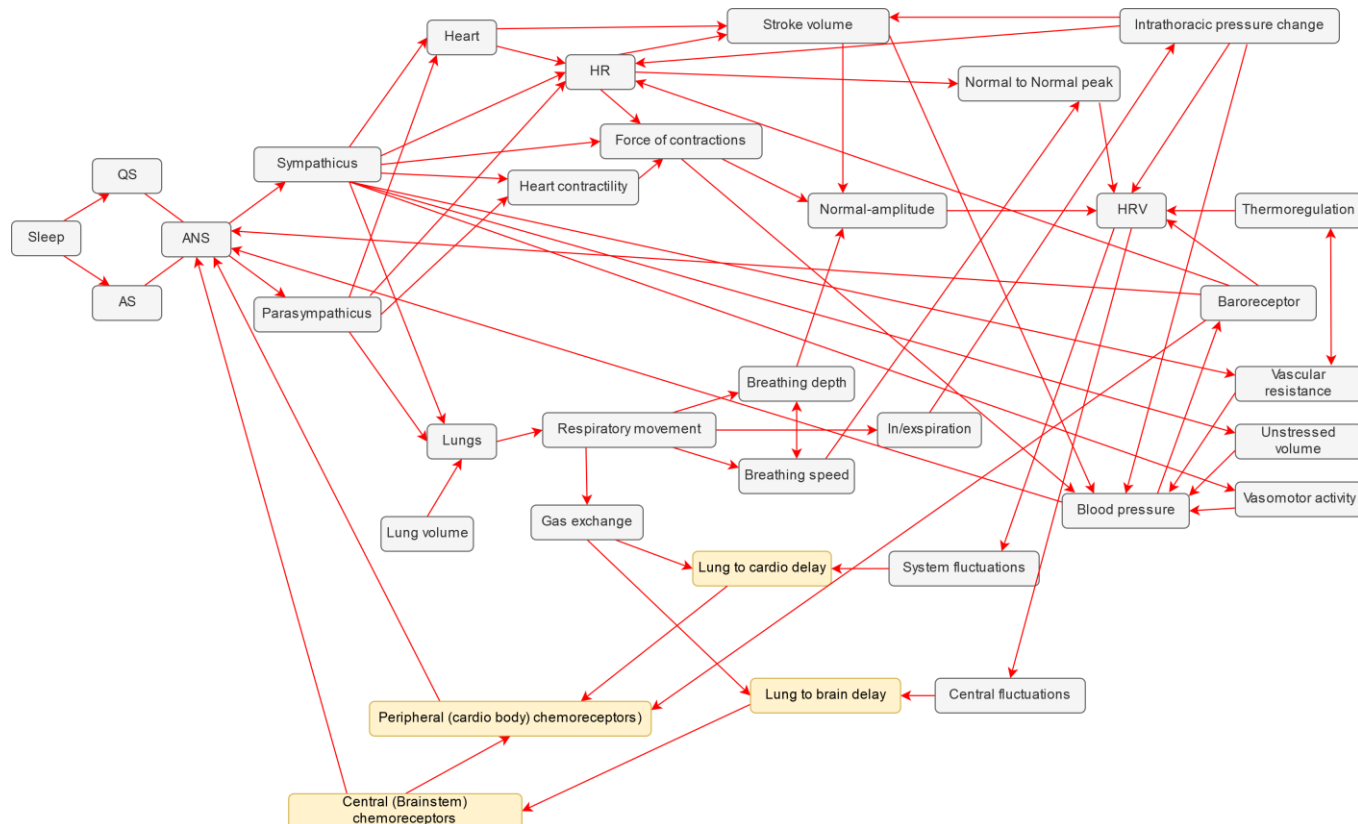
Figure 14 The connection between heart and lung and lung to the brain is delayed. Those delays influence the ANS via peripheral and central chemo-receptors closing the loop. The delays are influenced by the HRV-induced fluctuations and the process of gas exchange in the alveoli. These connections are marked yellow.

toward chemical changes (here the changes in pH and pCO_2 in the blood). The arterial baroreflexes have a powerful inhibitory influence on the chemoreflexes [98]. Very low-frequency periodicities, in the frequency range less than 0.04Hz, have been ascribed in different ways to modulation by chemoreception [99]. There is a strong negative influence of the peripheral chemoreceptors toward the central chemoreceptors [100].

1.7   Research Question

The main research question addressed in the presented thesis is as follows:

Can preterm infant sleep states be automatically classified only based on electrocardiography (ECG)-derived features?

Other groups have looked into (preterm) infant sleep classification based on different physiological signals such as EEG [54], [101], [102], activity [103], [104], video analysis, or respiration [105]. Nevertheless, the signals used in those studies are usually not continuously recorded in a NICU. We believe that continuous monitoring is necessary for sleep pattern analysis, sleep cycle prediction, and neural development monitoring. Moreover, a preterm infant is already struggling with the harsh conditions of its new living conditions, so burdening a preterm infant with an additional sensor for continuous monitoring should be avoided. For both stated reasons, we focused on the ECG signals for analysis as they are commonly used in the NICU for continuous heart signal monitoring. Focusing on the less developed countries, ECG systems are more broadly present and cheaper accessible in the health care environment than other monitoring systems such as video or EEG. Accessibility and affordability of neurodiagnostic tests are linked to the economic status of a country [106]. Therefore, an ECG-based sleep state classifier can potentially be adapted on a much broader scale, including low-income countries.

A further question focusses on the approach of classifying the sleep states. As laid out in the previous chapter, there are many cardiorespiratory influencing factors. It

should be investigated if it is possible to describe the sleep state differences with linear models or if these underlying cardiorespiratory factors influence the measured ECG signal in a way that nonlinear functions have to be implemented. Do these nonlinear functions have to be of such complexity that deep neural networks are in need to describe the interdependencies in higher dimensionality?

A subsequent question is the target of classification. Is it possible to classify all states to an adequate degree? Which state combinations can be classified with which grade of complexity? Is it possible to classify the two major states AS and QS only with linear classification methods?

In the following chapters, we address those questions to give insights into automated preterm infant sleep classification and provide information for the preterm infant research community.

2 Methods to obtain ECG signals unobtrusively in preterm infants

2.1 Abstract

Sleep is important for the development of preterm infants. During sleep, neural connections are formed, and the development of brain regions is triggered. In general, various rudimentary sleep states can be identified in the preterm infant, namely, active sleep (AS), quiet sleep (QS), and intermediate sleep (IS). As the infant develops, sleep states change in length and organization, with these changes as important indicators of brain development. As a result, several methods have been deployed to distinguish between the different preterm infant sleep states, among which polysomnography (PSG) is the most frequently used. However, PSG is limited by the use of adhesive electrodes or patches that are attached to the body by numerous cables that can disturb sleep. Given the importance of sleep, this chapter explores more unobtrusive methods that can identify sleep states without disturbing the infant. After a brief introduction to preterm sleep states, an analysis of the physiological characteristics associated with the different sleep states is provided, and various methods of measuring these physiological characteristics are explored. Finally, the advantages and disadvantages of each of these methods are evaluated, and recommendations for neonatal sleep monitoring are proposed.

2.2   Introduction

Sleep and its sleep states have a specific role in preterm infants. Active sleep (AS) is predominant during the first weeks after birth and is associated with an active developmental process such as forming of neural connections and developing specific brain regions [45], [52], [55], [56]. By contrast, quiet sleep (QS) is considered a resting and reenergizing state. In QS, errors during development are repaired, and the brain plasticity is used to reform brain regions to the demands of the situation at hand [52], [56], [63]. As the distribution of AS and QS changes throughout development, the state of development can be recognized by the sleep state distribution [39]–[45]. Therefore, continuous monitoring can provide an indicator of such development over time. Currently, this monitoring is mainly performed with polysomnography (PSG) and/or behavioral observations to determine the quality of sleep and potentially detect pathological sleep events. PSG contains many measurements that require the attachment of sensors and electrodes to the body of the infant. The use of adhesive electrodes on preterm infants can cause damage to their fragile skin, making them more prone to infections and leading to disfiguring scars in 10% of the cases [107], [108]. In addition, cords and patches can be an extra burden as the muscles of preterm infants usually are not strong enough to master the weight of the cables. Given the necessity of sleep monitoring and the obtrusiveness of commonly used methods, unobtrusive sleep measurement methods provide a much more comfortable alternative with a minimal burden on preterm neonates.

These techniques can also be used for long-term automated sleep and development monitoring due to their unobtrusiveness. Thus, long-term monitoring can lead to personalized sleep pattern identification, which can be implemented in the care plan around the preterm infant. Additionally, alarm sounds can be adapted to the sleep state of the preterm infant to reduce disturbance to the infant and alarm fatigue of the caretaker. Alarm fatigue can lead to an increase of missed or delayed attention to critical situations by desensitization of caretakers to alarms [109]. Furthermore,

continuous sleep monitoring can help detect sleep-associated events such as central apnea and may reveal medication effects on the sleep architecture and neural development of preterm infants. Sleep monitoring might also support identifying discomfort or stress by an altered sleep architecture.

In general, sleep monitoring will increase the awareness of the caretakers to sleep by integrating the moments of necessary neonatal (intensive) care without disturbing the indispensable sleep of the vulnerable preterm infant.

2.3   Measurements methods

Infant sleep is scored by two main methods: PSG and behavioral sleep measurement. The scoring parameters of both methods can be found in the corresponding publications, e.g., [51]. The three ways to perform classification with the two methods are: the use of PSG methods alone [36], [54], [110]–[112], the use of behavioral methods [43], [44], [113], [114], and the use of both methods in combination [37], [41], [47], [51], [115], [116]. However, the standard in preterm infant sleep classification is still the use of manually annotated EEG signals.

In the following sections, we give an overview of central signal modalities for preterm infant sleep analysis.

2.3.1   EEG signals

EEG signal patterns can be used to distinguish sleep states. The EEG signal patterns that separate the sleep states are described in detail in several publications [66], [117]. These patterns change with the development of the preterm infant and are also well discussed in the literature, e.g., [66], [117], [118]. Niemarkt et al. [117] explain that the EEG signal of a preterm infant before 32 weeks GA is formed by continuous (tracé continue) and discontinuous background patterns. During development, the inter-burst intervals (IBI) of the discontinuous patterns shortens, and the burst length increases. With advancing development, the discontinuous pattern can then be

separated into tracé discontinue and tracé alternant. The tracé alternant is detectable from 34 GA onwards and is connected with QS [117].

With manual EEG annotation sleep state separation is usually possible as early as week 32 to 34 GA (for <34 weeks GA with the use of other EEG patterns than tracé alternant [117]). In the review by Grigg-Damberger et al. [51], it is stated that sleep states cannot be determined earlier than 32 weeks GA on the basis of EEG alone. Therefore, sleep state separation at an earlier stage needs additional PSG measurements to be combined with the EEG measurement. After that, the sleep states can then be separated during 50%-80% of the time [39], while the remaining time is labeled as IS.

In Parmelee et al. [119], it is explained that not all parameters show the same variability at a certain age and are thereby less suitable to determine sleep states in premature infants at that particular age. They gave an overview of the different parameter usefulness over age (see Table 3).

The earliest separation of the sleep states was performed at 26-30 weeks GA by Sheldon [120] using only the behavioral signs. Curzi-Dascalova et al. [41] separate the states also early at 27 weeks GA by using EEG in combination with other PSG parameters, and HR. An overview of the selected group findings on first AS and QS separation can be seen in Figure 15.

Table 3 Usefulness of measurements for sleep state determination from Parmelee et al. [119]. This table shows the different usability of parameters for sleep state separation due to their variability at a certain gestational age (GA).

Age in weeks GA	22	26	30	34	38
Body movement	+	+	++	+++	++++
Eye movement		+	++	+++	++++
Breathing rate			+	++	+++
EEG			+	++	+++
Chin electromyography				+	+++
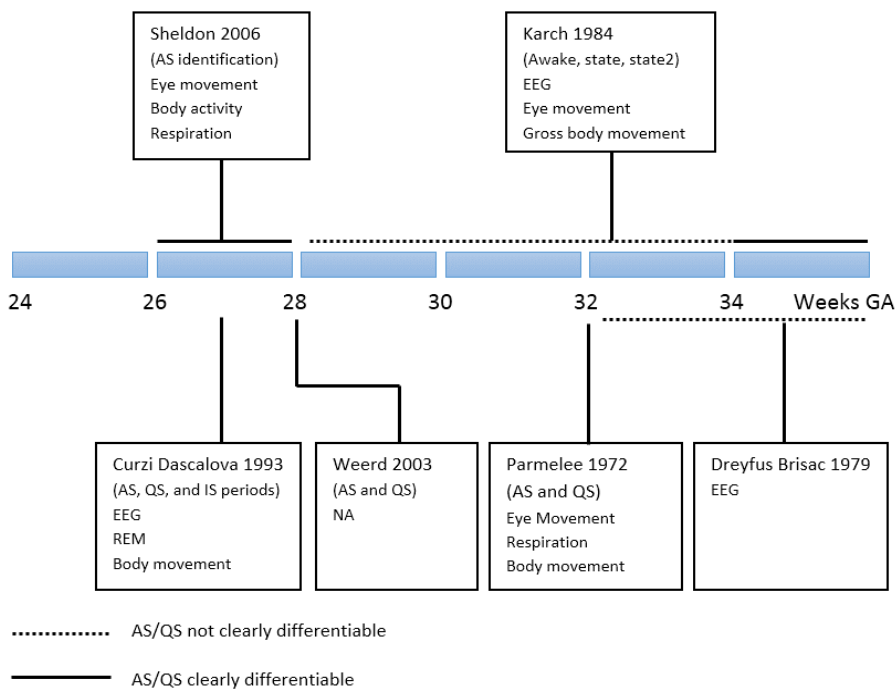
Figure 15 AS and QS separation over gestational age (GA).
The different time points at which active sleep and quiet sleep can be separated are shown here. The used method by which the sleep states where separated is included. Citations are from upper left to lower right: [120], [299], [41], [39], [119], [300].

While AS is always detected in an earlier stage of development [40], [120], QS can be separated at a later stage at normally 32 weeks GA [121]. The difference in the time of first sleep stage separation (as shown in Figure 15) may have occurred due to the used measurement method, the signal quality, the medical condition, and/or stability of the preterm infant. This difference could also be related to the number of annotators, the experience of the annotator, as well as the confidence of the annotators in their findings.

The main methods of PSG regarding the preterm sleep state separation are shown in Table 4.

Table 4 Polysomnography attributes.
This table shows how the different preterm infant sleep states are observed using different features. The states are active sleep (AS), quiet sleep (QS) intermediate sleep (IS), awake (W), and arousal. The vital signs/movements are brain activity, heart rate, eye movements, heart rate variability, muscle activity, and respiration.

	Brain activity	Heart rate	Eye movements	Heart rate variability	Muscle activity	Respiration
AS	Continuous mix of EEG patterns 40-80µV in amplitude	Mostly irregular	Slow and rapid bursts and isolated eye movement; 20s epochs of REM bursts	Low frequencies are dominant 0.03 to 0.39 Hz	Low amplitudes superimposed with twitches and phasic saccadic movements; Low tone between movements	Uneven; Faster in rate; Costal respiration; Higher minute ventilation during REM; Paradoxical breathing can appear
QS	<34 GA Discontinuous with delta bursts >32< 46 GA tracé alternant 50-150µV in amplitude	Mostly regular with some acceleration during startles	No Eye movement (infrequent eye movement can be seen)	High frequencies are dominant 0.4 to 1 Hz	Low amplitude; Tonic motor activity	Relatively regular and abdominal; Slow and constant in Amplitude; Can also be irregular
IS	Defined if no other state fits	"	"	"	"	"
W	Mixed EEG pattern with movement artifacts	High and irregular heart rate	Open and moving eyes; No REMs; Eyes can be closed during crying	Variable	High muscle tone superimposed with general body movements	Regular; Tracking motor activity
Arousal	Decreased amplitude and increased frequency	Heart rate acceleration	Open and fixed eyes	Increased beat to beat variability	General body movement incl. startle >1min = awakening	Increased respiration variability and pauses; Sighs can be heard

2.3.2 Heart rate variability

Preterm infant HR lies between 120 and 160 bpm [85], [122]. The HR and the heart rate variability (HRV) change with the development of the preterm infant with a decreasing HR and increase of R peak intervals [122], [123]. Thereby, the HRV is highly related to the developmental stage [124], [125]. With the development, not only the absolute values of the HRV and HR change but also the response time to ex- and internal changes increases over time [125]. In addition to their long-term trends, the HR and HRV are also directly connected with the immediate sleep states changes, resulting in immediate changes in the HR and HRV. Thereby, the HR decreases during QS and increases during AS [124], while the HRV reflects the function of the cardiovascular autonomic control and ANS, as described in section 1.6.3. Therefore, the different sleep states can be distinguished in the power spectrum and time-domain features of the HRV [123]. The HRV also changes with development due to the maturity of the ANS and, more specific the parasympathetic control of HRV [126]. Krueger et al. [127] showed that the high frequencies between 0.3 and 1.3 Hz increased within female patients from 28 weeks GA to 34 weeks GA. Other features like the low frequencies between 0.04 to 0.2 Hz, or the ratio between low and high frequencies did not change significantly over maturation. Other group findings confirm these trends for older preterm infant groups between 31 weeks GA to 41 weeks GA without gender differentiation but also a decrease in the ratio between low and high frequencies [128]–[130].

2.3.3 Respiration

A preterm infant breathes 40 – 80 times per minute [85], [122]. As with HR and HRV, the breathing patterns change throughout the preterm infant's development. The breathing/respiration rate (BR) gets more regular within QS with maturation [43], [46]. These breathing patterns can also be considered as an indicator of the sleep states, where the respiratory peak can be used to estimate the vagal activity and

separate the sympathetic and parasympathetic dominance respectively AS and QS [131]. Another way to identify sleep state changes is to use the mentioned regularity of BR. These changes are reflected in both frequency [43] and amplitude [46] of the BR signal. The BR is relatively even during QS while more irregular during AS [37], [43], [116]. Holditch-Davis et al. [43] measured the regularity in the BR and divided it into very regular, regular, and irregular. The following are the criteria they used to separate these categories:

Regular respiration:

- No more than one breath is between 20% - 50% of the max peak height.
- The narrowest peak to peak is at least 50% of the widest peak to peak interval.
- Very regular respiration criteria are not met.

Very regular respiration

- During a 10s epoch, the smallest breath is at least 80% of the height of the largest breath.
- The narrowest peak to peak interval is at least 67% of the widest peak to peak interval.

All other breathing patterns can be classified as irregular patterns.

2.4   Methods for behavioral sleep classification

Behavioral measurements for sleep state analysis are used to assist the PSG analysis for a more robust annotation. However, they can also be used on their own to separate sleep states. Behavioral classification for sleep uses general body movements [37], [41], [112], [115], [116], specific body movements like face, eye, chin, or limb movement [37], [47], [54], [110], [116], and BR [43], [51], [111]. General body movements (also referred to as motor activities) range from low activity in drowsy or

alert states to high activities during crying periods [46]. The sleep states can be behaviorally scored in the following manner:

AS: The eyes are closed [38], [69], [103] or slightly opening and closing [114]. They might open during REM in AS [69]. A wide range of motor activity can be detected [46]. Motor activity is sporadic and appears in bursts of 5-60 seconds [69], but the muscle tone is low in between the movements [43], [69]. The facial expressions include smiles, grimaces, and frowns; also bursts of sucking movements are seen [38], [103] as well as small twitches. Sighs and sobs can be heard [103].

QS: The eyes are closed [38], [69], [103]. Little or no motor activity is detectable [46]. Motor activity only appears with occasional startles, sighs, or other brief discharges. A tonic motor level is maintained [43], [103]. Mouth movements or sucking can be seen [38], [69].

An overview of the behavioral observations regarding the sleep states can be found in Table 5.

Preterm infant behavioral patterns are often considered static over time. However, Holditch-Davis et al. [43] found decreasing body movement with maturation. They realized that their finding contradicts other publications and account it with a longer follow up period of their trial [43]. If body movements are used to annotate sleep, and the sleep organization is changing with maturation, it can be assumed that the body movements undergo the changes as well. Therefore, age should also be considered as a factor in the behavioral annotation.

Table 5 Behavioral attributes.
This table shows behavioral attributes that are used to identify sleep/wake states. These states are active sleep (AS), quiet sleep (QS), intermediate sleep (IS), wake (W), and arousal.

	Movement / Motor activity	Eyes	Face	Audio
AS	Wide range; Small twitches; Sporadic motor burst of 5-60s; Low muscle tone in-between	Closed or slightly opening and closing eyes; REM	Smiles; Grimaces; Frowns; Bursts of sucking	Sighs; Sobs
QS	Little or none startles; Continuous tonic levels; Motor low	Closed eyes	Episodes of rhythmic mouth movements; Sucking	Sighs
IS	Defined if no other state fits	"	"	"
W	Head and arm movements; Orientation response; Motor high or low	Open eyes; Focused eyes; No REM; Scanning	In quiet wake relaxed; No smile or frown	Crying; Fussing
Arousal	Head and hand movements; Motor can have different states	Open eyes; Fixed and focused	Mouth movement	Sighs

2.5 Unobtrusive measurement methods for sleep state monitoring

Contrary to behavioral observations, other sleep monitoring methods require the attachment of adhesive electrodes, including cables running over their limbs. As mentioned in the introduction of this chapter and the general introduction, those are additional burdens for the fragile preterm infant. Therefore, unobtrusive methods for vital signs and movement measurements regarding preterm infant sleep state classification are described in the following sections 2.6 to 2.8. An overview of the

different measurement methods for vital signs and movements is given beforehand in Table 6.

In the following sections, we will discuss different methods to obtain the necessary vital signs and movement signals for preterm infant sleep state classification. Clustered in the following groups:

- PSG with electrical origin
    - EEG, electrocardiogram (ECG), and electrooculogram (EOG)
- PSG with mechanical origin
    - HR, BR
- Behavioral signals
    - Body movement, facial expression

Table 6 Unobtrusive methods.

The measurement methods are listed in relation to the vital signs and movement measurements. The methods are merged for several vital signs or movement. The degree of unobtrusiveness of the methods is indicated with colors and diagonal lines [for b&w prints]. Red (two diagonal lines) indicates high to medium obtrusiveness and green (no diagonal line) indicates minimal obtrusiveness. The methods are explained in sections 2.6 to 2.8.

	Unobtrusiveness	EEG-signal	ECG-signal	Eye movement	Heart rate	Respiration	Movement	Facial expressions
Dry electrodes	(red)	No conductive gel needed; High impedance; Low movement artifacts; Applying pressure needed				-	-	-
Needle patch	(red)	Minimal invasive but skin penetration; Very low impedance; Suitable for long term monitoring; Possible low acceptance by parents				-	-	-
Capacitive electrodes	(green)	Unobtrusive; Prone to motion artifacts					-	-
Laser Doppler vibrometry	(green)	-	-	-	Unobtrusive; Prone to motion artifacts; Heart rate, respiration, and movement has to be separated; Need clear field of view		-	-
Ballistocardiogram	(green)	-	-	-	Unobtrusive; Sensitive; Prone to motion artifacts; Heart rate, respiration, and movement has to be separated; Camera needs clear field of view			-
Photoplethysmogram	(green)	-	-	-				-
Accelerometry (behavioral)	(green)	-	-	-				-
Doppler radar	(green)	-	-	-				-
Camera	(green)	-	-					

2.6   Unobtrusive EEG/ECG/EOG signal measurement

In this section, unobtrusive measurement methods for electric originated signals with dry electrodes and capacitive electrodes are reviewed. Normal gel electrodes are not useful for long term monitoring of preterm infants because of impedance changes due to drying gel [132]. Additionally, there are inflammation risks at the electrode-skin contact [108]. In some rare cases, toxicological concerns of the gel electrodes were reported.

In preterm infants under 34 weeks GA, the epidermis is 2-3 layers thick with barely any protective outer skin layer (the stratum corneum) [107]. Therefore, all electrodes in contact with the fragile skin are considered obtrusive.

2.6.1   Dry electrodes

Dry electrodes create contact with the skin without the need for skin abrading to reduce skin resistance or any application of conductive gel to ensure correct electrode contact. There is a variety of dry electrodes like, e.g., different metal disks, conductive rubber, spring-loaded fingers, or conductive foam. The principle is that in most cases, the material follows the skin contour without leaving any high impedance air gap, as shown in Figure 16. The interface impedance with dry electrodes is higher than with gel electrodes as only the skin moisture/sweat is used as a conductive bridge, which has a lower ion conductivity than conductive gel [108]. Lower impedance means that smaller signal amplitudes can be measured (e.g., low EEG amplitudes). Also, the thermal noise (Johnson noise) of the gel electrode itself is lower, leading to a higher signal to noise ratio (SNR).

A dry electrode type with minimal skin contact is the Nanoneedle array (Figure 17). This electrode consists of 50 nm wide and 10-15 μm long carbon tubes [134], which can puncture through the high impedance 10-15 μm thick stratum corneum [135], reaching the underlying low impedance layers. Although the Nanoneedle arrays could
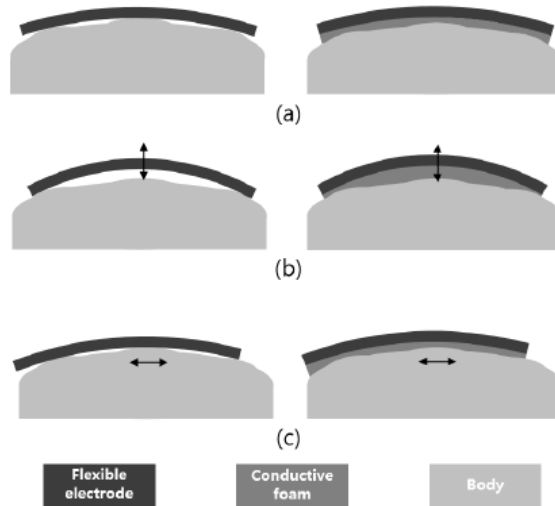
Figure 16 Use of conductive foam in combination with a flexible electrode (From [133] with the permission of the authors).
The regular flexible electrode contains many high resistant air gaps between electrode and skin (left side). It is also prone to movements. The conductive foam closes the gaps and reduces the movement (right side).

be considered invasive, Lopez Gordo et al. [132] state that due to the minimal diameter of the needles, the infection risk is minimal. This makes them less harmful while showing similar impedance and fewer motion artifacts than classic wet electrodes. Other dry electrodes have to be fixated with pressure or adhesives to the preterm infant; for that matter, they become almost as obtrusive as the Nanoneedle electrode, which only needs minimal applying pressure. Additionally, Nanoneedles have a minimal surface in contact with the skin and the skin penetration affiliates with only minimal infection risk. The needle stability also minimizes the infection risk and additionally stabilizes the impedance [132]. Due to their minimal penetration depth, the Nano- and Microneedle arrays are painless (so far only shown for adults). The underlying nerves are not stimulated enough to trigger pain [136]. If that holds also in neonates has to be investigated.
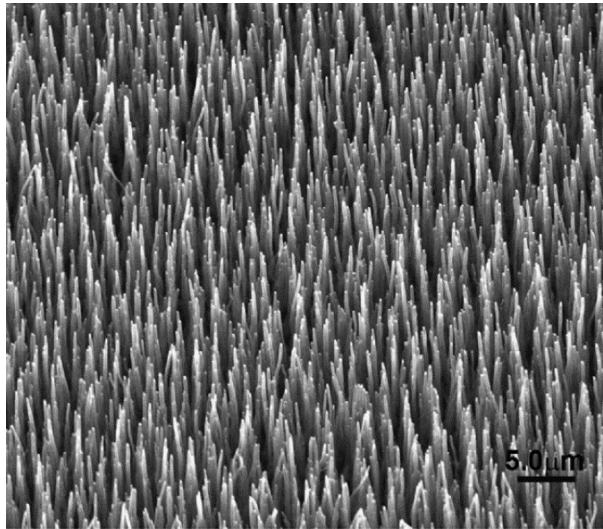
Figure 17 Electron microscope image of a multiwalled Carbon nanotube pad (From [134] with permission from the author).

Nanoneedles should not be confused with more routinely used subdermal needle electrodes. Subdermal needle electrodes are, with around 0.5 mm, relatively thick compared to Nanoneedles (factor 10000), posing a higher infection risk. They also can cause pain due to the penetration depth reaching the underlying nerves [137]. Even though they are routinely used, the American Clinical Neurophysiology Society does not recommend using subdermal needle electrodes on neonates or infants [138]. Therefore, we did not proceed with further investigations into subdermal needle electrodes.

2.6.2 Capacitive electrodes

Another unobtrusive vital sign measurement method is the use of capacitive coupling. This method allows the skin to couple capacitively to the measurement electrode without any direct contact with the skin of the preterm infant. However, the biggest problem with these electrodes is their sensitivity to movement and dielectric
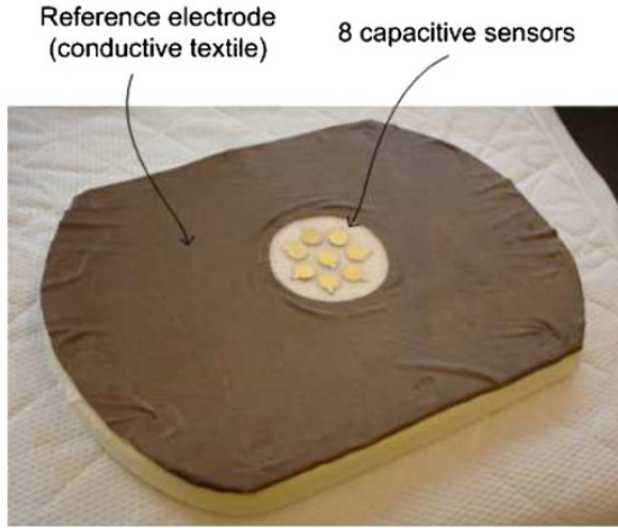
Figure 18 Inside of a neonatal mattress with 8 integrated capacitive sensors (From [107] with permission from the author).

artifacts. The coupling capacitance is altered by every movement [139], [140] producing artifacts. A problem, more for EEG than ECG measurements, is the small amplitude of signals and the high resistive elements of capacitive electrodes [140]. Therefore, the attached amplifier circuits need to have very low noise [141] not to mask the EEG signal, which makes the measurement unit more complex. Chi et al. [142] and Sullivan et al. [141] created contactless EEG/ECG electrodes with noise amplitudes comparable to gel electrodes. To lower the motion artifacts without increasing the complexity of the measurement unit Serteyn et al. [143] used an estimation of artifacts to lower them in their measurements, reaching a noise level 10 times smaller than the recorded ECG amplitude. This indicates that it is better to eliminate motion artifacts during post-processing rather than with more complexity in the amplifier circuits. For ECG monitoring alone, the electrodes can also be placed under the preterm infant, e.g., in a mattress, to detect the HR or HRV. Atallah et al. [107] used a sensor array (shown in Figure 18), which also takes into account that the preterm infant is not always in the same position. They developed a mattress with eight capacitive sensors and adaptive channel selection, allowing contactless ECG and
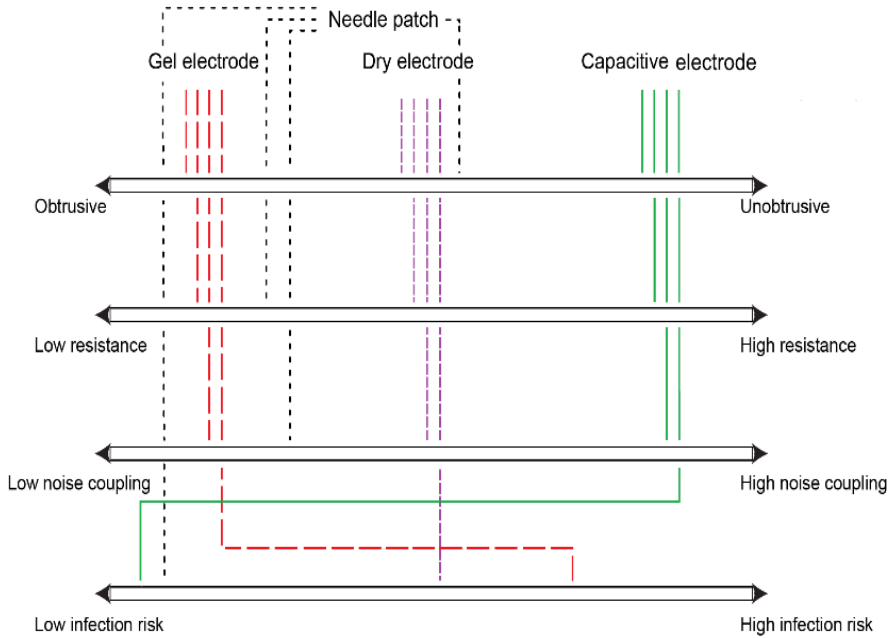
Figure 19 Comparison of different electrode types.
The four different electrode types: gel, dry, capacitive, and needle patches are illustrated for their properties: obtrusiveness, resistance, noise canceling, and infection risk. The gel electrode show poor results in unobtrusiveness and infection risk compared to capacitive electrodes. Dry electrodes and needle patches share a medium obtrusiveness, where the needle patch has almost no infection risk. On the technical side, gel electrodes and needle patches show low resistance and noise coupling compared to dry and capacitive electrodes.

HR measurements. With this method, they were able to achieve a sensitivity for ECG up to 89% dependent on motion, and a positive predictive value of 99% independent of motion, which can be interpreted as very high accuracy. Thereby, the results improve with fewer layers of cloth between the infant and electrodes [107]. Figure 19 looks into obtrusiveness, resistance, noise coupling, and infection risk of the electrodes discussed in this section. Capacitive sensors show excellent results regarding those parameters compared to gel electrodes, which appear to perform the poorest. This is due to their unobtrusiveness and low infection risk. On the technical side, gel electrodes perform well regarding resistance and noise coupling, but this does

not make them more suitable for preterm infants. The Nanoneedle patches are a good tradeoff between good technical properties and minimal harm to the infant. However, their use for the neonatal population needs to be carefully assessed.

2.7 Unobtrusive heart rate and respiration measurements

For signals of mechanical origin, e.g., simple HR measurement, without the ECG waveform, there are additional methods to dry and capacitive electrodes that also allow contactless measurements.

2.7.1 Ballistocardiogram

A ballistocardiogram (BCG) is the measurement of the cardiac activity, which aims to retrieve the mechanical signal from the platform on which the patient is lying. This could be, e.g., the mattress inside an incubator or the incubator rack itself [144]. The signal is created by the force of the heart pumping blood through the body. The propagation to the outside depends on blood pressure and vein elasticity [145]. The generated pulse wave can be monitored electrically with piezo elements (e.g., bed film sensors, Figure 20) [146], mechanically with load cells [147], [148], or accelerometers, in this form of application known as kinetocardiogram [149]. The difference between the methods is that measuring with accelerometers needs time-varying pressure, while piezo and load sensors measure an absolute pressure value. With any of these sensors, BCG can be used for both HR monitoring and BR monitoring [147], [148]. These methods have been successfully used with adult patients [146]–[149] and even tested in the neonatal intensive care unit (NICU) [150]. In adult patients, BCG showed the same result as classic PSG with a mean epoch to epoch agreement of 92.5% and a kappa of 0.62 between BCG and PSG [151]. The test on a preterm infant showed a high correlation between ECG reference with BCG of 0.91 and BR reference with BCG of 0.74 [150]. There are commercial products for infants using BCG, e.g., AngelCare [152], NannyCare [153], or Nanny [154], but they are mostly used as

sudden infant death syndrome alarm systems rather than continuous monitoring devices.

BCG methods pose a difficulty in designing algorithms for separating the HR and BR from motion, and noise. Brink et al. [147] found that for adults, even the resonance frequency of the mattress could lead to misinterpretation of signals. This might not be relevant for preterm infants but emphasizes that slight movements can already affect the signal. These problems may be solved by adaptive signal processing, as done by Liu et al. [155], who presented a new method to separate the BR in adults from movement with pressure-sensitive textile detecting body position, movement, and BR during sleep with high resolution. They achieved a low error rate of 5.7 – 1.7% for BR detection. Another problem in the NICU is the signal amplitude. The small body of a preterm infant does not produce large mechanical waves. Therefore, the sensors need to be placed as close as possible to the body and should not be shielded by damping material. On the other hand, preterm infants move in a smaller range than adults during sleep, which results in smaller amplitudes of motion artifacts and higher SNR.
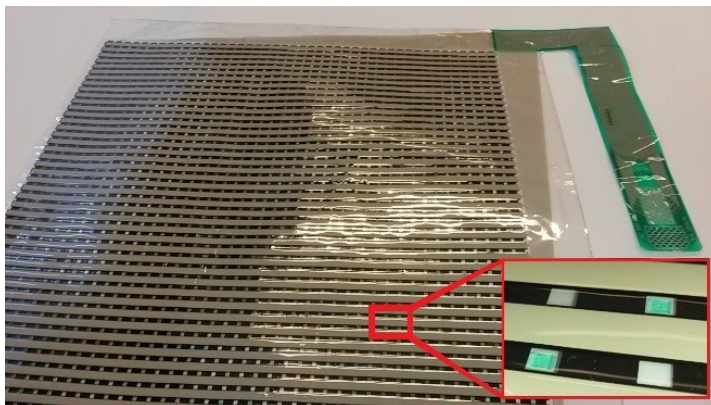
Figure 20 Bed film sensors.
The Sensor-mat of Tekscan [301] is made from numerous piezoelectric sensors and is placed between the mattress and the sheets to record movements. (Courtesy of Mohammed Meftah).

2.7.2 Radar

Unobtrusive monitoring of HR and BR is also possible via radar. The heart wave and the respiratory movements are propagated to the surface and can, therefore, be detected by a non-contact system. A radar signal is directed towards the body of the preterm infant, where it is reflected with a frequency shift. The moving body surface modulates the phase due to the Doppler effect (See Figure 21). This phase shift can be retransformed into HR and BR signals [156]. Thereby, the radar signals are reflected in different amounts by different materials, depending on the permittivity of the material. This is an advantage of the radar technology over other non-contact methods because the radar signal will be reflected by the preterm infant body (water permittivity= 50-88) but almost not by the clothing (cotton permittivity= 1-3) and incubator shield (PMMA permittivity= 5), making it independent of a clear field of view (e.g., under the incubator mattress). With the (adult) patient in the optimal position, the detection accuracy is up to 99.9% [157]. A preterm infant can easily be placed at the optimal position without a great range of random movements making this method more viable for the NICU than for adults. The use of radar for vital sign measurement suffers from similar effects as discussed for the previous methods; noise, movement, and even background movement from several meters away [158] can deteriorate the accuracy [156]. Recently, however, several algorithms have been proposed to deal with these problems. Solutions to the common radar problem of offset recalibration due to unwanted reflections from static objects have also been proposed [157]. Despite that, this method has, due to its experimental state, not been used in the NICU so far.

2.7.3 Laser Doppler vibrometer

A similar method to the Doppler radar is the Doppler laser vibrometer. The Laser beam is sent to, e.g., the chest of the preterm infant, where it is reflected with a shift in the phase due to the Doppler effect. The reflected light is then captured with a
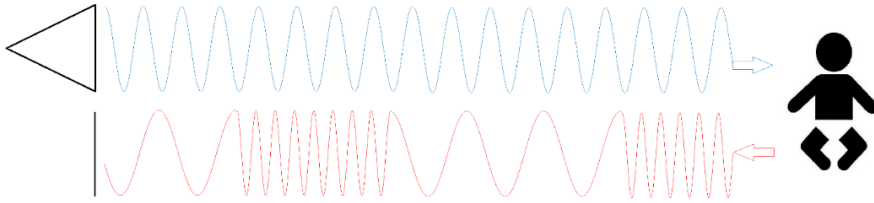
Figure 21 Basic principle of a Doppler radar/laser.
A radar or laser beam hits the target (upper signal) and is reflected by the target with a Doppler shift according to the target movement (lower signal). If the patient`s e.g., chest is moving out, the frequency increases and vice versa. From the shift of the incoming to the outgoing radar/laser signal, the target movement can be calculated.

photodetector (See Figure 21). With the captured light signals, the phase shift can be translated into movement (e.g., chest movement), which can be related to the HR and BR. Marchionni et al. [159] used this method successfully in the NICU, achieving very high correlations of HR/BR detection with the reference ECG and ventilation of 0.96 and 0.99. The energy of the laser beam has to be limited to avoid damage to the eyes or skin. However, this limitation does not interfere with the measurement. As with all mentioned non-contact measurement systems, the motion artifacts could reduce the accuracy of the method. The Doppler laser does not need direct skin contact to measure movement, but when used on textiles, noise is added to the measurement when the textile is slipping or moving in a different direction to the HR or breathing movements.

2.7.4 Photoplethysmography

Another unobtrusive way of HR and BR monitoring is via photoplethysmography (PPG). PPG is a common, simple, and low-cost optical technique that can be used to detect blood volume changes in the skin [160]. The PPG sensor emits light via LEDs into the skin and detects the reflected light. The determination of HR via PPG is based on the different absorption levels of light by hemoglobin and the surrounding tissue. The blood volume change in the veins, by the pumping heart, changes the amount of absorbed light and, consequently, the detected amount of reflected light.
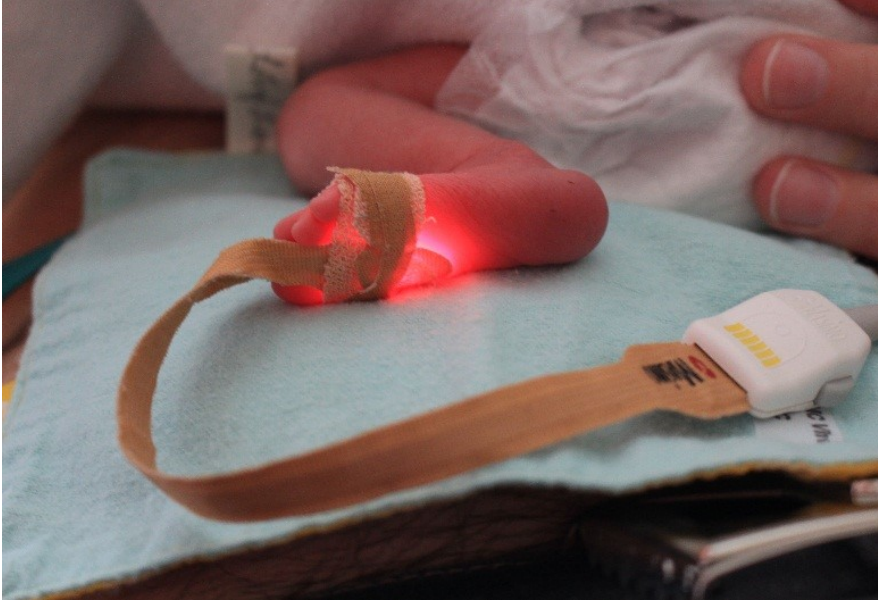
Figure 22 PPG on the foot of a preterm infant (With permission of the parents).
The PPG sensor is attached on the infant's foot. Here an adhesive plaster is used for fixation.
This can also be done with non-adhesives, e.g. a sock.

Thereby, the blood volume change can be measured and translated into HR and subsequently, BR. In addition to HR and BR assessment, PPG can detect respiratory sinus arrhythmias, which can be used to has been applied mostly to adult patients algorithms have been proposed [161]–[164]. Although they have not been tested on preterm infants, they have been used for newborn term infants [161], [165], [166] and term infants in the NICU [161], [167] with very high correlations of r=0.99 between the reference ECG and PPG as well as reference BR and PPG. The problems with this method for (possibly) preterm and/or term infants are the movement artifacts [167] and the complicated realization of a breathing protocol to calibrate the analysis. The calibration helps to separate the BR from other variations in the PPG around the same frequency range, e.g., temperature fluctuations, vasomotion, and/or baroreceptor oscillations (Mayer waves).

The classic PPG is performed with an attached light source and photodetector directly on the skin (Figure 22), where it is mostly wrapped around the foot of the preterm infant. It does not need any abrading of the skin and is thereby only slightly obtrusive. As described in the next section, a more recent and advanced method is the use of non-contact video PPG analysis.

2.7.5   Camera

The camera measurement method of PPG uses the same basic principles as classic ankle/foot PPG, which are described in the previous section, but avoids the use of any attached sensors. Another difference between the methods is that the camera measures the ambient light emitted from the skin of the patients instead of the LED light. An example is a work by Verkruysse et al. [168] who used a simple, inexpensive camera to detect the BR and HR with the use of ambient light as illumination. A simple webcam is sufficient and shows comparable results as a high-end camera system [169]. Koolen et al. [170] showed that analysis of HR is also possible in neonates using a camera. A pilot study in the NICU by Aarts et al. [171] resulted in 13 out of 19 preterm infant HR measurements with >90 % matching the reference ECG. However, they used a camera dependent on good illumination in the visible spectrum. Yet, the presented studies derive only HR, which is insufficient in this form alone for sleep state classification. ECG sleep state classification needs HRV, which needs precise R peak detection. Nevertheless, with improved camera techniques, the results in a laboratory setting can reach correlation values of 1.0 for HR, 0.91 for BR, and 0.97 for HRV [172].

This demonstrates that illumination is an important factor for camera-based approaches. This is especially important with monitoring preterm infants in an incubator. The light is mostly dimmed in the patient room, and the incubator is covered to avoid disturbance of the preterm infant. The light intensity is 3-6 Lux (a candle at a distance of one meter has 1 Lux) in a covered incubator and 60-120 Lux in an uncovered incubator with dimmed lights. Recently, Van Gestel et al. [173]

showed that for HR measurement, the use of cameras in the IR spectrum shows the same results as cameras operating in the visible spectrum, enabling the analysis in pitch dark incubator settings. This makes continuous and less obtrusive measurements possible. In this case, the camera has to be equipped with a near infra-red (NIR) light source. To protect the preterm infant's eyes, the maximum NIR radiation is required to be limited to 10 mW/cm^2 and the maximum total infra-red (IR) radiation to 60 mW/cm^2 [174]. When using an artificial light source, it has to be considered that not only the light amplitude is important for the monitoring via video but also the modulation of the light source, which can desynchronize the signals. However, Tarrasenko et al. [175] showed very recently that this artificial light flicker noise can be eliminated with autoregressive methods. An additional problem is that of movement artifacts [176]. However, movement artifacts can be removed with an algorithm to synchronize frames [168] or cross-correlate several frames, increasing the SNR for BR and HR [176].

2.8   Unobtrusive behavioral measurements

Video analysis can also offer a solution for behavioral sleep analysis. Limb movements, face twitches, and REM can all be detected visually. The illumination problem is the same as that discussed in the camera section above. Camera systems were used to annotate sleep in premature infants [63]. However, no automated algorithms were used in most of the cases. Instead, time-lapsed video analysis was used to annotate the videos manually [44]. One of the few exceptions is the work of Scatena et al. [177] in which the open-source video analysis tool ZoneMinder was used to automate the video analysis for sleep in adults. However, the results were only fair with a Cohens Kappa value of 0.47 and 0.654 between the software and reference PSG and the software and reference actigraphy. No automated video analysis has been conducted for preterm infants yet. Due to the similar vessel composition of preterm infants and adults, adaptation in the video analysis and actigraphy software might enable an automated analysis for preterm infants. Similar to HR and BR with

BCG, actigraphy can be acquired by the use of pressure sensors or accelerometry embedded in, e.g., the neonatal mattress [69], [178], detecting sleep, wake, or activity patterns [179]. Pressure sensors can detect slight changes in the g force (>0.05 g) [103], [177], making it suitable for the less intense movements of preterm infants. The signals of the movements are stronger than those of HR and BR, which implies easier monitoring. Limb movement measurement for preterm infants can be obtained by a wrist or ankle worn accelerometer [40]. In this case, the signal that was considered previously as noise can now be used for sleep state separation. However, the noise signal information needs to be combined with other measurements if small motion patterns need to be registered.
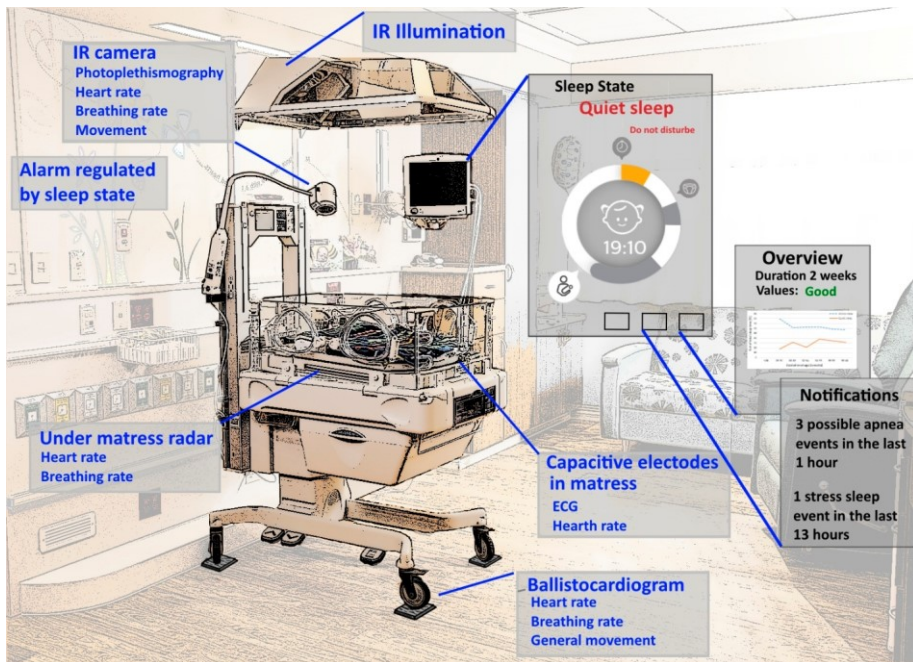
Figure 23 Visual illustration of a possible future NICU including unobtrusive vital sign measuring methods (Monitor graphic with courtesy of Mine Danisman-Tasar).
The vital signs could lead to a sleep state monitoring system reflecting the actual sleep state and additional information as, e.g., state development over longer periods, short term estimations, and/or warnings about sleep abnormalities.
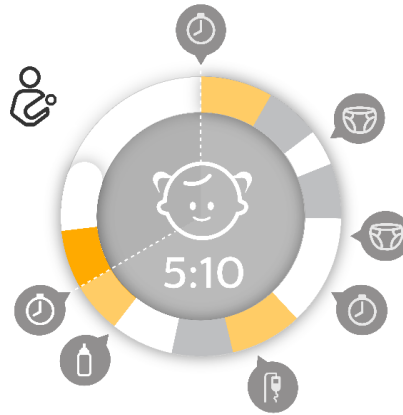
Figure 24 Detail of sleep monitoring visualization (Courtesy of Mine Danisman-Tasar).

The sleep states could be visualized over the past hours to get an overview of the state patterns of the neonate. Additionally caretaking, feeding, or other interventions can be highlighted, and optimal timeslots for caretaking can be suggested. Thereby caretakers can plan and coordinate their shifts according to the neonatal sleep and keep an overview over changes and incidences.

To conclude this chapter, Figure 23 illustrates a possible future NICU, including some of the discussed measurement features. The vital signs, measured with this method, can lead to a monitoring system, as in Figure 23 and Figure 24, presenting, e.g.,

- Actual and past sleep states
- Estimated time of state change
- Recommend time slot for optimal caretaking
- Long term display of sleep states with boundaries of "normal/expected" changes
- Display of sleep-related events (e.g., possible sleep apnea event)

2.9   Conclusions and recommendations

First, we want to repeat the most important points:

- Sleep plays a very important role in the development of preterm infants. Its disturbance can lead to negative long term effects.

- The variation of the sleep states AS and QS over time is an indicator of development.

- Sleep states can be separated in different ways. The classic method is EEG. However, EEG can separate the sleep states only at a later stage (clearly at 36 GA) compared with other PSG or behavioral methods (26-30 GA).

- There are several approaches for unobtrusive PSG and behavioral sleep state monitoring such as EEG/ECG measurements via capacitive electrodes or dry electrodes, as well as HR, HRV, BR, and movement monitoring via BCG, Doppler-radar, Doppler-laser, or camera monitoring.

- Sleep state separation should be performed via HR, HRV, BR, and/or movement. Multiple unobtrusive methods are available to measure these signals. Sleep states can be detected earlier with behavioral measurements than with EEG signals.

- The obtrusiveness of methods to obtain the vital signs for sleep state separation varies. The described methods differ in the way they come in contact with the preterm infant. Classic adhesive electrodes and cables are a burden for the preterm infant.

In order to rank the presented methods and give recommendations, we divide the discussed methods into the origin of the signal: electrical (EEG, ECG, EOG) and mechanical origin (HR, BR) plus behavioral signals (body movement, facial expression).

For measuring the signals of electrical origin such as EEG, ECG, and EOG, capacitive electrodes and Nanoneedle array seems to be optimal solutions. Capacitive

electrodes have good signal properties for ECG, while including EEG, the Nanoneedle array seems to be the better alternative. There is minimal harm involved in Nanoneedle arrays in contrary to the skin damage induced by classic adhesive electrodes. However, the acceptance of needle patches for preterm infants by the parents or legal guardians could be problematic due to the negative associations of needles with pain and discomfort.

For signals of mechanical origin, the measuring methods are quite similar in their (un)obtrusiveness and performance. BCG and radar are non-contact methods and do not need a free field of view. Radar has the drawback that, currently, in most cases, recalibration is needed for each individual change in the measurement setup; however, recently proposed algorithms can deal with this problem. For BCG, the small signal amplitude of preterm infants HR and HRV might be a problem, and it showed lower performance values in clinical trials than the other methods. Vital sign measurement via laser needs a free field of view on the blanket covering the neonate or the neonate itself to be able to operate. This is a drawback compared to the similar operating radar, favoring radar over BCG and laser. It remains to be investigated which of the methods deals best with the problem of motion artifacts. With the inclusion of behavioral observation for sleep state separation, the camera seems to be the optimal solution allowing simultaneous HR, BR, and behavioral measurements while additionally providing general visual surveillance. However, a free field of view is always required. If the camera is repositioned or blocked by, e.g., the staff, sleep monitoring is not possible. The incubator is often covered, and the lights in the NICU are frequently dimmed to minimize disturbance for the preterm infant. Therefore, IR illumination is needed to obtain a clear video image with sufficient resolution for a sleep state annotation algorithm or manual annotator.

In terms of costs, it is difficult to judge the above techniques, as most are still in the research phase and are not commercially available. Nevertheless, in most cases, the reusability will fix the expenses.

Most of the mentioned methods are currently not used in the NICU or only in first trials for preterm infants; this could be due to the long development and approval time for medical devices as well as the coherent cautious implementation of new approaches especially in the sensible and high-risk environment of the NICU.

Finally, each measurement method has different advantages and disadvantages, but the use of Nanoneedle patches, BCG, Doppler laser, Doppler radar, and/or camera enables to measure the vital signs HR and BR unobtrusively for preterm infant sleep monitoring. The performance of the methods compared to the golden standards are very high, making them excellent candidates for the use in the NICU. Further method-fusion could even increase the performance, to the end that, in the future, they possibly replace the standards without any loss off crucial information.

In the next chapters, we focus on ECG as a possible unobtrusive signal for sleep staging. However, before using unobtrusive measurement methods to obtain the ECG signals, we want to prove that ECG alone is suitable as a signal for a machine-learning algorithm to separate the main sleep states in preterm infants.

3   Machine learning on preterm infant sleep

3.1   Abstract

As an approach of unobtrusive assessment of neonatal sleep states, we aimed at an automated sleep state classification based only on heart rate variability. The signals are obtained from electrocardiography used for regular patient monitoring. We analyzed active and quiet sleep states of preterm infants between 30 and 37 weeks postmenstrual age. To determine the sleep states, we used a nonlinear kernel support vector machine for sleep state separation based on known heart rate variability features. We used unweighted and weighted misclassification penalties for the imbalanced distribution between sleep states. The validation was performed with leave-one-out-cross-validation based on the annotations of three independent observers. We analyzed the classifier performance with receiver operating curves leading to a maximum mean value for the area under the curve of 0.87. Using these sleep state separation methods, we show that automated active and quiet sleep state separation based on heart rate variability in preterm infants is feasible.

3.2   Introduction

Newborn infants show two distinct sleep states defined as AS and QS [117]. In full-term infants, AS is traditionally associated with REM, increased variability in cardiorespiratory rates, low muscle tone, and body movements in combination with specific continuous patterns of the EEG. In contrast, QS is associated with the absence of REM, decreased variability in respiratory rates, and fewer body movements in combination with a discontinuous EEG pattern. Even in very preterm infants, rudimentary sleep states can be identified from 26 weeks PMA [66].

The important role of sleep states in brain development is only beginning to be understood. It has been shown that sleep cycles are necessary for normal sensory and cortical development of the fetus and newborn [180], [181]. AS is important in providing the early stimulation and activity requirements of the growing brain. During AS, several organizational events take place such as the topographic alignment of the somatosensory, auditory, and the visual system and their connection to the cortex structures. [180], [181]

The time spent in AS and QS has been shown to be associated with maturation [43], [181], [182]. The distribution changes from 80% AS and 18% QS at early GA to around 60% AS and 30% QS at term age (see Figure 5) [183]. The NICU environment has a profound detrimental effect on sleep pattern development. Significant differences are found in sleep behavior between fetuses and preterm infants at the same postmenstrual age. It has been shown that preterm infants spend less time in AS and more in QS compared to fetuses at comparable age [178], [180], [184]. The difference can be related to the clinical condition of the preterm infant or to the interaction with the "hostile" NICU environment with a variety of noxious stimuli and painful procedures [184], [185].

Therefore, investigation of sleep states in preterm infants provides the opportunity to gain more insight into preterm brain development and identify which factors support or disrupt preterm brain development. Currently, PSG is considered the standard for sleep assessment. PSG employs audio and video recording of the infant as well as the typical recordings of respiration, HR, electromyography, electro-oculography, and EEG. However, the instrumentation required for these studies is only found in sleep laboratories and not in a typical NICU setting (Table 7).

Table 7 Review of literature on automated sleep staging.

This table gives an overview of the literature on automated sleep staging, from 1987 until 2017. Abbreviations: postmenstrual age (PMA) [gestational age + postnatal age], polysomnography (PSG), electroencephalography (EEG), electrooculography (EOG), electromyography (EMG), active sleep (AS), quiet sleep (QS), wake (W), sleep (S), heart rate (HR), heart rate variability (HRV), respiratory rate (RR), linear discriminant analysis (LDA), antedependence models (AM), learning vector quantization (LVQ), multi-layer perceptron (MLP), support vector machine (SVM), area under the curve (AUC).

Author Year	Population, PMA [wk]	Gold standard Annotations	Sleep states of interest	Used signals for analysis	Method	Results
Harper 1987 [186]	39-41	Manual scoring based on PSG	AS, QS, W	HR, RR	LDA	Agreement for separation of AS-QS-W HR: 82%; RR: 80%; HR+RR 85%
Haddad 1987 [105]	44-56	Manual scoring EEG, EOG, chin EMG and behavior	AS, QS	HR, RR	Kolmogorov Smirnov distances	Agreement for separation of AS and QS AS 99% and for QS 93%
Sadeh 1995 [103]	40-84	Manual scoring respiration and behavior	AS, QS, W	Actigraphy	LDA	Agreement for separation of AS-QS-W 75-87% depending on PMA, and for S-W 89-98% depending on PMA
Nason 2001 [187]	48-60	Manual scoring EEG, EOG, ECG, chest and abdominal movement	S, W	HR	LDA; AM	Agreement for separation of S and W LDA, 75%-90% depending on PMA AM: 89%-96%
Lewicke 2004 [188]	33-58	Manual scoring EEG, EOG, EMG	S, W	HR Actigraphy	Neural network (LVQ)	Agreement for separation of S and W HRV: Sleep: 90% Wake: 57% Actigraphy: S 92%, W 42%

Lewicke 2008 [189]	39-53	Manual scoring EEG, EOG, EMG	S, W	HR	Neural network (MLP; LVQ); Non-linear SVM	Agreement for separation of S and W MLP, S 86% and W 85% LVQNN, S 89% and W 80% SVM, S 90% and W 79%
Fraiwan 2011 [190]	40	Manual scoring based on PSG	AS, QS, W	EEG	Time-frequency analysis	Agreement for separation of AS-QS-W 63%-75%
Terril 2012 [111]	48-84	Manual scoring based on PSG	AS, QS, W	RR	Time-frequency analysis	Agreement for separation of AS-QS-W 80%-85%
Palmu 2013 [54]	25-32	Manual scoring EEG, EOG, chin EMG	S-W	EEG	Time-frequency analysis	Agreement for separation of S-W, extracted from table 90%
Isler 2016 [191]	37-44	Manual scoring EEG, EOG, chin EMG , respiration, behavior	AS, QS	RR variability	Time-frequency analysis	Agreement for separation of AS-QS AS 78%-90%, QS 87-100%
Dereymaeker 2017 [102]	27-42	Manual scoring EEG and video	QS	EEG	Time-frequency analysis	AUC 0.97 for detecting QS
Koolen 2017 [101]	24-45	Manual scoring EEG	AS, QS	EEG	Non-linear SVM	Accuracy for separation of AS and QS Accuracy: 85%, sensitivity 83%, specificity 87%

Furthermore, PSG requires the placement of multiple electrodes and sensors that are not tolerated by the skin of the vulnerable preterm infant.

Recent advances in technology allow to collect a variety of physiological data in the clinical setting and to process and analyze these in an automated fashion. Various methods have been explored to develop sleep state separation techniques that require only a subset of standard PSG measures [190]–[192]. Automated sleep staging (or sleep state classification) based on, e.g., HRV is already successfully implemented in adults [68], [193]–[195]. For newborns, however, automated sleep scoring is still in the exploration phase, while the development of stable EEG based algorithms is only recently emerging

The first research was introduced in the late eighties by Harper et al. [186] and Haddad et al. [105]. Harper et al. used cardiorespiratory signals with a discriminant analysis on 25 term infants over a period of 6 months to separate AS, QS, and wake. They created different models depending on age and achieved an overall agreement with the manual observations of 85%. Haddad et al. [105] exceeded the results of Harper et al. classifying only AS from QS based on respiratory variability with an accuracy of 99% on detecting AS and of 93% on detecting QS using Kolmogorov Smirnov distances. These good results might be explained by the age of the subjects, varying from 44 to 56 weeks PMA. Sleep state separation becomes easier with increasing maturation as each sleep state becomes more pronounced and can be separated more clearly. This was also found by Sadeh et al. [103] who separated AS, QS, and wake only using actigraphy. They created several movement-based features that were analyzed with linear discriminant analysis (LDA) for 41 term infants with age ranging from birth (term) to one year of age. The classification accuracy increased over the course of development from 89% to 97% for sleep and wake distinction. Nason et al. [187] confirmed this observation when they used wavelet analysis in combination with LDA and antedependence models (AM) to separate sleep and wake on one subject over a duration of 4 month. Performance increased over age from 75-90% with an LDA and 90-96% with AM.

In 2004, the group of Lewicke and Schuckers used the CHIME study for sleep staging in term infants based on HRV. Lewicke et al. [188] first compared the use of HRV against actigraphy for sleep staging with a learning vector quantization (LVQ) neural network. They reported that the use of HRV resulted in a correct detection of sleep in 90% and wake in 57%, respectively. The use of accelerometer measurements led to 92% for sleep and 42% for wake detection, respectively. The lower agreement for wake could be explained with the use of accelerometer measurements, which might not detect wake episodes with less or no movement. In addition, the generally lesser amount of data on wake episodes in term infants can reduce neural network performance as it is directly linked to the quantity of training data. In a second study [189], they applied two additional classification methods together with the LVQ on an extended data set of 190 early term infants. In that study, they used only HRV as input for the LVQ, Multilayer perceptron neural network, and a support vector machine (SVM). With a huge amount of 57000 30s epoch for each training, test, and validation set, they were able to increase the correct prediction for wake to 80%. The SVM created the highest scores with a detection accuracy of 90% for sleep and 79% for wake. This was the first time that such an amount of data was analyzed for automated sleep staging for early term infants. It could be postulated that this was the first stable sleep staging algorithm for early term infants. Automated analysis of preterm infant sleep using cardiorespiratory signals was published in 2016. Similar to Haddad et al., Isler et al. [191] created a threshold-based algorithm where the threshold was derived from the normalized instantaneous breathing rate variance and respiration variability. By replicating the manual scoring process and tailoring the analysis specifically to the dataset, they reached a 100% agreement with the observer annotations. The more general performance of their method ranges from 78% to 92% agreement with the observer annotations[54]

In 2017, Dereymaeker et al. [102] and Koolen et al. [101] published a classification algorithm of preterm infant sleep states using EEG signals. Dereymaeker et al. focused mainly on identifying QS, where they based their analysis on the heightened

discontinuity of the EEG during QS. Using a cluster-based adaptive sleep staging method, they achieved very high results from the Receiver Operating Characteristic (ROC) with an area under the curve (AUC) of 0.97 for detecting QS from other sleep states out of preterm infant data stream. Koolen et al. used six features on a nonlinear SVM classifier, resulting in an AUC of 0.83 for preterm infants <32 weeks PMA and 0.87 for infants >32 weeks PMA. A more complete overview of sleep state classification based on EEG is given by Dereymaeker et al.[196].

As most previous studies were based on term newborns, in this paper, we aimed to investigate the feasibility of classifying AS and QS based on HRV only for preterm infants. This would be complementary to the studies by Koolen et al. and Dereymaekers et al. (focusing on EEG analysis) as well as the work by Isler et al. (focusing on respiratory analysis).

3.3 Methods

3.3.1 Population

In this retrospective study, we analyzed eight healthy and stable preterm infants born at a mean gestational age of 30 $\pm$ 2.6 weeks, who were studied at a mean postmenstrual age of 34 $\pm$ 2.8 weeks. The mean birth weight was 1644 $\pm$ 309 g. More details can are presented in Table 8.Table 8 The infants were admitted to the neonatal department at Máxima Medical Center Veldhoven, the Netherlands. The clinical staff asserted the medical condition of the patients. Ethical approval was given by the medical ethical committee of the hospital, and written consent was given by the patient's parents.

Table 8 Patients weight, gestational age [GA], and postmenstrual age [PMA].

Patient	Weight [g]	GA [wk; d]	PMA [wk; d]
1	1845	31; 5	33; 6
2	2265	33; 6	34; 6
3	1740	28; 6	34; 6
4	1235	30; 1	31; 2
5	1700	32; 6	33; 5
6	1290	28; 6	29; 3
7	1460	25; 4	31; 5
8	1615	29; 6	32; 0

3.3.2   Annotation

Data were annotated by three independent observers for the following neonatal sleep states adhering the Prechtl system [197]: AS, QS, active- and quiet wake, vocalization, and body positions. The mean annotation time per infant was 334 ± 54 min with simultaneous video recording, ECG, and thorax impedance. After removal of caretaking episodes, the recordings were divided into 30s epochs, and sleep states were assumed to be coded correctly if at least two annotators agreed. The total number of annotations was 4057 30s epochs, of which 845 were discarded because of signal corruption or disagreement between observers. Of the remaining 3212 annotated 30s epochs, 283 had other sleep states than AS or QS. For the analysis, a total of 2929 30s epochs remained with AS (n=2617) or QS epochs (n=312). The Cohen's kappa ($\varkappa$) coefficient of agreement representing interrater variability ranged between 0.61-0.80 for AS (substantial) and between 0.41-0.60 for QS (moderate) [198].

3.3.3   Data acquisition

The ECG was recorded with three standard leads. The recording device was a Philips monitor (IntelliVue MX 800, Germany) using 250 Hz (n=2) and 500 Hz (n=6) sample frequency for the ECG. Sample frequencies of 250 Hz and higher are sufficient for linear HRV analysis [199]. From the ECG signal, the R peaks were detected with an in-house developed algorithm [200]. More on the analysis is

described in section 4.3.5. Subsequently, the R-R intervals were used for HRV feature analysis.

3.3.4   HRV features

We selected 20 HRV features using 18 commonly used features in adult sleep analysis [68], [194], [201], and two preterm infant specific frequency-domain features. The adult frequencies are divided into very low frequency (VLF), low frequency (LF), and high frequency (HF). We refer to Table 9 for details. To create the two additional premature frequency ranges we extended the HF feature to three HF features: HF, pHF1, and pHF2 with a standard adult frequency range from 0.15 to 0.4 Hz for HF, and additional ranges of 0.45-0.7 Hz for pHF1 and 0.7-1.5 Hz for pHF2 [202] to accommodate the increased cardiorespiratory rates in preterm infants. A sliding window of different lengths (1; 2.5; 5 min) centered on each 30s epoch was used to evaluate if different windowing gives additional information for separating sleep states in preterm infants.

R-peaks in ECG, needed for the HRV, appear inherently non-equidistant in time. To avoid introducing extra parameters in resampling the signal, the Lomb-Scargle algorithm for spectral analysis was applied [203].

Table 9 Heart rate variability features.

The HRV features are derived from the Task Force [201], and comments are added from neonatal studies.

Feature [unit]	Description	Connotation
BpE	Beats per Epoch	ECG R-R intervals detection. If ECG is noise distorted, BpE decrease. This mostly appears during AS and wake. Longer heart rate is reflected when long term windowed.
TotPow [ms^2]	Total power or variance of NN intervals of a defined window size.	Reflects overall heart rate variability [201], [204]
VLF [ms^2]	The power of the very-low-frequency band between 0.003-0.04 Hz of a defined window size.	Oscillations in VLF are attributed to peripheral resistance fluctuations caused by thermoregulation [205].
LF [ms^2]	The power of the low-frequency band between 0.04-0.15 Hz of a defined window size.	LF fluctuations are assumed to be related to baroreflex activity and under the parasympathetic and sympathetic influence [205], [206]. Fluctuations in the neonatal baroreceptor loop are at approximately 0.07 Hz [122], [125], [206].
LFnorm [%]	LF power in normalized units LF/(Total Power-VLF) x 100	Normalization, to correct for total power variability.
HF [ms^2]	The power of the high-frequency band between 0.15-0.4 Hz of a defined window size.	HF fluctuations are associated with activities of the parasympathetic system and respiratory activity [122], [205], [207]. Respiratory activity is closely linked to preterm sleep states [183], [191] and seems more prominent during quiet sleep [207].
HFnorm [%]	HF power in normalized units HF/(Total Power-VLF) x 100	Normalization, to correct for total power variability.
pHF1 [ms^2]	The power of the high-frequency band between 0.4-0.7 Hz	pHF1 fluctuations are associated with activities of the parasympathetic system and respiratory activity, especially in preterm infants [202].

pHF2 [ms^2]	The power of the high-frequency band between 0.7-1.5 Hz	pHF2 fluctuations are associated with activities of the parasympathetic system and respiratory activity number, especially in preterm infants [202].
LF/HF [n.u.]	Ratio LF/HF	This estimate claims to reflect the sympathovagal balance in adults, although the value has to be established in newborns [122]. Increased values may indicate greater sympathetic and/or lesser vagal modulation [206].
SDNN [ms]	The standard deviation of normal to normal R-R intervals of a defined window length.	Reflects the overall heart rate variability influenced by both the para- and sympathetic nervous system [201], [204].
RMSSD [ms]	Root mean square of successive differences between adjacent R-R intervals of a defined window length.	Influenced mainly by parasympathetic activity and respiratory activity.
NNx [count]	The number of pairs of successive R-R intervals that differ by more than 10, 20, 30 or 50 ms of a defined window length.	NNx reflects parasympathetic activity. While NN10 covers more overall changes, NN50 represents high-frequency variations with influence from respiratory activity [208].
pNNx [%]	The proportion of NNx divided by the total number of R-R intervals of a defined window length.	pNNx are directly linked to the NNx features. pNNx for values of x<50 ms may provide more robust estimates of cardiac vagal tone modulation even in the presence of outliers [208], [209].

3.3.5   Feature selection

Supervised learning for classification with an SVM needs a selection of normalized input features to avoid overfitting resulting in a decrease of performance. We implemented a linear and non-linear classification approach, including three feature selection methods (two filtering methods and a wrapper method), which were used to find the top feature subset separating AS from QS. First, the correlation-based feature selection (CFS) was applied, and second, because the classes are imbalanced,

the minority based feature selection method (FSMC). For both filtering feature selection methods, a greedy forward search was used to find the optimal feature subset. The CFS is sorting the features by the highest correlation between feature and class as well as the lowest correlation in-between features [210]. The FSMC is determining the difference between feature values for the majority and minority class. It sorts the features by the highest difference between the values for the majority and minority class [211]. Third, a mixed wrapper method was implemented (Figure 25). Starting the wrapper, a brute force forward search was used where all possible subset combinations of all features (without repetition) were generated, and for each subset, the sleep state separation performance was validated. With a search for highest performance, the optimal subset was determined. Due to the exponential increase of computation time per subset combination length (generation), subsets combinations
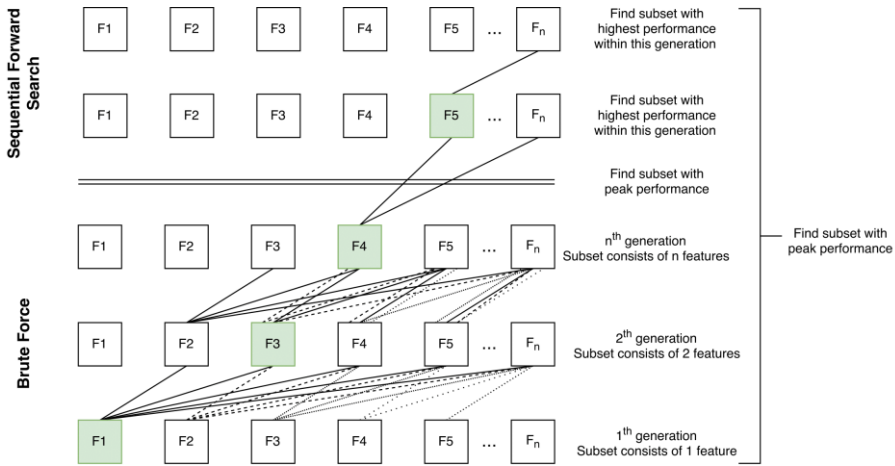
Figure 25 The brute force method creates subsets with feature combinations of only one feature (1th generation) up to combinations of n features (nth generation). Thereby, features in each subset are not repeated (e.g., F1, F2, F1) and the order is ignored (F1, F2 = F2, F1). Afterwards, all created features are evaluated for their performance and the feature set with highest performance is chosen. The sequential forward search (best first) adds in each iteration one feature to the chosen subset, finds the highest performance in this iteration and continues to the next iteration. Finally, the subset with the overall highest performance is chosen to proceed.

of all features were limited to the seventh generation (137979 combinations/iterations in total).

The remaining 10 generations of subset combinations were analyzed with a sequential forward search (best first), where only the next single best feature was added to the previous subset combination (105 iterations). The final subset with the maximum performance out of all combinations was determined.

3.3.6   Classifier

After the feature selection and within the wrapper method, an SVM classifier, embedded in Python`s Scikit-learn library [212], was used to train and test on the dataset. SVM is a widely used classifier [213], which is searching for hyperplanes separating the different classes with a maximum margin under the condition that misclassifications are minimized. Thereby, misclassification is penalized with a constant C, which is multiplied with the number of misclassifications. A large C value will increase the misclassification penalty (decreasing the misclassification), leading to a decrease of the minimum margin around the hyperplane separating the classes and vice versa. If the C value is chosen to be large, the generalization of the SVM model can get weakened. For this study, the C value was set to value 3.6 after a parameter search as a good balance between speed, accuracy, and generalization.

As the data is unbalanced for AS and QS states, the C value can additionally be multiplied with value pairs representing the distribution of a dataset to compensate for the class imbalance. In preterm infants the class distribution change over time and development, therefore, we clustered the participants into four different GA ranges and computed the class weight pair for each cluster not on the actual data distribution per cluster, but on the gross expected distribution known from the literature [183] as seen in Figure 5. The subjects were ranked into the following clusters in GA: 31-32 weeks (n=2), 33-34 weeks (n=1), 35-36 weeks (n=2), and 37-39 weeks (n=3). This

clustering should not be mistaken for an age clustering to create different feature/training sets, which is not feasible due to the limited population size.

Within the wrapper feature selection method, a nonlinear SVM classification was used. The SVM is optimally extendible for nonlinear tasks using the "kernalisation" or "kernel trick" where the training data are transformed into a higher dimensional feature space where the problem then becomes linear again [214]. The applied SVM used the radial basis function (RBF) kernel for non-linear approximation. The RBF kernel also uses a free parameter γ, which determines the inverse influence of the support vectors. We used a γ value of 0.2.

A leave one out cross-validation (LOOCV) was applied to attest the classifier performance. Cross-validation methods separate the total data into different parts of which one is used to validate the classifier and the remaining parts to train. Nevertheless, for small datasets, these methods can become biased as it is highly probable that training and testing sets are created from the same patients. To avoid a biased validation, the LOOCV uses the patients itself as separation. All but one patient are chosen for training and the remaining patient for testing. For a more detailed description, we refer to the literature [213].

3.3.7   Statistical analysis

For continuous distributed data, median and interquartile range (IQR) were calculated. Data were analyzed with Matlab 2014b software program (MathWorks, Natick, Massachusetts, US). The performance of sleep state separation was calculated with the receiver operating curve (ROC) and the corresponding area under the curve (AUC).

3.4   Results

In **,** the median of all 20 HRV features for AS and QS are presented together with interquartile range with 25 and 75 percentiles for comparability with the literature.

We used 1, 2.5, and 5 min windows to investigate the effects of window length in preterm infant sleep staging. The combination of different window sizes did not increase the classifier performance. On the contrary, the analysis of only 30s epoch based HRV features lead to a lowered sleep staging performance of 0.71. Hence, we used the 5 min window length for sleep state separation and classifier methodology recommended by the Task Force [201]. To evaluate the linear approach of sleep state separation, we implemented the CFS and FSMC filter methodologies. The performance of the optimal subset was evaluated with a linear SVM classifier and resulted in a mean AUC of 0.32 ± 0.16 for CSF. The FSMC method could not identify any suitable features for a subset conformable with its selection rules. These results showed that sleep state separation with a linear kernel is not feasible for this set of features.

After exclusion of sleep state separation with a linear kernel, we implemented a wrapper solution with a nonlinear kernel SVM. The optimal subset was obtained by training the SVM on seven preterm infants, while one is left out for cross-validation (Figure 26 and Figure 27). The performance analysis of sleep state separation was calculated with the ROC and AUC for each iteration of the wrapper and the following sequential forward search. The ROCs per patient and mean ROC of the chosen subset is shown in Figure 26, with a mean AUC value of 0.85 ± 0.46. As the sleep states were unbalanced, we adapted the classifier parameter C with weighting factors pairs calculated from the expected distribution, which grossly corresponds with the four age cluster (see Chapter 3.3). Using the different class weights, we achieved a better performance of the sleep staging with an AUC of up to 0.87 ± 0.42 (Figure 27). The optimal subset features resulted from the wrapper analysis for both classification types were BpE, NN20, SDNN, pNN20, and total power.

Features	Units	AS Median	AS IQR	AS Percentile 25	AS Percentile 75	QS Median	QS IQR	QS Percentile 25	QS Percentile 75
BpE	counts	770	49	749	798	745	72	704	776
NN10	counts	118	159	44	203	90	178	8	186
NN20	counts	37	95	11	106	10	35	-	35
NN30	counts	21	64	3	67	2	10	-	10
NN50	counts	10	33	-	33	-	5	-	5
RMSSD	ms²	18	30	8	38	12	16	6	22
SDNN	ms²	24	16	16	32	13	11	8	19
pNN10	%	16	21	7	27	13	25	1	26
PNN20	%	5	13	2	15	1	5	-	5
pNN30	%	3	8	1	9	10	1	-	1
PNN50	%	1	4	-	4	-	1	-	1
HF	ms²/Hz	38,962	63,774	16,817	80,591	15,898	28,401	3	31,289
Hfnorm	%	12	8	8	16	10	6	7	13
LF	ms²/Hz	117,274	148,547	59,518	208,065	49,878	125,967	22	147,946
Lfnorm	%	39	41	21	62	55	38	32	70
VLF	ms²/Hz	258,488	380,003	114,717	494,720	57,558	87,281	31	118,391
ratioLFHF	n.u.	3	4	2	6	5	5	2	7
pHF1	ms²/Hz	26,039	80,965	6,262	87,227	9,764	24,230	1	25,335
totpower	ms²/Hz	687,347	897,832	303,403	1,201,235	204,300	364,824	76	441,208
pHF2	ms²/Hz	59,147	271,635	10,071	281,707	8,087	25,003	2	27,097

Table 10 This Table presents the unscaled heart rate variability features for active sleep (AS) and quiet sleep (QS) as median, interquartile range (IQR) and [25th, 75th] percentiles. For feature abbreviations, please see Table 9.
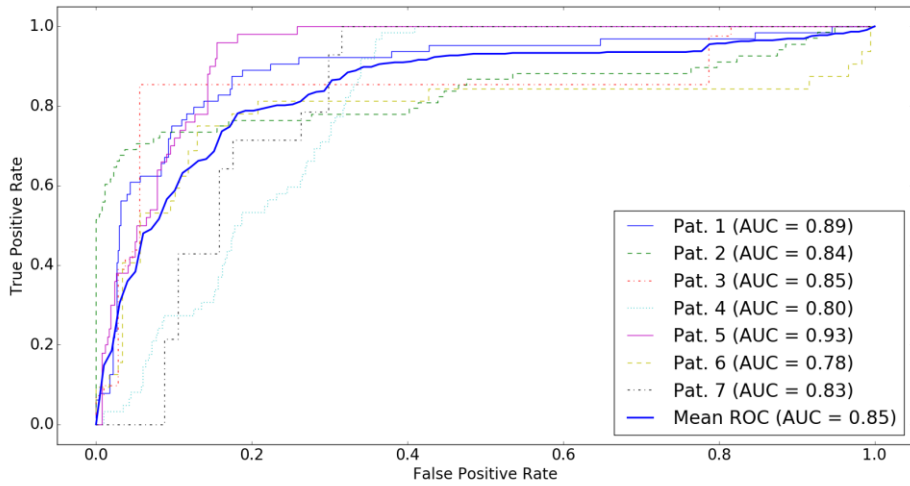
Figure 26 The "Receiver-Operating-Characteristic" for the eight fold leave one out cross validation of SVM sleep state separation. Patient 8 showed only active sleep and therefore this data was included in testing and training, but the ROC could not be calculated.
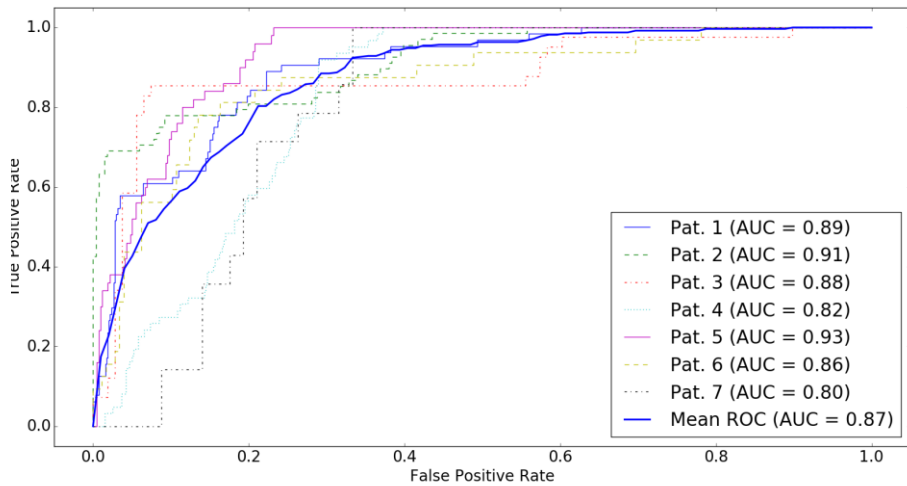
Figure 27 The Receiver Operating Characteristic (ROC) for the eight fold leave one out cross validation of the Support vector machine (SVM) sleep state separation. Here we used adapted missclassification penalty value pairs for parameter C to counter class imbalance. Patient 8 showed only active sleep and therefore this data was included in testing and training, but the ROC could not be calculated.

97

3.5   Discussion

HR is an important physiological parameter for sleep monitoring in newborn infants, children, and adults. In this chapter, we explored the feasibility of sleep state separation by automated analysis of HRV features obtained from standard patient monitoring. A nonlinear kernel support vector machine for sleep state separation between AS and QS was used. The classifier performance with receiver operating curves resulted in a mean value for the AUC of 0.85 (Figure 26) and under consideration of sleep state distribution, an AUC of 0.87 (Figure 27), indicating that HRV features are valuable for automated sleep state separation in preterm infants.

The early postnatal period for preterm infants is characterized by fluctuating periods of AS and QS states, with intermediate/undetermined sleep phases. These fluctuating states are associated with characteristic cardiorespiratory variability [43]. Although cardiorespiratory coupling is weak in very preterm infants, studies conducted in more mature infants show the presence of cardiorespiratory coupling [97]. In **,** the different HRV values for AS and QS per feature are presented. Overall, the high IQR of the various HRV values may reflect an immature autonomic nervous regulation. Also, recovery to a stable state of autonomous control after any disturbance is reflected by a relatively long time constant of approximately 10 minutes [215]. Our findings are comparable with other preterm infant HRV studies [206], [207], [216], and the time domain features show less complexity in HR regulation during QS, reflected in lower median values and IQR. As the respiratory sinus arrhythmia in preterm infants is less dominant than in term babies [207], the median increase with AS in the time domain can probably be linked to the dominance of the sympathetic nervous system in preterm infants [201], [217]. This might originate from the not fully developed adrenergic receptors in the sinus node [207]. It has been suggested that the limited cardiorespiratory coupling is stronger during QS than during AS [207]. This enhances the difference in median and IQR as more regular breathing patterns during QS reduce the signal complexity while the irregular breathing during AS does show only

limited influence. The predominance of the sympathetic nervous system is associated with an increase in LF power [205], increasing the feature values for LF and LFnorm compared to HF and HFnorm. However, also vagal modulation of baroreflex response to blood pressure disturbances is important [122]. In addition, the baroreceptor reflex sensitivity increases during maturation [125] with a reduced high-frequency contribution to HRV in preterm infants. Like other studies, the LF/HF ratio showed slightly higher values for QS, probably indicating a greater sympathetic or less vagal modulation [206]. In general, the spectral power values were decreased in QS compared to AS, suggesting that the autonomic modulation is lower during QS [216]. This could be explained as the AS state is seen as the most basic state which is regulated by a network of several forebrain areas and the controlling brainstem [52] and therefore, earlier developed and more pronounced than the QS [123].

3.5.1 Features

In general, the approach to automatically separate sleep states using a cardiorespiratory signal consists of a feature extraction approach, followed by a classification step. The features we chose are derived from adult sleep analysis [204], [218], as the main objective was to prove the feasibility of separating sleep states in preterm infants based on HRV. We also added two additional features with increased frequency ranges to accommodate the general higher cardiorespiratory rates in preterm infants. The newly used frequency ranges for preterm infants (See Table 9) were proposed in 2008 by Indic et al. [202]. The new feature pHF1 consistently appeared in the top feature subsets underlining the importance and influence on sleep state separation and also the assumption of Indic et al. that higher cardiorespiratory rates in preterm infants need adapted frequency ranges seems to be valid. This is not surprising as other groups already explained [43] and demonstrate [111], [191] the benefit of respiration analysis for sleep staging. The top five feature subsets all resulted in comparable performance (AUC: 0.85-0.87). In those top five feature subsets, total power and SDNN were always present. pHF1, NN20, NN30, and pNN20 were

present three out of five times. The time-domain feature BpE appeared only in the top feature subset.

To further increase classification performance and stability, recently published novel term [219] or preterm infant HRV features [220] can be implemented in future research. In addition, the preterm infant desaturation features presented by Kommers et al. [208] can be used to eliminate episodes of tachycardia, avoiding misinterpretation of autonomous activity due to, non-sleep-related, altered HRV. Also, instead of eliminating noise from the signal, it could rather be used as an additional information source supporting sleep state separation [85].

3.5.2 Sleep states separation

The nonlinear SVM kernel was chosen based on a good separation performance with a mean AUC of 0.85 to 0.87 (Figure 26 and Figure 27). The SVM classifier was selected as it is robust against outliers while showing high performance for single and multiclass classification problems [221]. Outliers have to be expected as the sleep state annotation is challenging by various covariates. The age clustering for compensating changing state distributions increased the performance. Nevertheless, while the performance improved for most subjects (4), some did not change (2), and one patient decreased. This could be interpreted as hitting the right cluster improves the classification, while a false clustering can lead to a decrease in performance. In our opinion, only clustering on age is the reason for the possibility of a decrease in performance. Commonly, it is postulated that sleep state changes over gestational age [183]. This is certainly correct, but age is only one indicator of neural development. The coupling between age and development (sleep state distribution) can change with neural miss-development. Therefore, we suggest that the correct clustering has to be determined and based not only on age but rather overall condition, including weight and size and other biomarkers.

Nonetheless, also biomarkers that are not directly linked to neural development can give miss-information and potentially create false sleep staging, which consequently would lead to false results of development monitoring. Generally, it can be said that sleep state separation without additional background information should be aspired to.

3.5.3   Unobtrusive HRV measurement

In this study, adhesive ECG electrodes were used as part of standard care and patient monitoring. As described in Chapter 2, there have been innovative developments in obtaining unobtrusively cardiac signal measurements from which HRV can be derived. In general, the unobtrusive HRV measurements can be classified as contact and non-contact methods. The contact methods include a variety of options, including a neonatal jacket embedded with smart textile electrodes for ECG measurements [222], a neonatal snuggle embedded with reflectance photoplethysmography based on near infra-red spectroscopy technology for pulse oxygenation monitoring [223], and intelligent bedsheet embedded with polyvinylidene fluoride or electromechanical film sensors for ballistocardiography measurements [224]. These contact ECG sensors are suitable for the sleeping scenario by integrating into a bed sheet or a mattress. The non-contact methods include HRV extracted from thermal imaging, video analysis, Doppler effect, and capacitively coupled ECG [171], [173], [183], [225]. The imaging and Doppler methods could be embedded into a neonatal incubator for non-contact measurements. For example, the sensory neonatal jacket and snuggle provide a natural platform for seamlessly embedding sensors for unobtrusive measurements. A major limitation of most unobtrusive methods is the sensitivity to motion artifacts deteriorating the signal quality. In general, the type of unobtrusive HRV method depends on various factors, such as the neonatal sleep monitoring scenario, the environments of incubator inside NICU, the reliability of measurements, and the suitability for long term sleep monitoring.

3.5.4 Methodological limitations:

The small sample size of analyzed preterm newborns is the main limitation of this study. Also, here we included only AS and QS states. However, AS and QS are the dominant states in the early weeks after preterm birth, and other states are less well defined in preterm infants. The objective of the study was to investigate the feasibility of separating AS and QS, which are most important for development monitoring, based on HRV measure alone. In follow up studies, described in Chapters 4 and 5, we integrated multi-stage analysis.

Further, the dataset was reduced by 28% due to unused sleep states, corrupted data, and disagreement between observers. As the data was not reduced randomly, the sleep states AS and QS were afterward in a slightly shifted distribution of 89% to 11%. This could have affected the classification if too much of the QS state would have been disregarded. However, the QS state is still in the expected range, and the imbalanced distribution with an inflated AS was corrected with weighted misclassification penalties.

To increase the performance of sleep state separation, several demographic variables such as gestational and postnatal age could be taken into consideration as these variables influence the HRV [226]. Note that the methodology of gestational age clustering in this study was used only for consideration of the sleep state distribution, not for age determined feature creation and training sets. In our case, age clustering for different classification was not feasible due to the small dataset. Clustering the data would have further decreased the training data. Finally, this study used mainly the (adult) recommendations from the Task Force as there is no consensus in newborns. The European and North American Task Force was formed in an attempt to set standards for future studies of HRV [201]. While important recommendations were made for the length of recordings and the required spectral indices, the conclusions were based on adult studies only. As the neonatal heart and breathing rate differs from adults, the recommendations of the Task Force may not be

applicable in preterm infants. However, until recommendations for neonatal standardized analytical methods are made, many fetal and neonatal studies use the recommendations of the Task Force [227], [228]. We followed this approach by using mainly the adult recommendations from the Task Force. Nevertheless, to account for the increased cardiorespiratory rates in preterm infants, we used two additional frequency-domain features, pHF1, and pHF2 [202].

3.5.5   Future recommendations

For a stable performance based on HRV features only and over a wide range in the preterm infant population, additional data is needed. For possible clinical application and a more holistic view on preterm infant sleep, the classification between all states should be considered. As earlier research showed good performance of respiratory analysis for sleep staging, we recommend using cardio and respiratory features in combination to achieve higher or more stable performance. Finally, as the R peak detection is essential for correct classification with HRV, we suggest to investigate and validate R peak detection algorithms specifically for preterm infants.

3.6   Conclusion

This study shows that using a nonlinear SVM classifier approach for HRV features provides good results for preterm infant sleep state analyses of AS and QS. While our findings cannot yet be seen as robust due to the limited population size, the classifier performance can compete with the literature [101], [111], [189]. Merging the different vital sign approaches, e.g., respiration [191] and activity [104] with HRV, will most likely lead to a robust, unobtrusive, and automated methodology for continuous preterm infant sleep monitoring.

As we could show that machine learning, in general, is capable of preterm infant sleep classification, we want to examine if similar results can be obtained using only genuinely unobtrusive signal modalities. In the following chapter, we use capacitive

ECG in combination with machine learning to separate preterm infant sleep states. We also consider different machine learning algorithms and include all states into the classification task.

4 Use of unobtrusive ECG measurement for the use with machine learning

4.1 Abstract

To date, mainly obtrusive methods (e.g., adhesive electrodes in electroencephalography or electrocardiography) are necessary to determine the preterm infant sleep states. As any obtrusive measure should be avoided in preterm infants because of their immature skin development, we investigate in this chapter the possibility of automated sleep staging using electrocardiograph signals from non-adhesive capacitive electrocardiography. Capacitive electrocardiography data from eight different patients with a mean gestational age of 30 ± 2.5 weeks are compared to manually annotated reference signals from classic adhesive electrodes. The sleep annotations were performed by two trained observers based on behavioral observations. Based on these annotations, the classification performance of the preterm infant active and quiet sleep states from capacitive electrocardiography signals showed a Kappa value of 0.56 ± 0.20. Adding wake and caretaking into the classification, a performance of Kappa 0.44 ± 0.21 was achieved. In-between sleep state performance showed a classification performance of Kappa 0.36 ± 0.12. Lastly, a performance for all sleep states of Kappa 0.35 ± 0.17 was attained. Capacitive electrocardiography signals can be utilized to classify the central preterm infant sleep states, active and quiet sleep. With further research based on our results, automated classification of sleep stated can become an essential instrument in the future intensive neonatal care for continuous brain maturation monitoring. Especially, being able to use capacitive electrocardiography for continuous monitoring is a significant contributor to reducing disruption and harm for this extreme fragile patient group.

4.2   Introduction

Preterm infant sleep is strongly connected to brain maturation. A more in-depth explanation of the importance of sleep on brain maturation and the classification of sleep in preterm infants is reviewed in the introduction of this thesis. As a short refresher, sleep in newborns can be separated into three sleep states, active sleep (AS), quiet sleep (QS), and intermediate sleep (IS) and wake and/or caretaking. AS is known to activate neural activities and is essential in synaptogenesis of the neural interconnections. AS is the most dominant state, with around 70% of the total sleep time in the first weeks after birth. During QS, neural activity is also seen but less than in AS. QS is mostly described as a resting or reenergizing state. Also, developmental errors are corrected during QS using the heightened brain plasticity of the preterm infants to reorganize the brain structure. In the preterm brain, the time spent in AS and QS fluctuates quickly and may be difficult to separate during the first weeks after birth. Many episodes cannot be explicitly allocated to one specific state, and therefore IS is more prominent at that early state in preterm infants. The distribution of the sleep states can be a biomarker of brain maturation.

In clinical practice, sleep staging is mainly based on manually annotated electroencephalogram (EEG) and/or behavioral analysis. These practices are time-consuming and not continuous. Therefore, several research groups work on automated sleep staging algorithms for preterm infants [101], [102], [110], [191], [229], [230]. To date, the research is mainly focused on EEG and electrocardiography (ECG) analysis. For both signal types, standard adhesive electrode settings are used. As mentioned in Chapter 2, all electrodes in contact with the fragile preterm infant skin are considered obtrusive [108], [183]. The epidermis of a preterm infant under 32 weeks' of gestation is only two to three layers thick with almost no protective outer skin. Disrupting the immature epidermis results in an increased risk of infection, healing, and scarring [107].

In this chapter, the focus is on the use of ECG measurements from capacitive ECG (cECG) electrodes to determine different sleep states in preterm infants automatically. It will be determined which R-peak detection method is superior for cECG analysis. Secondly, it will be investigated how well the previously used features perform in comparison to newly implemented features for AS and QS separation. Then, it will be investigated how multi-class classification performs comparing the use of the ECG and cECG signals. Finally, we examine the difference in the successfully used features for the ECG and cECG classification.

4.3 Methods

4.3.1 Population

In this retrospective study, eight stable preterm infants born with a mean gestational age (GA) of 30 ± 2.5 weeks were analyzed. These are not the same patients as researched in chapter 3. They were studied at a mean postmenstrual age (PMA) of 32 ± 2.6 weeks. More details can are presented in Table 11. The patients had a mean birth weight of 1652 ± 565 g. They were admitted to the neonatal intensive care unit (NICU) of the neonatal department at the Máxima Medical Center Veldhoven, The Netherlands. Ethical approval was given by the medical ethical committee of the hospital; written consent was given by the patient's parents.

Table 11 Patients weight, gestational age [GA], and postmenstrual age [PMA].

Patient	Weight [g]	GA [wk; d]	PMA [wk; d]
1	1606	30; 3	31; 2
2	2410	33; 6	34; 3
3	1160	27; 3	29; 0
4	1845	31; 5	34; 6
5	1840	30; 3	31; 1
6	1420	29; 2	30; 4
7	1110	27; 0	29; 1
8	755	33; 2	34; 5

4.3.2   Annotations

The data was annotated by two trained observers based on 30s epochs adhering the Prechtl system [197]. The observers used a reference ECG time series and video information for annotation. Their adapted annotation style was tested in another trial and proven to be on par with gold standard full PSG annotations [231]. They annotated the following states: AS, QS, IS, wake, caretaking, and unknown (unable to annotate). The total duration of annotated data was 40 h (4850 30s epochs) with a mean duration per patient of 5.2 ± 1.3 h (630 ± 157 30s epochs). The overall distribution of state was: AS: 62.7%, QS: 8.2%, IS: 13.7%, wake: 2%, caretaking: 11.4% and unknown: 1.9%. The mean interrater reliability was high with Kappa of 0.70. The highest agreement was found in AS, QS, and caretaking with a Kappa of 0.75, 0.74, and 0.71. Caretaking was not perfectly agreed on mainly due to some difficulties in determining the exact beginning and end of the caretaking procedure. While the differences were found mainly in-between states with different start and duration of the transition state resulting in a Kappa of 0.59. Wake was agreed on with a Kappa of 0.69. After clarifying differences in concordance with a third trained annotator, the observers reached consent. The concordant annotations were used for further analysis. As a reminder, the Cohen's Kappa statistic defines a score below 0.41 as poor, over 0.41 as moderate, over 0.61 as substantial, and over 0.81 as almost perfect [232].

As preterm infants are mostly awake during caretaking periods, generating very similar signal structures, the labels caretaking and wake were merged under the label caretaking + wake (CTW).

4.3.3   Data recordings

For each infant, ECG, cECG, and videos were recorded in three to five sessions within one week. The reference ECG was recorded with three standard leads via a Philips Monitor (Intellivue Mx800, Germany) at a sampling frequency of 500 Hz. The

cECG sensors were connected to a data acquisition system (DAQ) system. The cECG data were recorded at 8 kHz, then downsampled to 500 Hz. The videos were recorded with a standard, medium resolution, grayscale camera. The camera was mounted either for facial view or total body view. All signals were time-synchronized to the video recordings.

4.3.4   Capacitive ECG

To obtain the cECG, a special incubator mattress with eight capacitive sensors was used. The electrodes were connected to the DAQ. The top layer of the mattress itself was made out of conductive material to function as a reference to the eight capacitive electrodes (Figure 28 and Figure 18). The mattress was covered with a polyurethane cover to withstand fluids (e.g., urine) and for easy cleaning and disinfection. The polyurethane layer was then covered with a cotton bed sheet for comfort and to reduce the triboelectric effect appearing on both electrodes types [233]. The triboelectric effect creates an electric charge when friction is applied between two different materials, e.g., due to the movement of the neonate, and is not the signal of interest but rather an artifact. The baby should be positioned as to cover both, at least a part of the sensors array and a part of the reference electrode. The position (supine, side, prone) does not influence the heart rate determination but changes the shape of the ECG signals. To provide further comfort, the neonate was placed on the mattress in a cotton snuggle.

A vectorcardiogram and its projection on the Einthoven leads [234], were constructed from the raw data after a channel selection process. Before the channel selection, neutralization was used on the analog signal eliminating the input impedance of the sensor amplifier to reduce motion-induced noise [235]. The channel selection process first ranked the channels by its coupling factor between sensors and body. The coupling factor was determined by injecting a low-current1 kHz, whose signal amplitude decay is proportionate to the coupling into the electrodes [236]. Secondly, bad channels, i.e., those with a variance higher than a certain threshold,
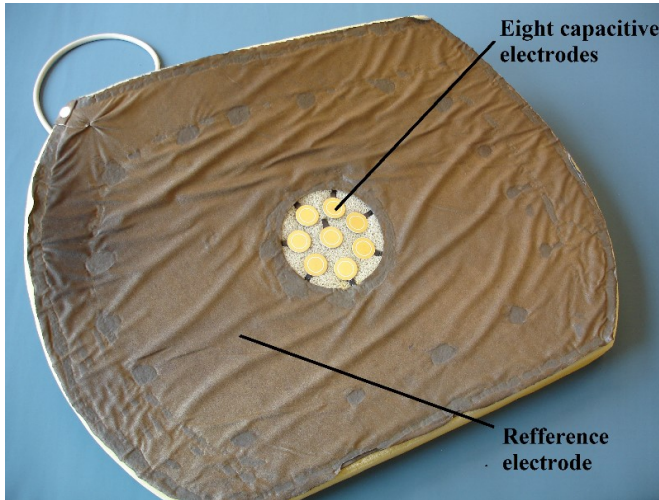
Figure 28 Incubator mattress with eight capacitive sensors and a conductive cover material acting as the reference electrode. The mattress would be protected with a polyurethane cover to withstand fluids (e.g. urine) and for easy cleaning and disinfection.

were eliminated. For details about the channel selection, we refer to Atallah et al. [107]. The ECG was then reconstructed from at least three selected channels [237]. The signals from the selected channels were first down-sampled, and a bandpass filter of 3 to 35 Hz was applied, thus removing all frequency components outside the (dominant) ECG band, e.g., the well-known 50/60 Hz common-mode interference was rejected. To subtract the common mode, which is significantly affected by people walking by, the signals were averaged, and the average signal subtracted from each channel signal.

4.3.5 R peak detection

To analyze the heart rate variability (HRV) from the cECG, R-peak detection, and normal-to-normal beat (NN) interval determination is very important. Slight variations in peak detection would introduce false sleep staging as the difference between the sleep states is only minimal. In this paper, we compare two R peak detection methods to determine which yield better results for the presented data. One

method is from Wijshoff et al. [200] and a second from Rooijakkers et al. [238]. The algorithm from Wijshoff et al. was not initially intended for preterm infants but was confirmed to be working well in this patient group [229]. Rooijakkers' method was created and confirmed for fetal monitoring. Wijshoff et al. calculated the first derivative of the ECG signal to search for the steepest ascent and descent of the QR and RS slopes. A variable threshold was applied to detect the peaks in the QRS complex. They then used a sub-peak detection to verify the peak position at the real max by interpolating around the found peaks. The sub-peak detection assured that there is no shift from the real peak due to off sampling. Rooijakkers et al. band passed the ECG signal locally with the use of time-discrete continuous wavelet transform with the peak wavelet frequency centered in the 10-25 Hz frequency band. They then segmented the ECG signal to obtain one QRS complex per segment. Those segments are run through a variable threshold to find the R peaks within a set time window, which is based on the previously found R peak.

Cross-correlation was determined between the HRV signal created from the ground truth ECG and the cECG signal for each R peak detection method. The features for sleep state classification were created for both methods, and the classification performance was compared using the two different methods.

4.3.6   Features

To be able to separate the different states, in total, 34 ECG and HRV features in the time, frequency, and non-linear domain were determined. The features were calculated on the base of 300s windows centered on 30s epochs. The features are calculated for each 30s epoch and averaged over the corresponding 300s window. An overview of all used features can be found in Table 12.

For the HRV, the needed ECG R-peaks are fundamentally non-equidistant in time. To avoid resampling the RR signal, and thereby introducing extra parameters, the Lomb-Scargle algorithm was applied to generate the frequency spectrum [203]. As a

base, a set of existing HRV features are reused, which were introduced in Chapter 3. These are linear HRV features focusing mainly on direct representation of changes in the para-/sympathetic nervous system expressed in cardiorespiratory changes. To refresh, for the two preterm infant features pHF1, and pHF2, introduced in Chapter 3, the frequency band was extended based on adult sleep staging to 0.45-0.7 Hz for pHF1 and 0.7-1.5 Hz for pHF2, and then the spectrum power in these two bands was computed.

To investigate the movement induced noise (e.g., body movement artifacts) of the capacitive electrodes, features calculated directly from the ECG signals (ECG and cECG) were introduced. Movement artifacts can be a direct and indirect indication for certain sleep states. Beats per epoch (BpE) counts the R peaks in an interval of 300s. Line length (LL) and mean LL (aLL) calculates and averages the length of the ECG time series signal over a window of 300s. Also, the standard derivation is calculated over LL (SDLL) and aLL (SDaLL) in 300s windows. The LL is calculated with linear piecewise approximation using numerical integration for each segment of 30s to calculate the arc length, which is then summed up over 300s epochs to gain the cumulative chordal distance.

Kommers et al. [208] designed two new features specifically for preterm infants to capture regulatory changes during kangaroo care: the percentage of HR decelerations (pDec) and the magnitude of HR deceleration (SDDec). The two features indicate how many HR decelerations occur and the extent of those decelerations. They showed that pDEC and SDDec are strongly affected by kangaroo care, supporting that HR decelerations are a consequence of the autonomic nervous system response. Thereby, these features are expected to be influenced by sleep state changes.

Lucchini et al. [220] suggested that non-linear sample entropy (SE) and quadratic sample entropy (QSE) describe the preterm infant autonomic response. As the autonomic response is directly linked to the sleep states, those non-linear features were incorporated. In the literature, it is described that the standard value for the

tolerance parameter r of 20% times the standard derivation is not accurate enough anymore, especially in preterm infants where instabilities are the norm [239], [240]. Therefore, the SE was calculated on an adaptive tolerance parameter *r*. The embedded dimension *m* was fixed to 2 following [220], [239]. The r-value per epoch was determined by calculating the SE over a range of r from 0.05 to 0.3 times the standard derivation (SER) for 300s epochs, which also includes the standard value for r. The calculation to find the optimal r-value was done as described in the following points:

- Calculate SE over a range of r values (SER).

- Generalize the SER curve by low pass filtering (e.g., moving average).

- Find drop off/turning point of the SER curve where the amount of matches increases and entropy decreases.

- Find the minimum value of the SER curve.

- Calculate the mean SE value between the SER curve turning point and the minimum value.

- Find the r with the closest entropy value to the calculated mean r-value in the original SER curve.

In addition to SE and QSE, another fluctuation measure was calculated: the Lempel-Ziv complexity measure (LZ) for the ECG (or cECG) and corresponding HRV signal. With LZ, the different signal structure is measured using the ECG (or cECG) and HRV time series as analytic signals. The LZ is more prone to signal length than SE and QSE, but when used on a fixed time window, the effect will not have any influence on the calculated LZ values. Another novel feature is the sample entropy area under the curve (SEAUC). Here the previous calculated SER curve is extended to a range of r of 0.1 to 20, creating a longer tail. Then the curve is integrated to show the difference between the curve shapes for different sleep states. With a higher entropy in the signal, matches are found later with a higher tolerance r, shifting the drop off/turning point and thereby the AUC value.

Table 12 Overview of the used ECG and HRV features for classification.

NR	Feature [unit]	Description
0	BpE [count]	Beats per Epoch
1,2	LL, aLL [mV]	Line Length / mean Line Length
3-6	NNx [count]	The number of pairs of successive R-R intervals that differ by more than 10, 20, 30 or 50 ms of a defined window length.
7-10	pNNx [%]	The proportion of NNx divided by the total number of R-R intervals of a defined window length.
11	RMSSD [ms]	Root mean square of successive differences between adjacent R-R intervals of a defined window length.
12	SDALL [mV]	Standard derivation of averaged line length
13	SDANN [ms]	Standard Deviation of averaged NN intervals
14	SDLL [ms]	Standard derivation of line length
15	SDNN [ms]	The standard deviation of normal to normal R-R intervals of a defined window length.
16	HF [ms^2]	The power of the high-frequency band between 0.15-0.4 Hz of a defined window size.
17	HFnorm [%]	HF power in normalized units HF/(Total Power-VLF) x 100
18	LF [ms2]	The power of the low-frequency band between 0.04-0.15 Hz of a defined window size.
19	LFnorm [%]	LF power in normalized units LF/(Total Power-VLF) x 100
20	LF/HF [n.u.]	Ratio LF/HF
21	pHF1 [ms^2]	The power of the high-frequency band between 0.4-0.7 Hz

22	pHF1norm [%]	pHF1 power in normalized units pHF1/(Total Power-VLF) x 100
23	TotPow [ms^2]	Total power or variance of NN intervals of a defined window size.
24	pHF2 [ms^2]	The power of the high-frequency band between 0.7-1.5 Hz
25	pHF2norm [%]	pHF2 power in normalized units pHF2/(Total Power-VLF) x 100
26	VLF [ms^2]	The power of the very-low-frequency band between 0.003-0.04 Hz of a defined window size.
27,28	SEN, QSE [n.u.]	Sample entropy / Quadratic sample entropy
29	SEAUC [n.u.]	Sample entropy area under the curve
30	pDEC [%]	The percentage of HR decelerations
31	SDDec [ms]	Magnitude of HR deceleration
32,33	LZNN [n.u.], LZECG [n.u.]	Lempel-Ziv complexity measure on HRV and ECG

4.3.7   Preprocessing

Before feeding the features to the learning routine, preprocessing steps were performed. The data were first normalized per recorded session using the Python scikit-learn standard-scaler and Min-Max-scaler for comparison [212]. Selected features were averaged per session with a moving average using different window lengths. As the data was highly unbalanced, the Synthetic Minority Over-sampling Technique (SMOTE) [241] and Adaptive Synthetic Sampling Approach (ADASYN) [242] were applied to increase the number of data points for more stable and generalized learning. To be able to distinguish non-linear and inter-feature correlations and to increase linear and non-linear separability of the sleep states, feature transformation was applied to elevate the feature space to a higher dimension.

A second-order polynomial feature transformation and radial basis function kernel (RBF) were used with gamma grid search between 0.001 and 10 to transform the selected features. The polynomial function for feature x results in a new feature set: x, x^2, $x*y$, y, y^2, …, $x*n$, n, n^2. The polynomial feature transformation is used to generate a new feature space including feature interactions. For the selected features, the quadratic and mixed features were kept. The additional created features n and n^2 were removed as only the transformation of the selected feature x is of interest. The not intended feature transformation n and n^2 in the new feature space of x could lead to overfitting and redundant information, which can result in decreasing classification performance. If n or n^2 is generating valuable information, n should be selected separately for transformation. Then also $n*x$ should be removed as it is already included in the transformed feature space of x.

Further, the input parameters were averaged with a moving window. The size of the moving window was chosen between 0 and 50 30s epochs to incorporate short and long-term averaging factors.

The data was separated into training, validation, and test sets, split by patients to be used later in a leave one patient out cross-validation (LOOCV). The validation set is used during the training phase to update the parameters. The test set is kept aside to perform an unbiased test on never seen data after the training is finalized. As the data set is small, a classic data splitting based on a fraction into training, validation and test set could not be performed as it would have introduced bias and lead to overfitting on the classifier side. To avoid bias and overfitting, it was chosen to split the data between patients to assure that the validation and especially testing set represents unseen information. In the next step, a classifier is empirically chosen for each class-set. The classifier is fed with its parameters and preprocessed data. Following, the performance is validated with LOOCV, and the classifier parameters and preprocessing are adapted to optimize the classification performance.

4.3.8   Selection path strategy

As only a small group of patients was available with an unbalanced state distribution, a selection path model for each state was implemented. The goal is to increase the multiclass classification performance by separating the classification in smaller sub-classification problems, which can be optimized with the use of a wrapper, including data preprocessing, different classifiers, feature selection, and parameter tuning. Also, by creating those intermediate classifications, it can be determined which factors influence the classification of a particular state. The final target is a full class classification. Using intermediate steps until full class classification, the classification was separated into smaller groups of classes to identify the classification performance for individual state groups: AS-QS, AS-QS-IS, AS-QS-CTW. The classification of each state-group focuses on a sub-state-group composed of one minority state (QS, IS, CTW) and the majority state AS, e.g., AS-QS out of AS-QS-IS. Within the wrapper, the chosen classifier, feature selection, and parameters were individually optimized for each of the sub-state-groups. Each sub-state-group classification results in a prediction. Later, the wrapper merges the sub-state-predictions under a ruleset to a final prediction per state-group as wrapper output (Figure 29). The ruleset is such constructed that if AS, as the majority state, is below a minimum probability threshold, the state with the highest probability among the minority states is chosen. This ruleset is performed for each class-set probability output, which is optimized for a specific minority state. The minimum probability threshold is determined via a grid search. The class-set optimization uses the F1 score on the validation sets. The implemented probability cut off for AS is used to limit the influence of the majority class AS. Finally, the class-set optimized predictions are merged. Upper and lower probability ruling thresholds decide which sleep state is chosen for the final class prediction per epoch. The whole process is assessed by evaluating the final sleep state identifications from the test set against the annotations using the Kappa score ($\varkappa$) [232].
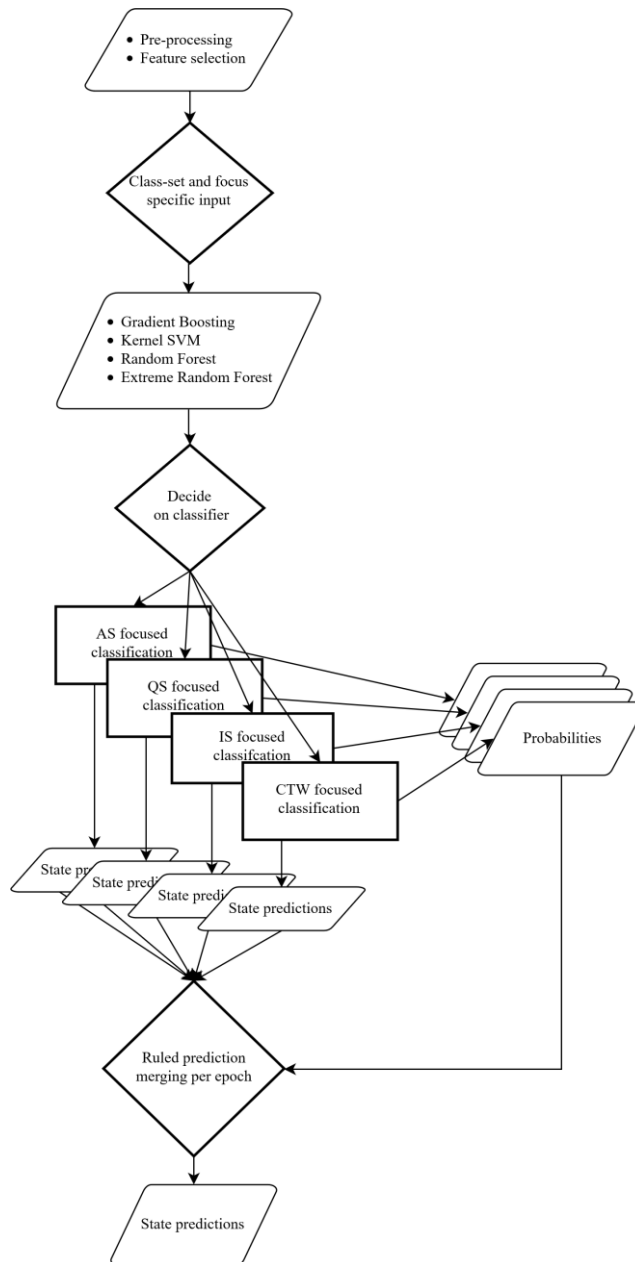
Figure 29 Overview of the multi-classifier approach. In the first step, parameters for each class are selected separately which are then fed to a chosen classifier. The classifier predictions for each state are merged into a single prediction using the state probability for each epoch. The merging uses a ruleset using the probabilities per class-set to decide on the final state prediction. The outcome are joint state/label predictions.

4.3.9   Classification

Five different classifiers from the Python scikit-learn library [212] are used for the wrapper to choose, which are decision tree (DT), RBF kernel support vector machine (SVM), random forest (RF), extra tree random forest (ERF) and gradient boosting (GB) [243], [244]. The classifiers' input parameters are specific for different kinds of classifiers. The adapted input parameters for the tree learners DT, RF, ERF were the number of trees, the measure of branch splitting quality, the depth of the tree, the minimum amount of samples representing a leaf node, and the minimal number of samples allowing for a split. Also, the GB tree can be run with logistic regression or AdaBoosting loss function. Another parameter is the learning rate (also shrinkage or eta), which corrects for prediction errors from the existing trees by weighting each tree contribution. The SVM needs the misclassification penalty parameter C and the hyperparameter gamma, which is determining the influence of the support vector on the class decision. Both parameters were found with a pre-grid search, including all features, to find the overall optimal values. As a starting point, the values 3 for C and 3.8 for gamma were used as good balance between speed and accuracy without losing generalization properties.

4.3.10   Feature selection and parameter determination

To choose the right feature set and classifier parameters for each separate sleep state classification, an adapted greedy-bi-directional (backward and forward) search was implemented. In the first step, all features are fed into the system without any dimension expansion (by feature transformation). After classification on the validation set, the feature importance is determined in the first validation run, and the best features are boosted with dimension expansion and/or the worst features can be removed. The classification performance is determined, and the feature transformation and/or the feature removal extended or reversed. The process is repeated until the performance on the validation set does not increase any further in
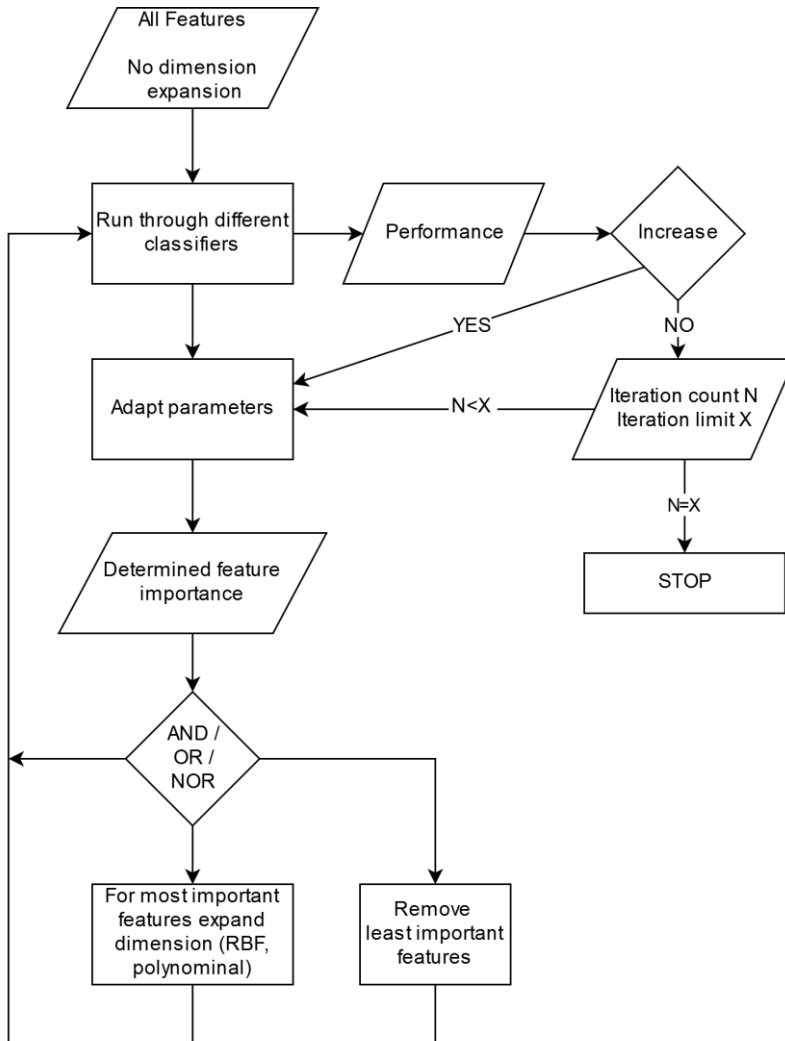
Figure 30 The flow diagram shows how the Features are selected. At start all features are used without dimension expansion like n order polynomial or the Radian Bases function. After classification with a chosen classifier, the most important featured are transformed to increase their class distinction. The least important features can be dropped. A best first greedy approach was used to adapt the parameters with the chosen features. The parameters can be found in Table 14. This is continued until the performance does not increase anymore.

the third decimal place. As the combinations of different parameters, classifiers, and chosen/ boosted features are extraordinary, this selection was not executed via an automated exhaustive grid search but partially manually with experimental alteration of feature selection and best first greedy search for parameter optimization. An overview of the adapted parameters can be found in Table 14. A simplified overview of the whole process can be seen in Figure 30.

4.4 Results

4.4.1 R peak coverage comparison for ECG and cECG

The R-peak intervals for the normalized ECG and cECG signals were determined per session, and the cross-correlation was calculated to see how well both signals align with their reference counterpart. The overall mean correlation per patient using the R-peak detector by Wijshoff et al. is 0.63 ± 0.04. The method of Rooijakkers et al. shows a slightly increased correlation of 0.692 ± 0.22 between the cECG and ECG R peaks. The discrepancy between the R-peaks detected from cECG, and the ECG mainly comes from the different impact of noise on the signal. Noise removal treatment did not enable reliable R peak detection and was therefore left out to avoid the creation of false signal episodes regarding R peak detection. Episodes, where the preterm infant was lying still and covering the electrodes, generate good ECG signals for peak detection (Figure 31).

For other episodes, the signals were corrupted with noise similar to the ECG in composition but without actual R-peaks (Figure 32), creating partly false R-peak detections. Nevertheless, using one or the other R-peak methods does not show any significant impact on sleep staging performance. To create consistency with Chapter 3, the R-peak method of Wijshoff et al. was used in this work.
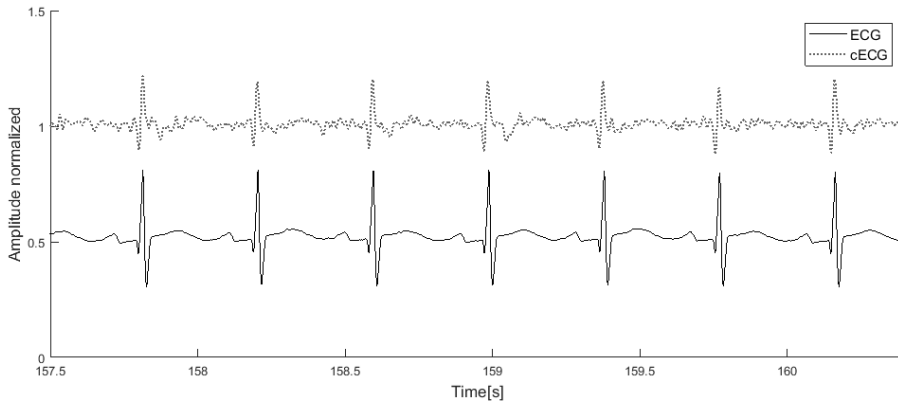
Figure 31 ECG and capacitive ECG (cECG) simultaneously recorded. The ECG signal is more distinct but the cECG shows good potential for R-peak detection. For a more discriminable display of signals, the cECG is plotted with an amplitude offset.
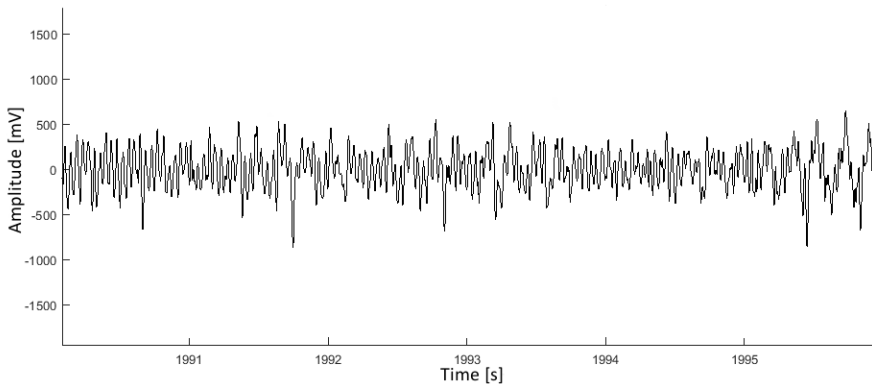
Figure 32 This image shows a cECG signal which can easily be mistaken by the R-peak detector for a signal with QRS complexes modulated with noise. The R- peak detection is not working properly on such epochs resulting in decreased correlation between ECG and cECG HRV signals.

4.4.2   Feature importance

The 34 features have different importance regarding the different state classification and for the used signal type of ECG or cECG. Limiting the input on the most important features can reduce overfitting, by reducing decisions based on noise, and increase accuracy by removing misleading data. The feature importance is determined

using the Random-Forest feature importance function. In Figure 33 and Figure 34, the feature importance for each sub-classification used for classification, including all states, are displayed. The features are displayed before feature transformation to recognize which main feature types are of key importance. To identify which feature types are important for ECG and which for cECG, the overall important features per signal are listed below. Overall important features are features exceeding for at least two subsets the threshold of 3.4% from the total distributed importance, which sums up to a total of 1. The threshold was found by a sweep analysis resulting in the highest performance based on the ECG features. These features are listed in Table 13 per used state-groups and signal type. Here, only the last state-group AS-QS-CTW-IS corresponds to the displayed Figure 33 and Figure 34.
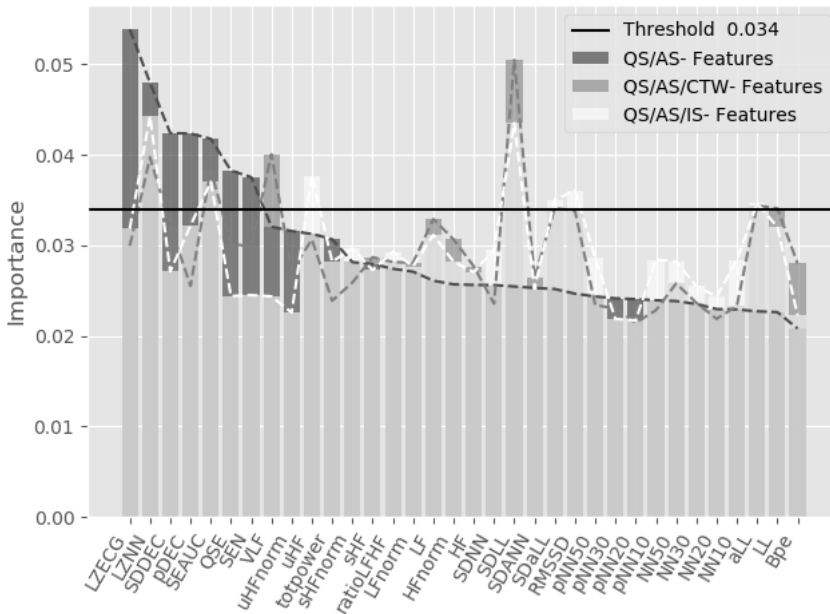
Figure 33 Feature importance for each class-set sorted after AS-QS most important features using the cECG signal. This figure represents the feature importance for classifying between AS, QS, CT/W and IS (state-group). Each of these feature sets was used to classify specifically focused on one class-set to later merge the predictions. The threshold determines which features are deemed as important. For the overall importance per state-group, features that exceeds the threshold for at least two sets are chosen.
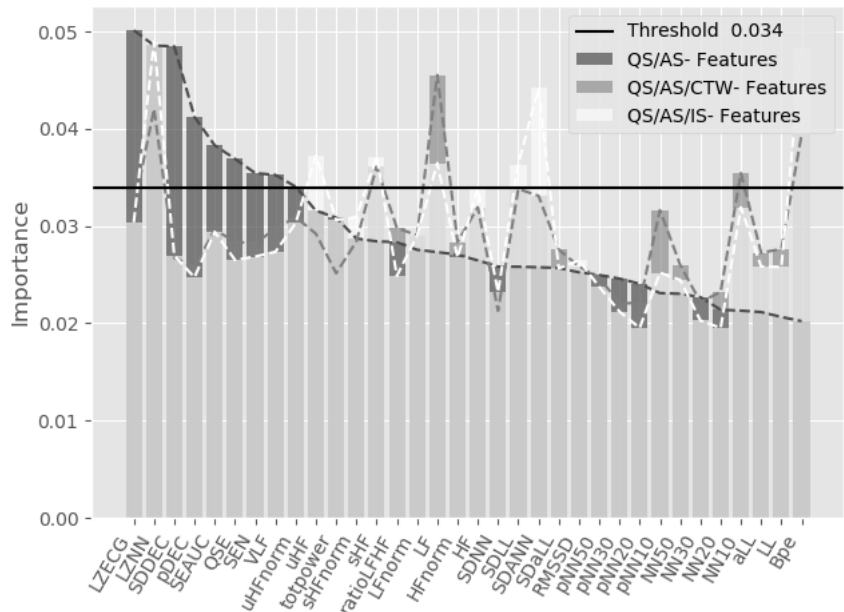
Figure 34 Feature importance for each class-set sorted after AS-QS most important features using the classic ECG signals. Framework same as in Figure 33.

Table 13 Overview of most relevant features for classification per state-group.

State-group	Signal	Features
AS-QS	ECG	LF, SDANN, LZNN
	cECG	LL, aLL, SDANN, LZECG, SDLL, SEAUC, LZNN
AS-QS-CTW	ECG	BpE, LZNN, SDANN, HF, pHF1, LF, SDLL, SDNN
	cECG	SDLL, SDaLL, LL, ZECG, aLL
AS-QS-IS	ECG	SDANN, BpE, LF, totpower, pHF2, pHF2norm
	cECG	SDLL, LZECG, LL, SDaLL, aLL, LZNN, SEAUC
AS-QS-CTW-IS	ECG	LZNN, BpE, LF, pHF1
	cECG	SDLL, LZNN, SDaLL, aLL, SEAUC

4.4.3 Parameter selection

The used parameter values varied for each task. Depending on the state combination, bi- or multi-state classification, and the used dataset, the performance changed based on the used parameter combination. The used parameter ranges can be found in Table 14. Feature groups were left out in the averaging step or polynomial transformation. Also, not all features were always used (See Table 15).

The parameters are presented in ranges as presenting all combinations in detail would exceed the purpose of this Chapter and would not benefit the reader as the specific parameter values are tied to the here used datasets.

Table 14 Parameter overview.

Parameter ranges used in different combinations per task. The parameters were adapted in the process shown in Figure 30.

Parameter	Range
Length of the moving window for averaging	None – 60 epochs on selected features
Estimators/Trees	80-500
Min. sample leave	2-5
Split criterion	Gini or Entropy
Probability Threshold for the majority class	0.65-0.9
Polynomial transformation	See Table 15
Used features	See Table 15

4.4.4 Capacitive vs. classic three lead ECG

To identify the sleep states in a genuinely unobtrusive manner, a cECG system was used to capture the ECG and HRV. Looking at AS and QS only, the analysis resulted in a κ of 0.42 ± 0.26; 0.49 (mean $\kappa \pm$ std; cumulative pooled κ) for the ECG and a κ of 0.56 ± 0.20; 0.51 for the cECG. The performance changed when more sleep states

were added to the classification task. A performance overview of the following $\varkappa$ values can be found in Figure 35 and Table 15.

Next, the CTW states are added to the classification task. Due to the low amount of wake but the similar physiological state (being awake, increased heart rate, increased number of noise), both states are merged. Adding CTW to be differentiated from AS and QS, the performance for both signal types, ECG and cECG, reduces. The classification using the ECG signal leads to a performance of $\varkappa$ 0.30 ± 0.12; 0.32. Using the cECG signal, the performance resulted in $\varkappa$ of 0.44 ± 0.21; 0.41. Trying to differentiate between AS, QS, and IS based on the ECG signal, a performance of $\varkappa$ 0.23 ± 0.15; 0.29 was achieved. Using the cECG signal here, the performance reached $\varkappa$ 0.36 ± 0.12; 0.34.
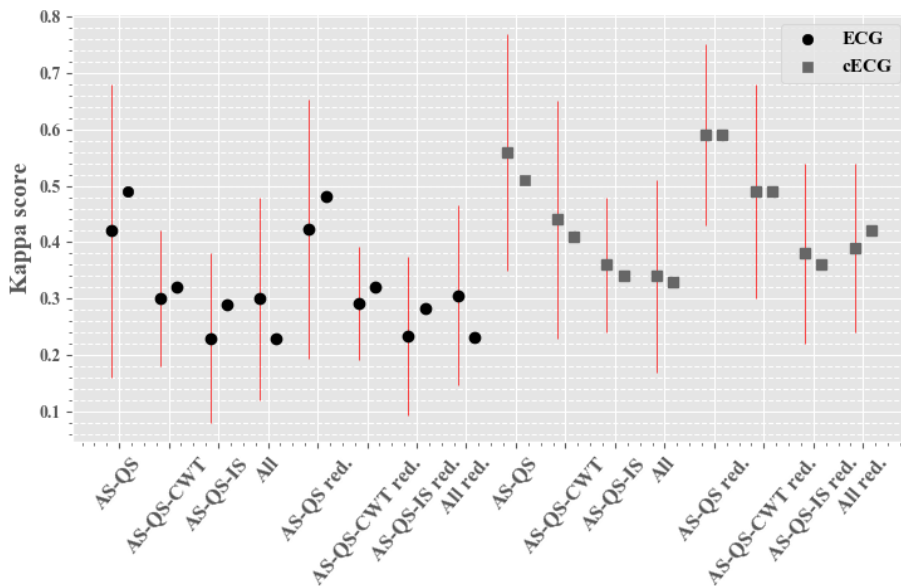
Figure 35 Classification performance for the ECG and the cECG for each class-set with mean κ ± std and cumulative pooled κ. Also displayed is the performance using the reduced set, were two outlying patients were excluded. The results from the reduced sets are labelled with the extension red.

Table 15

Mean kappa performance and used Features per dataset and per class-set. Feature index used as in *Table 12*. Features in bold were transformed.

Task	Signal	Selected features	Performance [κ]
AS |QS	ECG	0,..,5,7,11,..,16,18, 19,**21**,..,25,30,32,33	0.42 ± 0.26
	cECG	0,..,5,7,11,12,14,17, 19,20,22,24,..,29,32	0.56 ± 0.20
AS | QS | IS	ECG	0,..,**21**,22,**23**,**24**,..,33	0.23 ± 0.15;
	cECG	**1**,2,3,6,7,10,..,19, 21,..,29,**32**,33	0.36 ± 0.12
AS | QS | CTW	ECG	**0**,..,10,12,**13**..,18,21, 23,..,26,30,..31,**32**,33	0.30 ± 0.12
	cECG	0,**1**,**2**,5,..,9,11..,18, 20,..,29,32,33	0.44 ± 0.21
AS | QS | IS | CTW	ECG	0,4,..,29,32	0.30 ± 0.18
	cECG	1,2,**4**,**5**,6,7,10,11,12,14,**15**,.. ,**18**,..,**27**,28,29,**32**	0.34 ± 0.17

When differentiating between all states, using the ECG a performance of $\varkappa$ 0.30 ± 0.18; 0.23 was achieved. Using the cECG, the performance reached $\varkappa$ 0.34 ± 0.17; 0.33. Each classification used different input parameter values found via grid search and manual extraction. As those parameters are specific for the here presented data, it was chosen not to present the single parameter values. The general concept, use of the parameters, and selection method were described before.

To determine the maximum performance, two main outliers were dropped. The two outliers were the patient's lowest in age and weight. Both patients were ventilated with continuous positive airway pressure (CPAP). For the ECG signal, dropping the

outliers did not bring any improvements except that the standard derivation stabilized on a lower value for the Kappa per patient. When dropping two outliers for the cECG, increased performance could be measured. The mean increase was ϰ of +0.04 for the mean Kappa performance and ϰ of +0.07 for the cumulative performance measure. Reaching ϰ 0.59 ± 0.16; 0.59 for AS and QS classification, ϰ of 0.49 ± 0.19; 0.49 for AS, QS, and CTW classification, ϰ 0.38 ± 0.16; 0.36 for AS, QS, IS classification and ϰ of 0.39 ± 0.14; 0.42 for all state classification.

The feature sets used for the different class-set can be seen in Table 15. Those feature sets were used for all patient data and reduced dataset with excluded outliers.

4.5   Discussion

4.5.1   Annotations

The ground truth annotations were also based on ECG signals, but mainly video observations were used for the manual sleep state annotations. Thereby, features were included in the annotations which are not captured by the ECG and subsequent cannot be analyzed, creating a general difference between the ground truth and the ECG/HRV approach. Unfortunately, the videos were not always of high quality regarding patient visibility. In few cases, the movements and breathing could only be seen passively through a moving blanket, and the annotators had to rely solely on the vital signs, knowledge of the sleep cycle, and their experience with the patient. Therefore, some annotations might be slightly distorted compared to the actual state. This circumstance is only noted for completion, as in our appraisal this slight distortion does not majorly affect the classification performance.

Additionally, in the annotations, sections were found with very unlikely sleep state sequences. This led us to believe that the ground truth is, to some extent, not representing the actual situation and possibly leading to a minor decreased performance of the classifier.

4.5.2   Features

By identifying the features that are most important for the classification, the standard time domain and frequency domain features are prevailing when using the ECG data. Low- and high-frequency features are prominent in almost all state-group combinations. Low frequencies are associated with baroreflex activities, and the high frequencies link to the parasympathetic system and respiratory activities. In contrast, the dominant cECG features represent more the general signal structure and account for noise and movement artifacts. We believe that this is due to the cECG being more sensitive to movements and therefore picking up the difference between small jitters and no movement expressive for AS and QS. CTW is mainly represented by movement artifacts and external noise from nurse handling, making this state more susceptive for noise/movement based features.

It is known that respiration is a distinct indicator of sleep states. The respiratory sinus arrhythmia, known in adults, is not pronounced in preterm infants. Therefore, breathing might only be picked up through breathing motion artifacts rather than modulated ECG signals as in adults [245]. High-frequency features, which are linked to breathing activities, are seen as essential features with the ECG while in the cECG, the motion features dominate as they possibly resemble breathing stronger than the HRV frequencies. That breathing is picked up by the cECG system is quite likely as it is highly sensitive to movement artifacts.

From the literature, we assumed that pDEC and SDDEC would have more impact on the classification, specifically for discriminating CWT [208]. Nevertheless, as their impact increases mainly in stressful periods, they are possibly distorted by the noise, which naturally appears during stressful periods (e.g., caretaking). Possibly, they will be more prominent in the ECG when more data is available, and distinct deceleration patterns can emerge.

4.5.3   Classifiers

Trying different classifiers, the RF came out on top each time. Classifier benchmarks appear to be not representative in this case due to the low numbers of analyzed patients. However, we will attempt to discuss the probable causes of RF being superior in this case. In Chapter 3, we used the RBF kernel SVM successfully. As SVM, in general, performs very well on binary class problems, it was here confronted with multiclass classification where RF generally is suited better. SVM is known to be particularly good at dealing with outliers. As the data had two patients who were considered as outliers, SVM would be expected to perform better in this case. Nevertheless, RF can also handle outliers and in addition, tends to generalize better, creating a more stable model due to the randomization of data samples, especially on non-sparse data. Important is also that the RF needs much less parameter tuning than a kernel SVM. With the here chosen approach, many parameters fed into the system, which increased the calculation time. Hence, a more exhaustive parameter search for the SVM was not feasible, which possibly lead to suboptimal parameter settings and ergo performance.

The same arguments hold for the RF versus the GB tree classifier, which needs well-tuned parameters to outperform an RF approach. Choosing the RF decreases the chance of overfitting due to the parameter tuning to the dataset at hand. In contrast to an RF classifier, a GB tree learner aims to reduce the bias introduced by the use of shallow trees. For the RF, the bias is as high as the bias of the single sample trees and cannot be reduced. As our train and test sets are created based on patients, the bias is already significantly reduced, limiting the impact of the GB approach. Another benefit of using RF rather than SVM and GB is the robustness of RF against overfitting [246] by the random selection process. Overfitting can further be reduced by limiting the tree depth using *min sample leaf* and *min sample split,* which are adding more regularization.

The ERF performs similarly to the RF classifier. The waiving of bagging increases the performance of ERF over RF for large datasets. Nevertheless, here, the RF still outperformed the ERF slightly. For larger datasets, we would recommend looking further into ERF also because of the reduced calculation time compared to RF.

4.5.4   Classifications

Coming back to the classification results per state and the underlying rationale, the central sleep states AS and QS are mostly defined by the para/-sympathetic nervous system activities and its response to external triggers. When adding CTW or IS, the differentiation becomes more difficult, especially with such a small amount of training data. It can be seen in the confusion matrices that IS is falsely classified mostly as AS and some QS but least as CTW. Logically, those misclassifications rise from the IS definition as an in-between state. IS carries elements of all states and is not well defined regarding vital sign and observation boundaries. CWT is mainly misclassified as state QS or state IS. Those misclassifications seem to break ranks as it is expected that via the movement elements CWT would mainly be misclassified as AS. This is due to several reasons.

First, CWT has the least training data, which results in decreased pattern recognition. Additionally, caretaking induces hugely varying patterns making it more difficult to train and re-identify on specific patterns. In the videos, it could be identified that between handling episodes during caretaking, the patient would show no signs of movement, possibly leading to a false classification as QS. Last, misclassification is linked to the way the predicted labels were joined under probability threshold rules. AS and QS state predictions are determined to focus on the minority class. Next, IS and CTW are joined under probability ruling, which can lead to those misclassifications. The miss-prediction of states in both steps means that there might be further room for optimization by adjusting the probability thresholds even further while they are already tuned in a fine balance accepting this tradeoff.

As the dataset is limited in the amount of data available for training, we removed two outliers to determine the performance on a somewhat regular patient group, even though regular is a rather inept term regarding preterm infants. We noticed that removing those outliers actually only impacted with the cECG. While investigating this matter, it was confirmed that the video quality was decent, and therefore, the video annotations could not pose the core problem. Most prominent is the fact that the two preterm infants with a poor performance were both 27 weeks GA in contrast to the others with 29 weeks GA and older. One of the two patients also had a very low birth weight of 755 g leading probably to very unstable general conditions. Also, the preterm infants with poor performance, which were partly excluded, had CPAP devices in use. This could indicate that the use of CPAP obscured the SNS control of the breathing. As the effect through the included outliers could only be seen in the cECG and not the classic ECG, sleep state alteration through CPAP can most likely be suspended. It seems more likely that the CPAP masks the actual breathing pattern, substantiating, again, the importance of breathing pattern analysis for sleep staging.

With this small amount of data, the results cannot generally speak for preterm infant sleep classification but can draw a picture of the possibilities using the cECG and shows the impact of different factors on the classification such as ECG-features resembling breathing rhythm. We assume that using sole ECG signals will be on par with cECG when more data is available. Additional data will produce a more stable model for classification. The use of a *min sample leave* value of below 10 shows that this model is slightly tailored to this particular dataset. This does not lead to overfitting, but having little training data single epoch outlier has a more dominant impact on the classification performance.

Furthermore, throughout this analysis, we included and excluded a variety of parameters impacting the classification performance. This manually back- and forward approach is predestined for missing the superior classification paths by the wide range of tuning factors. Therefore, this approach would be better suited for a deep learning strategy, where complex and nonlinear cross dependencies are all taken

into account. Ansari et al. [247] showed the successful classification of QS in preterm infants using deep learning algorithms based on EEG signals. Unfortunately, for this preliminary study with limited data points, a deep learning strategy was outside the possible. In the next chapter, we will look into deep learning methods based on a larger dataset.

In Summary, the bi-state classification is possible, especially for the majority classes AS and QS, which are the most important classes for an automated neural maturation monitoring. All state classification is, at this point, not feasible. More data is needed for a stable all-state classification. All state classification would be necessary for a holistic view of the patient's sleep and improved predictions of sleep cycles.

4.6   Conclusion

It was shown that cECG signals can be used for preterm infant sleep staging, separating AS and QS in an acceptable manner. However, analysis of all states, including IS, caretaking, and wake becomes challenging. Movement alone is a strong indicator/separator for preterm infant sleep states. Incorporating features reflecting movement may help to detect sleep-associated respiratory activities, as there exists a strong connection between breathing patterns and preterm infant sleep states. As this study included only a small number of patients, further investigations for more generalized and improved ECG based sleep staging algorithms should be started.

AS multistate classification was not favorable due to the high dimensionality and complexity of the problem, we suggested looking into possibly better-suited deep learning approaches. In the following chapter, we examine different artificial neural network architectures and approaches for preterm infant sleep staging.

5 Deep learning approach for sleep state classification in preterm infants

5.1 Abstract

Preterm infant neural development is related to the distribution of their sleep states. The distribution changes throughout development. Automated sleep state monitoring can become a powerful aid for development monitoring in preterm infants. Three datasets, including 34 preterm infants and a total of 18018 30s, manually annotated, sleep intervals (sleep-epochs) were analyzed in this study. The annotation of sleep states includes active sleep, quiet sleep, intermediate sleep, wake, and caretaking. Four different recurrent neural network architectures were compared for two-state, three-state, and all-state analysis. A sequential network was used to compare long- and short-term memory and gated recurrent unit models. Other network architectures were based on the popular ResNet and ResNext architectures utilizing residual connection for more depth. The most essential sleep states, active and quiet sleep, could be separated with a kappa of 0.43 ± 0.08. Quiet versus caretaking and wake showed a kappa of 0.44 ± 0.01. The three-state classifications of active versus quiet versus intermediate sleep resulted in a kappa of 0.35 ± 0.07 and active versus quiet versus wake and caretaking resulted in a kappa of 0.33 ± 0.04. The all-state classification was underperforming, probably due to difficulties in separating subtle differences between all states and a lack of sufficient training data for the minority classes.

5.2 Introduction

Preterm infant sleep shows several distinct sleep states. They are defined mainly as AS, QS, and wake. In very preterm infants, sleep states may be rudimentary and show transitionary shifts from one to the other, often with patterns of both AS and QS, defined as intermediate or undetermined (IS) states [183]. AS is often compared to the adult REM sleep states because it shows similar increased neural activity [45], [52], [55], nevertheless, the role of preterm infant sleep states seems to be different. It is assumed that the sleep, the sleep states and the sleep cycles of the fetus, preterm and term infants all play an essential role in the sensory and cortical development [180], [181], [248]. Initially, AS is providing stimulation to the newborn brain in a sensory-reduced environment triggering the development of brain regions with reduced sensory input [55]. Furthermore, during AS, the development, integration, and alignment of specific neural tasks/regions into the cortex structure is taking place. During QS, it is reasonably assumed that developmental errors are corrected, and reorganizations are conducted with the use of increased brain plasticity. During QS, the parasympathetic nerve activity is dominant, blood pressure and heart rate are lowered and, therefore, often seen as the resting and re-energizing state [66], [248].

AS sleep is dominating the sleep cycle of preterm infants with about 80% of the total sleep time at early gestational birth. QS is seen as the minority state, with about 18% of the total sleep time. The distribution changes in the course of development with decreased AS and increased QS [39], [40], [43]. The states can be observed by differences in many electrophysiological signals, such as vital signs (e.g., heart rate and respiratory rate), movement, and EEG activity. During active sleep, increased cardio-respiratory activity and increased motor activity with sporadic eye movements can be observed. During QS, all cardio-respiratory activities are lowered in amplitude and dynamic range [43], [46], [183].

From our experience, at present, manual, sporadic sleep annotation is still the clinical standard for sleep classification and analysis. Continuous and automated

monitoring of sleep in newborns may provide clinical decision support by optimizing the workflow of the caretakers, avoiding disruption of AS/QS and, most importantly, to better safeguard the preterm infant´s developmental process. To provide such a monitoring system in a NICU setting, no additional sensor(s) should be introduced. Therefore, the system should concentrate on already existing, continuously monitored parameters such as ECG. To date, the most successful approaches regarding preterm infant sleep monitoring are using EEG signal analysis [101], [247]. Unfortunately, EEG and even the reduced Amplitude integrated electroencephalography (aEEG) have to introduce additional sensors to the routinely used monitoring solutions, such as ECG, to enable sleep monitoring. As the preterm infant skin is highly sensitive with only three layers thick epidermis and almost no outer protective skin layer [107], [108], [183], additional electrodes should be avoided and EEG/aEEG is not used for a regular and continuous solution at this point. To overcome these limitations motivated us to investigate ECG regarding preterm infant sleep monitoring. Further motivation for the use of ECG as a signal modality is the fact that ECG can also be obtained unobtrusively, with, e.g., capacitive ECG [249], and could thereby be utilized for a future non-contact sleep monitoring solution.

5.3   Related work

Machine learning opens up the possibility of creating an automated algorithm based only on ECG. In the adult sleep research, machine learning algorithms have already been successfully used [194], [195]. Radha et al. [250] compared several machine learning methods such as random forest and ensemble SVM to show promising real-time EEG sleep analysis. Further, full PSG analysis [68], [251], and unobtrusive actigraphy methods [252] were investigated for adult sleep state separation.

Sleep analysis for preterm infants is more difficult as the states are less distinguishable. Nevertheless, machine learning for automated EEG [101] analysis, ECG for sleep vs. wake [189], and HRV for sleep states analysis [229] have been

investigated, demonstrating the potential of machine learning for preterm infant sleep state classification.

A step further in the automated analysis is the use of artificial neural networks (ANN), and respectively, deep learning [253]. The advantage of ANNs over more traditional machine learning is that ANNs can learn richer representations and model complex non-linear relationships. This is of specific importance when dealing with human data as there appear many nonlinear and complex relationships between the in- and output. In addition, ANNs are easy to change in their architecture to adapt to the complexity level of the problem at hand. Also, ANNs are non-parametric models that have the advantage over parametric models that any given input distribution can be learned without prior knowledge. This comes especially in handy when multi-source input features are used with unknown or different distributions. Furthermore, convolution derived features could, in theory, be superior in separating sleep stages than handcrafted features based on ad-hoc decisions such as thresholds, filter cutoffs, or similar approaches. The latest development in ANN originated specific units for time series analysis in recurrent neural networks (RNN) [253], the long short-term memory (LSTM) [254], and gated recurrent unit (GRU) [255]. Those are specifically designed to find patterns in time, such as sleep architecture, to generate improved performance. As the development focuses currently on ANNs, more groundbreaking renovations, also benefitting sleep analysis, might be expected in the future. The application of ANNs to the topic of sleep classification is on the rise. In 2017, Bishwal et al. [256] presented the annotation tool SLEEPNET using a large dataset to train a deep RNN reaching human-level annotation performance. During the same period, Chambon et al. [257] published the implementation of an algorithm that is independent of crafted features using convolution in combination with spatial filtering for classification. A similar approach was chosen by Supratak et al. [258] using an ensemble of convolutional neural networks (CNN) and RNN networks to be able to classify sleep from raw EEG data. At the beginning of 2018, Olesen et al. [259] presented an approach with an adapted model using transfer learning from the

ResNet50 architecture. In the following, Sano et al. [260] used long- and short-term memory (LSTM) classifier to identify wake versus sleep from multimodal data. One of the most recent publications on the topic from Radha et al. [261] used LSTM classifier to classify sleep from HRV features, overcoming the temporal limits of non-temporal models. Also, they used transfer-learning to enable the utilization of other signals (here photoplethysmography) with the same trained model enabling different application areas.

Sleep classification in adults is well handled with deep learning, less studied in preterm infants. So far, Ansari et al. [247] are the first to implement a CNN network for preterm infant EEG signals successfully. This chapter tries to investigate the possibility of using an RNN approach for the more difficult preterm infant sleep classification. We hypothesize that with deep learning algorithms which incorporate time-domain analysis, such as with RNN architectures, preterm infant sleep states can be classified in an acceptable accuracy only based on ECG features. We predicate this hypothesis on the base that ANN networks are distinguished on analyzing highly complex systems which are influenced and dependent on multi-system factor interrelations. And preterm infant sleep is such a multi-factor influenced system. Generally, ANNs should outperform classic machine learning methods; however, based on limited data in this study, we believe that our ANN approach will perform equally to previous machine learning methods but having overall greater potential.

5.4   Methods

5.4.1   Population

Datasets from three retrospective studies were combined. These datasets comprise the data used in chapters 3 and 4 (with one new patient) and an additional unseen dataset of 17 patients. The dataset recordings have a timespan of several years in-between them. The infants were admitted to the NICU of the neonatal department at the Máxima Medical Center Veldhoven, The Netherlands. Ethical approval was

given by the medical ethical committee of the hospital, and written consent was given by the patients' parents. In those three retrospective studies, 34 (8, 9, 17) stable preterm infants were analyzed during 39 sessions. The preterm infants were born with a mean gestational age (GA) of 29 ± 2.1 weeks. They were studied at a mean postmenstrual age (PMA) of 33 ± 2.0 weeks. The patients had a mean birth weight of 1338 ± 473 g. More details can be found in Table 16.

Table 16 Patients weight, gestational age [GA], and postmenstrual age [PMA].

Patient	Weight [g]	GA [wk, d]	PMA [wk, d]
1	1845	31; 5	33; 6
2	2265	33; 6	34; 6
3	1740	28; 6	34; 6
4	1235	30; 1	31; 2
5	1700	32; 6	33; 5
6	1290	28; 6	29; 3
7	1460	25; 4	31; 5
8	1615	29; 6	32; 0
9	1606	30; 3	31; 2
10	2410	33; 6	34; 3
11	1160	27; 3	29; 0
12	1845	31; 5	34; 6
13	1420	29; 2	30; 4
14	1110	27; 0	29; 1
15	755	33; 2	34; 5
16	2080	27; 3	35; 6
17	2180	29; 6	31; 3
18	1180	29; 6	34; 3
19	1290	29; 1	32; 0
20	1020	29; 1	35; 1
21	610	29; 2	34; 0
22	1210	27; 5	32; 6
23	1250	26; 4	32; 5
24	1070	26; 6	33; 7
25	1880	30; 3	33; 1
26	1130	28; 6	33; 0
27	1050	28; 4	35; 1

28	1170	27; 3	34; 0
29	890	31; 4	33; 5
30	920	30; 3	33; 2
31	1050	28; 4	31; 0
32	600	27; 6	33; 1
33	790	29; 6	35; 0
34	680	26; 2	28; 1

5.4.2 Data recordings

Vital signs recordings for all studies were performed with a Philips patient monitor (Intellivue MX 800, Germany) at a sampling frequency of 500 Hz (n=32) or 250 Hz (n=2). The 250 Hz data were cubic interpolated to meet the 500 Hz. The vital sign signals consist of EEG, ECG, cECG, PPG, and EOG recordings. For all datasets combined, only ECG was consistently recorded. For this chapter, only the ECG signals were used. The signals were recorded in all studies with three-lead ECG using standard adhesive ECG patches.

Each preterm infant was also video-recorded. Videos were recorded either of the face or the total body view. The used cameras were standard, consumer-grade, greyscale devices.

5.4.3 Annotations

Per dataset, two trained observers annotated the data based on 30s intervals (sleep-epochs) adhering the Prechtl system [197]. The observers used a reference ECG and respiration time series and video information for annotation. Vocalization, which is part of the Prechtl system, could not be used due to the lack of audio recordings. They annotated the following states: AS, QS, IS, wake, caretaking, and unknown (unable to annotate). The more specific states from Prechtl active wake and quiet wake were merged into wake due to lack of data. The total duration of annotated data was 167 h (20021 30s intervals) with a mean duration per patient of 4.28 ± 1.5 h

(513 ± 179 30s intervals). The overall distribution of state was: AS: 51.45%, QS: 12.7%, IS: 16.5%, wake: 6.6%, caretaking: 2.2% and unknown: 10.5%. Subtracting the unknown epochs, a total amount of around 18018 30s intervals were left for analysis. The detailed distribution of all trials can be found in Table 17. The median inter-rater variability lay at 0.8 ± 0.1 for AS and 0.6 ± 0.1 for QS. The overall median inter-rater variability was 0.7 ± 0.1.

Table 17 State distribution per dataset in percent.

States	Dataset 1 [%]	Dataset 2 [%]	Dataset 3 [%]
Unknown	1,6	37,9	0,0
AS	65,4	47,7	46,0
QS	7,5	6,7	18,9
Wake	2,0	1,6	11,9
CT	8,4	0,0	0,0
IS	15,1	6,0	23,3

As preterm infants are mostly awake during caretaking periods, generating very similar signal structures, the labels caretaking and wake were merged under the label caretaking + wake (CTW) to equalize for the low amount of data from each state.

5.4.4 ECG R-peak detection

The R-peak detection algorithm of Wijshoff et al. [200] was used to determine the NN intervals and the resulting HRV signal. To determine the steepest ascent and descent of the QR and RS slopes, they calculated the first derivative of the ECG signal. Then the peaks in the QRS complex were detected with a variable threshold. By interpolation around the detected peaks, they verified that the position of the peak is at the real max. This sub-peak detection assured that there is no shift from the real peak due to off sampling.

5.4.5   Features

For each dataset, 47 features from HRV, ECG, and patient information were created. The features were calculated based on 30s intervals. The HRV features include the time, frequency, and nonlinear domain. The ECG features were calculated in the time and nonlinear domain, while the ECG derived respiration (EDR) features were calculated in the frequency and nonlinear domain. For HRV and EDR the signals are fundamentally non-equidistant in time. The Lomb-Scargle algorithm [203] was used to generate the frequency spectrum as resampling for classic Fourier transformation would have introduced extra parameters.

Table 18 Overview of the used ECG and HRV features for classification.

NR	Feature [unit]	Description
0	BpE	Beats per Epoch / mean Beats per Epoch
1,2	LL, aLL [mV]	Line Length / mean Line Length
3-6	NNx [count]	The number of pairs of successive R-R intervals that differ by more than 10, 20, 30 or 50 ms of a defined window length.
7-10	pNNx [%]	The proportion of NNx divided by the total number of R-R intervals of a defined window length.
11	RMSSD [ms]	Root mean square of successive differences between adjacent R-R intervals of a defined window length.
12	SDALL [mV]	Standard derivation of averaged line length
13	SDANN [ms]	Standard Deviation of averaged NN intervals
14	SDLL [ms]	Standard derivation of line length
15	SDNN [ms]	The standard deviation of normal to normal R-R intervals of a defined window length.
16	HF [ms^2]	The power of the high-frequency band between 0.15-0.4 Hz of defined window size.
17	HFnorm [%]	HF power in normalized units HF/(Total Power-VLF) x 100
18	LF [ms2]	The power of the low-frequency band between 0.04-0.15 Hz of defined window size.
19	LFnorm [%]	LF power in normalized units LF/(Total Power-VLF) x 100
20	LF/HF [n.u.]	Ratio LF/HF

21	pHF1 [ms²]	The power of the high-frequency band between 0.4-0.7 Hz
22	pHF1norm [%]	pHF1 power in normalized units pHF1/(Total Power-VLF) x 100
23	TotPow [ms²]	Total power or variance of NN intervals of defined window size.
24	pHF2 [ms²]	The power of the high-frequency band between 0.7-1.5 Hz
25	pHF2norm [%]	pHF2 power in normalized units pHF2/(Total Power-VLF) x 100
26	VLF [ms²]	The power of the very-low-frequency band between 0.003-0.04 Hz of defined window size.
27,28	SE, QSE [n.u.]	Sample entropy / Quadratic sample entropy
29	SEAUC [n.u.]	Sample entropy area under the curve
30	pDEC [%]	The percentage of HR decelerations
31	SDDec [ms]	Magnitude of HR deceleration
32,33	LZNN [n.u.], LZECG [n.u.]	Lempel-Ziv complexity measure on HRV and ECG
34	HF_R	The power of the high-frequency band of the respiration signal between 0.48-1.1 Hz of defined window size.
35	HFnorm_R	HF respiration power in normalized units. HF/(TotPow_R-LF_R) x 100 of the respiration.
36	MF_R	The power of the medium frequency band of the respiration signal between 0.56-0.84 Hz of defined window size.
37	MFnorm_R	MF power in normalized units of the respiration. MLF_R/(TotPow_R-LF_R) x 100
38	LF_R	The power of the low-frequency band of the respiration signal between 0.56-0.3 Hz of defined window size.
39	LFnorm_R	LF power in normalized units of the respiration. LF_R/(TotPow_R) x 100.
40	LF_R/HF_R	The ratio between the low and high respiration spectrum. LF_R/HF_R
41	MF_R/HF_R	The ratio between medium and high respiration spectrum. MF_R/HF_R
42	TotPow_R	The total power of the respiration frequency spectrum.
43	Age difference	Difference between age at birth and age at measurement.
44	Birthweight	Weight at the time of birth

| 45 | GA | Gestational age. Age at birth calculated from the last gestation. |
| 46 | PCA | Conceptional age. Age at time of measurement |

As the RSA and cardiorespiratory coupling is not very pronounced in preterm infant and can only be seen in more mature infants [97], [207], Joshi et al. [262] confirmed very recently that coupling between heart rate decelerations and accelerations exist in preterm infants but not vice versa. They assume that RSA in preterm infants is not present, mainly due to insufficient breathing depth. The respiration can consequently not be determined from the RSA but rather via superimposed chest movement on the ECG signal. Therefore, the EDR signal was calculated using the ECG envelope. In the frequency domain, the frequency band was limited to max 1.1 Hz (66 breaths per minute) and min 0.3 Hz (18 breaths per minute), which is described in the literature as the min and max respiration rates of preterm infants [263].
The frequency bands were then separated into high (1.1 - 0.84 Hz), medium (0.84 - 0.56 Hz), and low (0.56 – 0.3 Hz) bands.

From the patient information data, gestational age (GA), age at measurement (PCA), and birth weight (BW) were taken. To gain the time span between birth and data recording, PCA and GA were subtracted from each other. All this information was combined into a stability score. This score would indicate either an unstable, medium, or a stable patient condition.

All features are listed in Table 18. The normalized features with mean zero and standard derivation of one were combined into 3D tensors, which were fed as input into the deep learning models.

5.4.6   Preprocessing for deep learning

For the classification of the preterm infant sleep states, the neural network API Keras [264] was used with TensorFlow [265] backend. For sleep state classification, time series analysis was used. Therefore, the input was cast in the form of a 3D time

series tensor [samples, time step, features]. After testing, the time step was chosen as the total length of one recording session with the batch size set to 1. Thereby, long and short-term patterns can be recognized. To achieve uniform length, the tensors were padded to the length of the most extended session. Later, a masking layer and sample weight distribution of zero for the padded values was used to prohibit the padded values to influence the learning process. The data was separated into train and validations sets to exclude significant bias. The data was split per patient to reduce the bias for the train and validation process further. The split per patients was set to 70% training data and 30% validation data (ceiled). A 3-fold cross-validation process was used to ensure the proper generalization of the model.

As we have a majority (AS) and minority classes (QS, IS, CW), a class weight has to be calculated to balance this unequal class distribution. The sample weight was calculated depending on the sleep state and normalized to the majority class. Sample weight was used instead of class weight as class weight is converted to sample weights on the Keras backend. Using *sample_weight_mode temporal*, sample weight fulfills the class weight task and can as well be used for masking-padded-values with a sample weight of 0. So far, this can only be used for smaller datasets as *sample_weight_mode* does not work currently for the function *fit-generator*.

5.4.7   Classification models

Four different model types were compared: a deep residual model, a wide residual model, a wide residual model using transfer learning from sequential models, and the sequential models as standalone architectures.

The base residual architecture itself was adapted from the residual architectures ResNet [266] and ResNext [267]. Both approaches tackle the problem that an increase of model depth creates a sudden and rapid decrease of accuracy, which is not caused by overfitting but rather shattered gradients [268]. The shattered gradient appears in none residual networks with large depth like white noise. Both networks combine

multiple sequential models to one larger model. The sequential models thereby fit a residual map, which is easier to optimize than a larger model. The connections between the sequential models are ensured with skip layers performing identity mapping and feeding the output of a sequential model block into the next block (see Figure 36 and Figure 37). This process enables large networks with rather low complexity. As preterm infant sleep is a highly complex problem due to the many in- and exogenous influences on the sleep and ANS, residual architecture approaches might be able to constitute and handle this complexity. The ResNet and ResNext architectures were both developed for image classification and object detection. Therefore, they are using mainly rectified linear unit (ReLu) activated CNN layers coupled with dropout and pooling layers. For this chapter, the CNN and pooling layers were replaced with recurrent layers to capture patterns and connections in time series data rather than in image data.

Both approaches are similar in the core residual idea and compare equal to each other in object classification tasks with a slight advantage in floating-point operations per second (FLOPS), speed, and error rate of the ResNext over the ResNet architecture. As it cannot be directly deducted how the architectures would compare in a time series analysis task and in this specific setting, a comparison is appropriate.

Deep residual model: the deep residual network is made of an initiation block followed by five residual blocks of five connected GRU layers (Figure 37). Each connected block ends with a dropout [269] and a dense layer. The architecture is finished with a softmax activated dense layer. The initiation block consists of first a masking layer, which is needed as the data is padded. The masking layer is followed by a 1/2 dropout layer connected to a dense layer, which, both combined, function as a feature selection phase. This combination first randomly reduces the input nodes, trailed by reducing the dimensional space, forcing the focus on the most distinguishing input information. This acts as the replacement for the pooling layer used in the classic ResNet and ResNext architectures for reduction. The last layer of

146

the initiation is a batch normalization to avoid vanishing/exploding gradients by the scale of backpropagated weights.
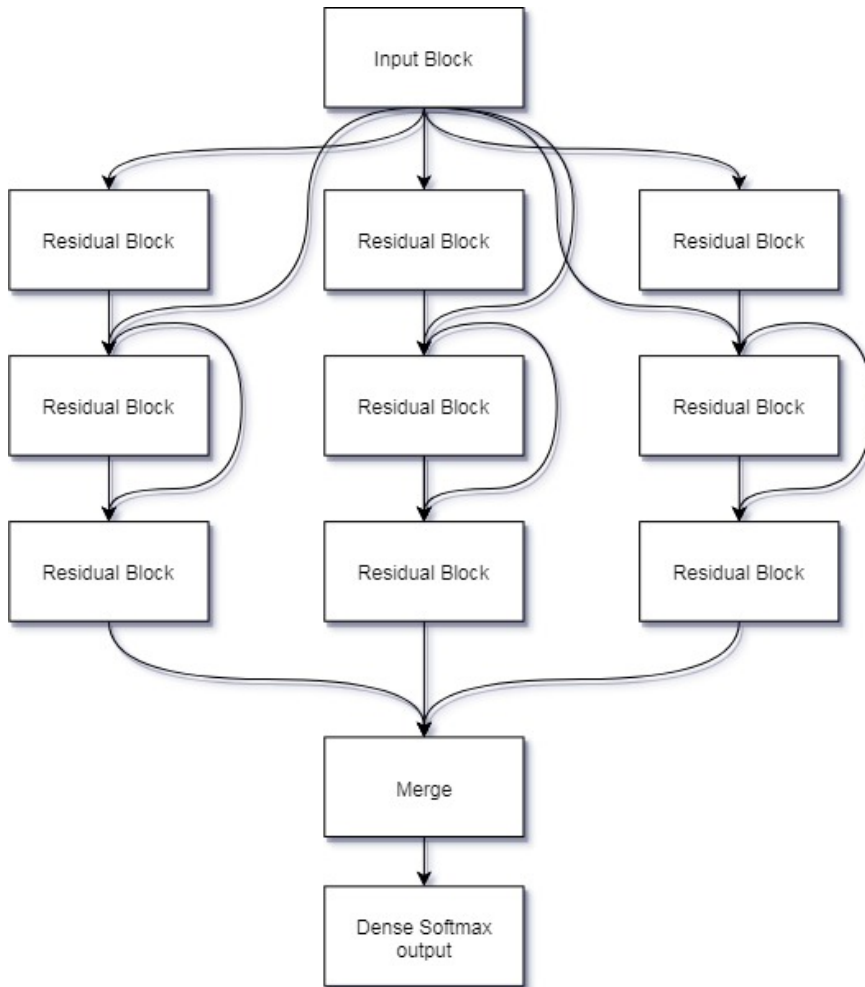
Figure 36 Exemplary Wide residual model structure.

Wide residual model with Initiation block of masking layer, dropout layer, and following dense layer. Afterward, the architecture is split into three paths, where each path consists of connected bi-directional gated recurrent unit layers which are later concatenated again. The layers are connected with feed forward and skip connections to help simplifying the network optimization. Each path uses different hidden units to incorporate more and less complex relations.
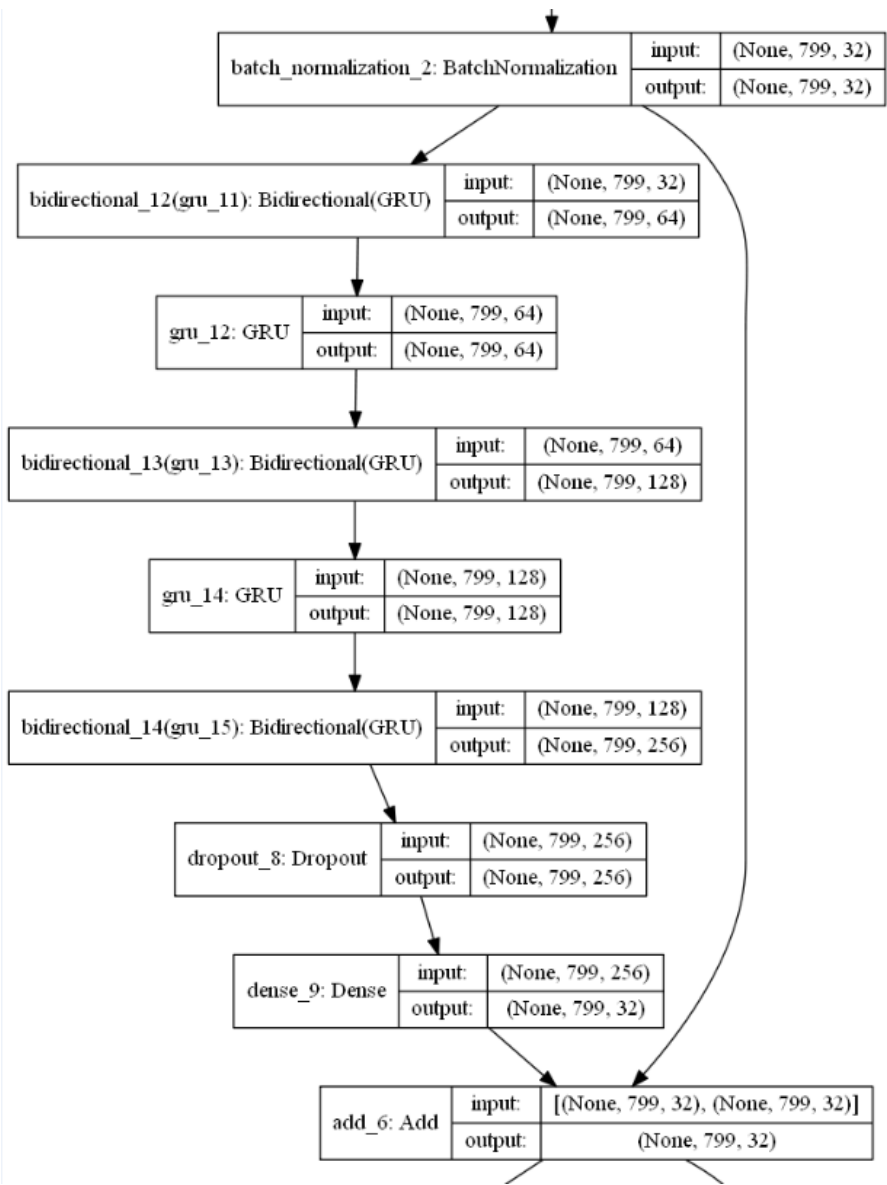
Figure 37 Residual block of deep model.

An exemplary block from the deep residual model. Here gated recurrent units and bidirectional gated recurrent unit layers alternate each other with increasing hidden units (neurons). Thereby, the hidden units increase from 32 to 256 covering simpler to more complex feature-state connections. The block is bypassed with a skip connection.

Wide residual model: the wide residual model (Figure 36) uses three parallel paths preluded by an initiation block. The initiation block is of the same structure as in the deep residual model. The parallel paths are each made of blocks from bi-directional GRU layers. Each GRU layer uses dropout and recurrent dropout to minimize overfitting. As direct regularization of the L2-norm, a kernel constraint is used over all axes. Direct kernel constraints work well in combination with dropout [269]. Each path has two of the described blocks which are connected via skip layers in the same way as with the deep model. All parallel paths are concatenated at the end leading to a dense layer with softmax activation to achieve final state predictions.

Wide residual model with transfer learning: transfer learning uses pre-learned information, resembled in the weights which are fixed in a model architecture. The fixed weights add to the training process without being changed during the training. Thereby, pre-learned information can be used to reduce the overall learning computation effort or improve the learning process by adding specific information. In image classification or object recognition, transfer learning is used to fix earlier learned universal information of shapes in images, e.g., general shapes in a face such as lines and edges. Top (later) layers then learn more complex compositions of such general shapes for a specific set of images, e.g., the composition of chimpanzee faces.

Here the same wide residual model architecture was used, but additional paths were added. The loaded weights of the pre-trained models were fixed into those additional paths Figure 38. The pre-trained models were trained on bi-class problems, always training two classes versus each other, learning the specific differences between only those classes. In the concatenation step, they are then used for decision-making. To avoid any bias, the bi-models are trained separately only on a fragment of the data, which is later not used for further training.
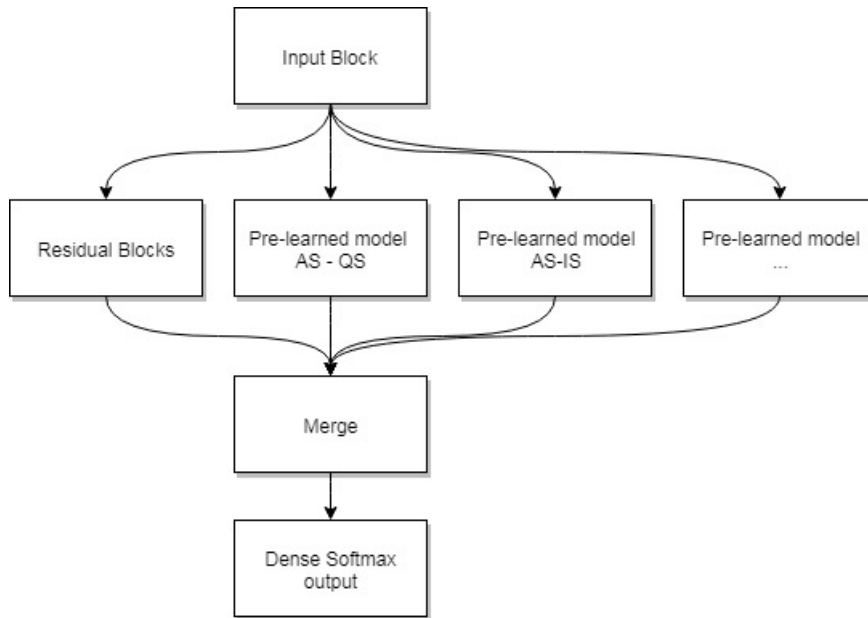
Figure 38 Architecture of the residual model utilizing transfer learning.

The input block and the Residual block are the same as in Figure 36. Also, pre-learned blocks of sequential architectures are added parallel and fixed until a various point. All weights are concatenated and added into a dense layer with softmax activation.

Sequential models: the sequential models, which are also used for pre-training, have an initial masking, dropout, and dense layer, which is followed by four bidirectional GRU layers. The models are closed with another dropout layer and a dense, softmax activated layer. The hidden Units for the dense and GRU layers were set to 32. All other parameters followed the main transfer learning model. The model architecture was compared in performance and speed to the same architecture using bidirectional LSTM layers.

5.4.8   Model parameters

The wide and deep residual model, use a range of hidden units for the GRU layers to learn and model a wider range of complex non-linear relationships in the input data

stream. The wide network uses a different hidden unit for each path ranging from 4 to 128 hidden units. The deep architecture increases the hidden units with each block from 32 to 256. The hidden units of the Dense Layers were set differently to accommodate the previously mentioned feature selection. The values ranged from 16 as roughly 1/3 of the input feature dimension and a power of 2, to a max of 47 representing the full input feature dimension.

To further handle the data imbalance, the earlier mentioned class weights were used in a weighted categorical cross-entropy loss function to increase the misclassification gravity for minority misclassification. Therefore, the normalized weights multiplied with the loss function where ll(s) is the a priori likelihood of s in the training data.

$$L(s)_w = L(s) \cdot (1 - ||(s)).$$

The *Sigmoid* function was selected as the activation function for each residual GRU/LSTM block. In the ResNet, and following residual architectures, ReLu function is used as the activation function. Using the ReLu activation function is very difficult and not advised with GRU/LSTM as it diverges, but mostly not necessary, as the gating scheme of the GRU/LSTM itself deals with the vanishing gradients. Therefore, the *Sigmoid* activation, which is optimally designed for the GRU/LSTM structure, can be used.

For the optimization algorithm, the Adaptive Moment Estimation (Adam) optimizer [270] was chosen as Adam shows to be generally very effective while also removing the manual setting of the learning rate and learning rate decay. The Adam also was tested against other optimization algorithms where it stood out superior.

As already mentioned, the timestep (or lookback) was set to the longest recording session and the batch size to 1.

To avoid overfitting, dropout, L1, and L2 regularizers were applied. The maximum dropout value was 0.6 before the results dropped off out of proportion. Combinations of L1 and L2 regularizations were implemented as kernel and activity regularizers in

different places. L1 was implemented mainly as an additional feature selection in the first dense layer with a value between 0.0001 and 0.001. The L2 norm was used mainly on each LSTM/GRU layer. The L2 norm for direct kernel constraint was set to 0.3 on each layer.

5.5 Results

Here, three different types of classifications are presented. The two-state classification, which is a classification between two states. The three-state classification, which tries to separate three states against each other such as AS, QS, and IS. And all-state classification, which tries to separate all states from each other.

The two-state classification with a sequential architecture shows promising results for using GRU and LSTM layers. Both show similar mean results with a difference in performance of 0.01 ± 0.02. Due to slight faster training with the use of GRU layers, all results are presented using GRU layers. The most robust performance is reached with the majority states AS and QS with a mean kappa over the folds ranging from 0.43 ± 0.07 to 0.40 ± 0.06 (Figure 39) and between QS and CTW with a mean kappa of 0.44 ± 0.01.

Then the combinations AS-IS and IS-CTW show similar results with 0.33 ± 0.03 and 0.32 ± 0.03. AS-CTW and QS-IS classification have the lowest performance of 0.28 ± 0.005 and 0.25 ± 0.03 (Table 19). Where Kappa score is defined as slight within $0 - 0.20$, fair between $0.21 - 0.40$, moderate with $0.41 - 0.60$, substantial between $0.61 - 0.80$, and as perfect within $0.81 - 1$ [232].

The classification of three states shows results (

Table **20**) between two- and all state classifications (compare Table 21). The majority classes AS and QS were compared with the minority classes IS and CTW. The mean performances in Table 20 indicate that the Majority classes are better differentiable together with IS resulting in a mean kappa of 0.35 ± 0.07 rather than

with CTW with a mean kappa of 0.33 ± 0.04. Both deep and wide residual models show similar results for all state classification with a mean kappa of 0.30 ± 0.06 and 0.25 ± 0.02 (Table 21). Same as with the sequential model, the majority states AS and QS are separated best, followed by QS and CTW using wide and deep residual models. The overall performance is lower than for bi-class classification. The use of transfer learning did not improve the performance even though the pertained models showed decent results. In contrary, it showed the lowest overall performance with a mean kappa of 0.13 ± 0.02 (Table 21) even though the performance of the pre-learned models were acceptable (see Table 22).
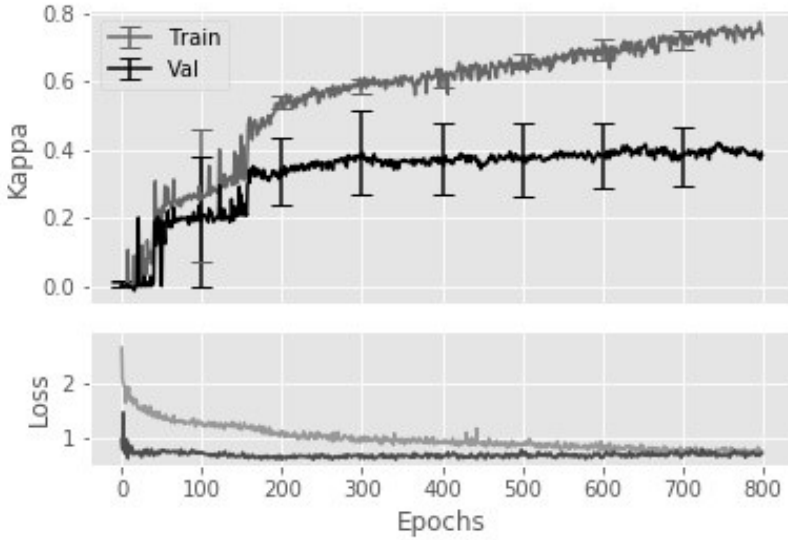
Figure 39 Mean kappa and loss of active versus quiet sleep classification over epochs. Kappa and weighted categorical cross entropy loss over epochs using a sequential architecture. Initial dropout is 0.2, and dropout/recurrent dropout per bidirectional gated recurrent unit layer is 0.5. Kernel constraint per layer is set with max norm 0.3. Hidden units of 32 is used per layer.

Table 19 Mean performance for bi-state classification using different model architectures.

The Kappa value of two-state classifications of different model types. The residual models used a kernel L2 regularization of 0.01 and an activity L2 regularization of 0.001. Initial dropout is 0.2, and dropout/recurrent dropout per bidirectional gated recurrent units layer is 0.5. Kernel constraint per layer is set with max norm 0.3. Hidden units of 32 are used per layer. The sequential model used four long short-term memory or gated recurrent units bi-directional layers with sigmoid activation, Adam optimizer, and weighted categorical cross-entropy loss function. The Kernel constraint was set to 3. 50% Initial and Recurrent dropout was used with both 0.5. Initial and Recurrent dropout was used. L2 activity and kernel regularization were set to 0.01 and 0.001 per layer.

State pairs	Residual Wide (κ ±std)	Residual deep (κ ±std)	Sequential (κ ±std)
AS-QS	0.37 ± 0.07	0.38 ± 0.04	0.43 ± 0.08
AS-IS	0.31 ± 0.03	0.30 ±0.03	0.33 ± 0.03
AS-CTW	0.26 ± 0.005	0.25 ±0.01	0.25 ± 0.03
QS-IS	0.28 ± 0.03	0.27 ±0.007	0.28 ± 0.005
QS-CTW	0.39 ± 0.005	0.40 ±0.001	0.44 ± 0.01
IS-CTW	0.30 ± 0.04	0.29 ±0.03	0.32 ± 0.03

Table 20 Mean kappa performance of sequential models for three state analysis.

Mean performance of three-state classification using a sigmoid activated model with four bi-directional gated recurrent units layers, Adam optimizer, weighted categorical cross-entropy loss function. The Kernel constraint was set to 3. 50 %. Initial and Recurrent dropout was used with 0.6 and 0.5. L2 kernel regularization per layer was set to 0.01.

State pairs	Kappa
AS-QS-IS	0.35 ± 0.07
AS-QS-CTW	0.33 ± 0.04

Table 21 Kappa performance on all state classification for different models. Mean kappa performance over different models using the same parameters as described in the other tables.

Model	Kappa
Deep Residual	0.30 ± 0.06
Wide Residual	0.25 ± 0.02
Sequential	0.25 ± 0.05
Wide Residual using Transfer learning	0.13 ± 0.02

Table 22 Kappa performance of sequential model architectures used for transfer learning. *Kappa results without cross-validation using the same parameters as the sequential model in* Table 19*, utilizing bidirectional gated recurrent unit layers.*

State pairs	Kappa
AS-QS	0.51
AS-IS	0.36
AS-CTW	0.27
QS-IS	0.29
QS-CTW	0.55
IS-CTW	0.38

5.6 Discussion

5.6.1 Features and patient demographics

The feature set was partly used in Chapters 3 and 4. It was reused and adapted as it showed a good representation of the underlying processes of preterm infant sleep. Movement features were included as movement is generally a strong differentiator between sleep, wake, and caretaking. For future research, improved measurement techniques for movement detection, as presented by Joshi et al. [271], or enhanced extraction methods for ECG would probably improve classification further. Additionally, patient information features were added to the set as it was noticed in previous tests that outlying values of patients can seriously influence the performance of classification. Such values are, in general, not random but often occur at very young, immature preterm infants and/or with low birth weight. Also, it makes a difference in the development at which age a preterm infant was born and at which timespan after birth the data were recorded. If the measurement takes place at the same time with different GA at birth (or vice versa), the development state and consequently the feature appearance can look sufficiently different to influence the learning. The same holds for the birthweight. A heavier baby tends to be more stable. As the values of age, age difference, and weight are almost continuous data, they were categorized into a stability score between 1 and 3. With a significantly larger dataset, either the values can be used directly, or a finer grid can be used to categorize the preterm infants. Further, if sleep cycles are to be measured, we would recommend splitting the patient groups at the age of 38 weeks GA increasing stability of training, as the not fully developed CNS as the circadian oscillator will probably lead to more erratic sleep cycles in preterm infants younger than 38 weeks GA. In Chapter 4, it was noticed that the use of respiration devices influence the classification performance. Unfortunately, it was not possible to gather this information for all patients.

5.6.2   Parameter choices

To avoid overfitting, various settings were applied with dropout, recurrent dropout, L1, and L2 norm for activity (Ar) and kernel (Kr) regulation. Dropout and activity regularization had the most effect on overfitting. Different combinations of these regularizations were useful in reducing overfitting. Here only two are mentioned as examples. Either using overall lower dropout (e.g., 0.3) in combination with a stiffer Kr and Ar L2 norm (e.g., 0.01) without any other regularization in the initiation block helped fighting overfitting. Another variant is to use an L1 norm as kernel regularization in the first dense Layer to help with feature selection in the initiation block in combination with an overall dropout/recurrent dropout of 0.5-0.6 but no other Kr or Ar in the following layers. Choosing too high values for the dropout and/or regularization would lead to a drastic reduction in overall performance on the validation data. There were plenty of combinations that all resulted in reducing overfitting. Nevertheless, further investigation and proper comparison will go beyond the scope of this chapter, as overfitting was not the primary problem in this analysis.

The lookback was chosen as the total duration of one session for the LSTM/GRU as long-term sleep cycles can influence the overall learning process. As LSTMs/ GRUs can only learn the variations in time on the information of one batch, long-term patterns such as total sleep cycles or specific sleeping patterns need at least 30 min of data up to 70 min [53]. In regular cases, sleep states changes follow the pattern wake-AS-QS-AS-wake with IS patterns in between. Irregular patterns are, for example, wake-QS-(AS)-wake. This pattern is called a stress sleep pattern, showing signs of the preterm infant's immediate need for rest. With more of such recorded patterns outside the norm, future research could try to detect anomalies in sleep cycle patterns to inform the responsible caretakers. Either regular or irregular cycle patterns cannot be learned with batches of insufficient length below at least one sleep cycle.

5.6.3   Classification performance

The classification between the majority classes AS and QS shows moderate performance with kappa 0.43 ± 0.07 and generally promising results. For a clinical monitoring device, this would not be an acceptable performance, but these results should be seen as proof of concept for an unobtrusive automated monitoring system. Especially with the earlier described benefits of ANNs, these results are a foundation that has a high chance of increasing in performance, stability, and generalization with an increasing amount of accessible data. To put the result in context, it has to be beard in mind that the general interrater variability is relatively low in that specific population. In early studies about the reliability of polysomnography in term infants, the kappa score reached 0.68 [272]. For adults, the mean kappa score after the Rechtschaffen & Kales standard also reaches 0.68 and following the AASM standard 0.76 [273]. There is no overview accessible for preterm infants, but it is very likely to be lower than for term infants.

It is difficult to directly compare the here presented result to recent results from other groups as they are mostly based on EEG analysis, which is optimal for sleep state analysis as they directly represent the state of the autonomous nervous system. Koolen et al. [101] presented good results in 2017 using a support vector machine on EEG signals with 85% accuracy for AS and QS separation. Latest results using EEG signals were presented by Ansari et al. [247] using a CNN to classify QS and NonQS with a ROC-AUC of 0.92. Also, Dereymaeker et al. [102] were able to detect QS with an AUC of 0.97 based on EEG. The most similar approach from Isler et al. [191] used only respiration signals for successful classification of AS and QS with an agreement of 78% to 90% for AS and QS separation. In a previous publication of our group [229] AS and QS were separated with a ROC of 0.87 based only on ECG features. As more states were included for separation, performance decreased. Fraiwan et al. [190] tried to separate AS, QS, and wake with a performance of 63 % to 75 % using also EEG analysis. Also, it has to be kept in mind that the here used

kappa performance takes into account the unequal distributed states with increased expected accuracy. Using only the accuracy measure would be biased towards the majority class. ROC-AUC could have been used to describe the performance of the trained ANNs, but as we have a multi-classification problem, ROC determination for the ANNs is more difficult as more than one output node threshold has to be varied, generating a multi-parameter problem. This could be handled by creating ROC curves for each individual class-pair. A better approach is to scale the bias weights of the first hidden layer note as proposed by Woods et al. [274]. However, when using a ROC-AUC analyses the number of data points per class should be roughly equal, which was not the case here. Precision-Recall analysis [275] is a suitable alternative in this case, but none of the compared studies used precision and recall, which has let us to choose the Kappa statistic.

Summarizing the results and methods used for sleep classification, Ansari et al. [247] can be considered the current state of the art for preterm infant sleep state analysis as they are using the latest analysis methods on EEG signals, which are the standard signals for sleep analysis, and achieved very high classification results for QS and nonQS states. Both of which are important for neural development monitoring.

Our moderate performance on AS and QS classification is also promising as the state distribution of AS and QS is one of the primary indicators for neural development in early preterm infants. The bi-state classification for AS and QS can be utilized for neural development indication and clinical decision support. As the minority states naturally occur less, they are of lesser importance to the course of development in the early stages of preterm infancy. In term infants, wake versus sleep becomes more important, but at that point, wake also has a more significant presence, which can be utilized for training.

Compared to the current state of the art by Ansari et al., our method still lags in performance for QS and nonQS (AS) classification. However, it has to be kept in mind that ROC-AUC performance measure can be overestimating due to imbalanced

data for the majority and minority classes, which is incorporated in the here used kappa score. Again, a direct comparison is not in order as the methods are based on different physiological signals and follow a different goal. We believe that Ansari aims to classify QS and nonQS states in preterm infants with the highest possible performance and therefore chose the most promising and sleep-related physiologic signal to analyze. This thesis, on the other hand, tries to classify preterm infant sleep states with a novel approach specifically based only on ECG derived features to ensure a more accessible and constant sleep monitoring. To emphasize our motivation again, the downside with using EEG for sleep analysis is the use of auxiliary sensors in addition to conventional monitoring sensors, which should be kept to a minimum as the preterm infant skin is highly sensitive. ECG, on the other hand, is a standard measurement in the NICU; therefore, this system is easily implementable in the current clinical practice. As the ECG signal is less interlocked with sleep compared to EEG, it is consequential that the results will fall behind, especially as this is the first ANN approach of this signal modality and patient group to date.

Nevertheless, it is insightful to look at the publication from Ansari et al. as they also used an ANN approach and outperformed all former attempts on EEG based QS and nonQS sleep state classifications. This indicates that ANN approaches for preterm infant sleep classification are capable of outperforming traditional signal analysis and standard machine learning algorithms in general. Translating this to ECG signal analysis, we believe that with further tuning and additional data, also ECG based analysis with ANNs can reach outstanding performance levels. To our knowledge, this is the first approach to use an ANN approach in combination with ECG derived features for preterm infant sleep analysis. With further research, it is reasonable to believe that ECG based sleep analysis will become on par with human annotators and maybe EEG analysis.

The overall low performance on all state classification has to be explained by two rationals — first, the general difficulty in separating human physiological events that

often show only a nuance of difference between its states, especially in preterm infants. This manifests in the existence of the sleep state IS, which, by definition, is a mix of the main sleep states AS, QS, and wake. IS incorporates patterns of all those sleep states, making it very difficult to pinpoint the beginning and the end of neighboring states. Over the course of development, IS occurs less frequent, leaving a clearer picture of sleep state boundaries. This can also be seen in the generally low interrater variability, showing that trained observers have difficulties in uniformly identifying the sleep states. Some annotators had years of experience, having seen plenty of training data. The other core problem in preterm infants is the immaturity of the autonomic nervous system. This immaturity of the regulatory mechanisms leads to instability in regulation and control of ex- and internal stimuli. In preterm infants, those instabilities result in common heart rate decelerations that are not connected with an autonomous response to such as sleep state changes [208]. Nevertheless, they can easily be misinterpreted as such. The combination of those often only nuanced differences between autonomous states and the general instability in preterm infants inducing non-state-related heart rate changes makes the classification of preterm infant sleep a difficult task.

The second rationale is the low amount of data, especially for the minority classes. Preterm infants sleep around 70% in 24 h [39]. Therefore, the wake state is naturally underrepresented, and caretaking also takes only a portion of the day. Interestingly, QS is as well underrepresented among the three datasets but shows enough difference to AS to be sufficiently distinguishable. Generally, CTW shows differences in the patterns to the QS and IS states resulting in a heightened performance for QS - CTW and IS - CTW classification despite the lack of data. The activity in both AS and CTW, and thereby signal similarity, makes it harder to classify, resulting in the lowest performance. Another influence could be wrong annotations, as during CTW the preterm infant moves similarly to AS. If the eyes are not open or caretaking cannot be directly observed in the video frame, CTW could be mistaken for AS. Furthermore, AS – IS is better separable than QS – IS, which could be due to the reduced breathing

and movement during IS. This reduction results in similar patterns for IS and QS, making the correct classification more difficult. Same as before, another reason could be the manual annotation. As changes to the heart rate variability indicating a state change without visible clues like twitches, eye movements, or rapidly changing breathing, IS could be easily mistaken for the onset or continuation of QS.

ECG-based sleep state classification is far less studied compared to EEG. Most of the cardiorespiratory based work in preterm infants considered AS/QS or wake/sleep states, as all state classification is a challenging matter. To our knowledge, all state classification has not been investigated regarding classification, despite in our group. Reliable all state classification in preterm infants have yet to be presented.

The tri-state classification is expected to show slightly lower performance than the bi-state counterparts. Here, the more difficult states, IS and CTW, reduce the combined performance. The slightly higher performance between AS - QS - IS, despite IS being a more difficult state to differentiate, has to be explained with a higher amount of training data. Despite the lack of data, AS - QS - CTW classification shows only slightly reduced performance as noise, instability, and increased movement dominate the ECG patterns and create a clearer differentiation.

The performances on all state classifications using the residual approach are underwhelming. Nevertheless, the use of a simpler, sequential model also did not generate reasonable results. Here again, the data to train on, especially for the minority classes, was considerably small with very early, unstable, and fragile patients. However, due to higher performance on the majority classes AS and QS, general different feature modalities between the single states, and the very high difference in the amount of data between majority and minority classes shows that the problematic performance is directly linked to the data amount and not fundamental problems with the used model architectures. This generally indicates again that the correct track is to utilize deep learning for preterm infant sleep classification as deep learning has mostly a higher performance potential than machine learning with increasing data size, as

explained before. Generally, all-state classification is not of main importance for early preterm infant development monitoring but is vital for a holistic view on the patient's sleep rhythm and possible predictions of sleep patterns.

To summarize, the separation of AS and QS show the general potential of using deep learning for sleep classification based only on ECG derived features, as stated in the research hypothesis. Similar results for the main sleep states were achieved compared to classic machine learning approaches [229], [249] Nevertheless, for a complete picture and overall sleep monitoring, a wider study is necessary to gain a stable model including training on extreme outliers.

5.6.4   Model architectures

Recently, GRU networks were found to have similar performance as LSTM networks. The GRU network uses less computational power than the LSTM network, as it generates fewer parameters. Nevertheless, both units perform almost equally, and one cannot be generally favored over the other. We tested architectures with both units and found that in our case both, LSTMs and GRUs layer use, performed equally. Due to lower calculation time, GRU layers were used further on.

The wide and deep residual model architectures show similar results (Table 21). The total amount of layers after the initiation block is similar to ten layers in the deep model and 12 for the wide model. Both architectures consider low and high complexity relations between the features and sleep states with increasing hidden units. In the ResNext model approach, the idea was also to introduce cardinality, an increase of parallel structures per residual block. At this point, we only used a cardinality of one as the model architecture could not be enhanced further due to overfitting spiraling out of hand. A deep model with 25 GRU layers after the initiation block was run with massive overfitting problems. Compared to a model with 4 GRU layers after the initiation block, the residual structured models showed weaker performance (Table 19). The overall disappointing results of the residual architectures

show that probably they overreach with the complexity of the analysis on that task at hand. The complexity cannot be put to use, as too few training examples for the more complex feature conjunctions are available. Additionally, the increased complexity tends to lead to overfitting due to training onto complex appearing noise structures. Regularization and dropout have to be set in place that can lead to a performance restriction. A solution to finding the right model architecture might be an evolutionary approach for architecture search. Generally, using this novel combination of a ResNet/ResNext architecture with GRU layers and corresponding activation is promising and might be wide-ranging and valuable in highly complex time series analysis, such as adult sleep analysis, where vastly more data is available.

Transfer learning did not result in an acceptable performance and did not improve the performance as hoped. Probably this is connected with the fact that the saved weights from the pre-trained models were taken from a single fold. Even though they showed reasonable results (Table 22), they lacked generalization on the validation data. Secondly, the data was further reduced by splitting the data pool for pre-training and later transfer learning for bias control.

5.6.5 Strength and limitations

It is a challenge to gather sufficient data in very preterm infants in the high-risk NICU environment. With a mean age of 29 ± 4.6 weeks GA, the study group is very realistic and generalizable for a NICU population. Human annotators performed the annotation of the dataset with a moderate interrater-variability. This may limit the performance of an automated system from the very beginning.

On the other hand, the trained model incorporates the different experiences and knowledge of different annotators creating a more stable, integral, and reliable model again. Interesting would be a wide range of annotators and annotation styles in future research, backed with sufficient data, to incorporate the derivations of different annotation techniques in the model. Due to this limited amount of data, the ANN approach could not develop its full potential; nevertheless, compared to the general

low interrater variability in this patient group, the results are acceptable. As said, the general strength of this approach is the potential to enhance its performance with increasing data.

Further, we believe that we have not yet found the features which describe the preterm infant sleep states in full detail representing their complexity. As novel feature extraction methods, such as using a CNN based on spectrogram, were used successfully in different ECG based applications (e.g., atrial fibrillation [276], [277] or arrhythmia detection [278]), those methods used for a better representation of the ECG signals, and a better way of extracting richer features from the ECG signals merits investigation.

5.6.6   Future perspectives

As the main reason for the low performance of the all-state classification can be linked with the low amount of data, considering the vast difference in preterm infant stability and development, we suggest that more preterm infant data has to be gathered to surmount the threshold where data size becomes not the primary influence on performance. Following, the gross amount of needed data is estimated. The assumption is that the classification performance would be similar for the data poorest state if such a state would have the same amount of data as now the data richest state. With wake as the data poorest state having 6.6 % (not considering caretaking as it results from external influence) and AS as the data richest state now with 52.41 %, the needed amount would be seven times higher as here present. This results in roughly 200 preterm infants with the same mean recorded time of 4 h. Alternatively, 50 patients with 24 h recordings, which would be more optimal regarding full sleep cycle analysis. As ANN performance is not linked linear to the data amount of a single state, it can be assumed that less data is sufficient.

Even though the transfer learning approach did not show the intended results, we suggest that term infant data instead of rare preterm infant data is used for pre-training

as signal patterns and sleep architecture are still very similar to preterm infants and much more data is available for this patient group.

Another approach could be to look at unsupervised learning for preterm infant sleep staging. So far, we rely on human annotations, which are in itself not perfect and show large interrater variability. The general shift of data patterns from unsupervised learning could indicate brain development in the same way as classified state distributions from supervised learning. Unsupervised learning would demand even more data but will reduce the necessity of manual annotation. Not annotated, preterm infant sleep data is already freely available, for example, from the CHIME study [279].

5.7 Conclusions

Active and quiet sleep can be moderately separated using a deep learning approach solely using ECG derived features. Nevertheless, all state classification is, so far, not possible and is hindered mostly by limited preterm infant training data as well as training data of very young and unstable patients. There is a level of data that has to be reached so that the data amount is not the significant factor for performance. For highly complex time series analysis, backed up with sufficient data, an RNN-ResNet architecture with sigmoid activation can be chosen for a deep network approach, avoiding the problem of shattered gradients.

After diving into unobtrusive signal modalities and classification methods, we now want to take another look at the entire concept and finish with the conclusion of our research in the final chapter.

6   Conclusions and future perspective

In the here presented thesis, we looked into different signal modalities and methods to classify preterm infant sleep. We focused on the use of ECG even though alternative unobtrusive methods are available for automated preterm infant sleep classification. To name only two, camera-based solution or motion detection solution could be considered if we are looking at the development of ANNs in the domain of image recognition. The ANN techniques will develop further and will likely outperform human classifiers in the near future. However, the primary motive why we chose ECG as the base for our research was the widespread availability of ECGs in the NICU. This availability reduces the need to implement new signal acquisition systems into the NICU. In particular, a low-level solution is preferable in low-income countries, where camera systems might not be as available as in high-income countries. Developing an algorithm that needs only basic signal inputs to operate enables a majority of NICUs, independent of their technical status, to utilize sleep and development monitoring.

6.1   Sleep state classification

In the thesis presented here, we demonstrate that preterm infant sleep can be separated automatically with the use of different approaches. Notably, the most prominent sleep states AS and QS can be separated rather well with a high AUC of 0.87 ± 0.42 using and nonlinear Kernel SVM. The high standard derivation results from two outlying patients in the early threshold segment. The curve overall has a standard curvature and does not artificially increase the AUC by spikes or outliers. Using a random forest approach, AS and QS classification performance reaches Kappa of 0.42 ± 0.26. Using capacitive ECG, the performance reached a Kappa of 0.59 ± 0.16. This result is comparable to the nonlinear SVM performance. With a sequential RNN model Kappa of 0.43 ± 0.08 was achieved, which compares to the random forest ECG results. Adding the CTW or IS states to the task, reduced the

performance in most cases. CTW reduces the performance thereby less than adding IS by losing around 0.1 and 0.2 Kappa values respectively. All-state classification also reduced the performance to Kappa values between 0.3 and 0.38, leaving all-state classification not suitable for clinical use in the near future. AS and QS classification performance is also not yet sufficient for clinical use, but we are convinced that with further research and relatively little effort a clinical solution is within reach.

In chapter 3, linear features were derived from the HRV in the time and frequency domain. Only one ECG feature was used in the form of beats per epochs. Those basic features were used in combination with a nonlinear kernel SVM classifier. Using the SVM was successful but needed a considerable amount of fine-tuning and manual parameter choice. We attempted to limit the manual choices by using a greedy forward search to find the best feature combination. Given that over one million combinations are possible out of all 20 used features, manual trial and error were impossible. Unfortunately, this method is very unlikely to find the optimal combination as already the starting point can be ill-chosen. The forward search resulted in a five-set feature combination, including time (NN20, SDNN, and pNN20), frequency (total power), and ECG features leading to the best performance. In Chapters 4 and 4.3.8, more features were added to determine unsteady signals indicating movements, jerks, and jitters, or generally active states. When we included those additional features, we obtained better representation than before, as can be seen in Figure 35. The better representation may also be related to the use of a different data set. However, as the subsequent features with slightly lower representation from the feature ranking were again the same linear features, we can assume that the improvement with the added features holds true. The results themselves cannot be directly compared due to different learning approaches, distribution of states, different amounts of training and testing data, and different annotators presenting different base conditions. In Chapters 3 and 4, using linear SVMs did not succeed, and instead nonlinear kernel SVMs were used before the RF classifier outperformed the SVM approach. This result confirmed once more the difficulty of preterm infant sleep state separation. The

premature autonomous nervous system shows only a slight difference between the states leading to vague state borders and making a clear separation difficult. A very good combination of right features, parameters, and machine learning algorithms to discriminate those vague state borders is difficult to find. The sleep states are influenced by a variety of interconnected systems, as outlined in Chapter 1.6.3. Those interconnections are partly represented though highly complex and nonlinear combinations of the features used in our approach, reflecting the influence of preterm infant sleep on the ANS. This description of the problem calls for ANNs due to their ability to find nonlinear and highly complex dependencies in the input data. In small data sets, the use of classic machine learning solutions is generally advisable, as presented in Chapters 3 and 4. Very likely, they can outperform ANNs at small amounts of structured input data. This general assumption is confirmed when we compare the results of Chapters 3, 4, and 5, where none of the RNN models outperform the classic approaches. However, we must remember that the underlying problem and difficulty in distinguishing the sleep states remain the same with a growing amount of available data. What changes is the opportunity to train on an increasing amount of examples. Thus, the ANNs are in clear advantage over classic machine learning algorithms. ANNs are hardly limited in the dimensionality to describe a given problem compared to classic machine learning approaches. With such a divergent problem as preterm infant sleep, high dimensionality is needed to describe the problem in full.

In this thesis, we have laid the ground for the use of ANNs in the problem of preterm infant sleep and achieved similar performance to traditional machine learning. As soon as more data become available, the models can be retrained, resulting in increased performance. As described in Chapter 5.6.6, about fifty 24 h recordings should yield very good performance for an all-state classification. An additional benefit of ANNs is the easy implementation of pre-learned models to boost the base knowledge of the ANN on which specific learned information can significantly increase performance. As in Chapter 5.6.6, we suggest utilizing more easily available

term infant sleep data to pre-learn models. ANNs are often praised for the reduced need for feature engineering, as feature engineering requires domain knowledge, knowledge on signal processing, and time. These requirements are only true for CNNs and not for RNNs, as used in this thesis. As we are positive that not all autonomous relations and effects are represented in the features presented here, we also suggest to further examine feature extraction using a CNN network. Moreover, the visual description of the ECG patterns is under-represented in our feature set. The measures of complexity can also be considered a pattern description. For future approaches, we highly recommend adding CNN feature extraction in addition to features built on domain knowledge. In Chapter 5, we present several deep learning architectures. None of them prevails over the others by far. For the selection of the chosen architectures, the optimal architecture might not have been found yet. The developed automatic machine learning algorithms (autoML), such as AutoKeras [280] and AdaNet [281], can help create network architectures with reduced computational costs to find optimized network architectures for a given problem. AutoML has proven to be very effective on known, public datasets such as CIFAR-10 [282], MNIST [283], and FASHION [284]. Future research should investigate autoML approaches to optimize the data structures on preterm infant sleep. Further, we have described the highly complex interactions of the cardio-respiratory system in chapter 1.6.4; using many of the measurable responses for multi-modality ANN input could describe the sleep states in more holistic and robust way than only focusing on HRV. Using multiple input modalities will also allow including information about the sleep states that are not existent in the ECG signal and therefore cannot be retrieved with the most complex model architectures. However, unobtrusiveness should be kept in mind.

To summarize the sleep classification, we are confident that, based on the results presented in this thesis, a continuous AS-QS monitoring system might be established in the NICU after further investigation and improvements. Unfortunately, no full state monitoring system can be implemented at this point. Substantial research effort

has to be made into all state classification before a monitoring system becomes realistic in a clinical setting. The focus should thereby be put on the IS state classification, as IS classification performed poorest with all compared methods. Our presented results are an advanced starting ground and booster for further advances in preterm infant sleep research.

6.2   Unobtrusive sleep state monitoring

First, we demonstrated the possibility of classifying AS and QS using different classification methods with the use of classic ECG signal recording. Second, we determined that real unobtrusive signals from a capacitive ECG are feasible for sleep state classification. With capacitive ECG signal recordings, we verified that the same classification results can be achieved as with a traditional ECG arrangement, referring to the human annotations. The capacitive ECG signal was more prone to noise-induced by movement as the capacitive elements are sensitive to area, distance, and dielectric changes. Nevertheless, the increased noise did not harm the classification performance, as this type of noise itself is information. Noise represents an active or passive movement. The amount of movement can be related to AS, wake, or caretaking. On the other hand, electronic or electromagnetic noise such as thermal, or 1/f noise have the known impacts on the recorded signals and analysis. The used capacitive electrodes were implemented in a NICU mattress fulfilling all clinical standards and having no influence on the preterm infants whatsoever.

The mattress itself is not of specific material regarding industry standards and production. It has to match to hygienic clinical standards, which include withstanding alcohol cleaning solutions. The mattress cover should be skin-friendly as the preterm infant skin is extraordinarily fragile. Preterm infants generally sweat for thermoregulatory purposes even though the efficiency on thermoregulation is poor [285]. Before 36 weeks GA, preterm infants do not initially sweat but build up the ability to sweat within typically 13 days after birth. The intensity of sweating and

initiation as a response to increased temperature depend on age. The younger the age, the higher the threshold to initialize sweating. Subsequently, a mattress is not required to absorb large amounts of sweat in a temperature-controlled incubator. When replaced or cleaned regularly, bacterial cultivation in the mattress cover is inconsequential.

The ECG signal composition, with the rules described in Chapter 4, and preprocessing of the signal can be hardcoded into the mattress hardware to enable simple connections to preset systems for ECG sleep state classification.

In summary, the combination of a successfully trained deep learning model and capacitive electrodes for ECG can yield an unobtrusive monitoring system that can realistically be deployed for preterm infant sleep analysis.

6.3   Patient group

In this thesis, three different patient groups were presented. The age ranged from very early to late preterm infants (see Figure 40). Most preterm infants were born at a very low gestational age, with the majority below 30 weeks GA. Measurements had to be made at 1 or 2 weeks after birth to ensure the stability of the patients. In some of the very early preterm infants, the measurements had to be postponed up to 8 weeks before stability was reached (see Figure 41). Stability is also related to weight. Following the WHO's definition [1], all of the preterm infants analyzed in this thesis are small for their GA with the maximum at 2500 g (see Figure 42). The majority is actually of very low or extremely low birth weight with less than 1500 and 1000 g, respectively.

Age at the time of recording is not the only point of interest for sleep classification. Preterm infants develop differently depending on the age at birth. As explained in the introduction, the earlier the preterm infant is born, the more comorbidities and general lower health conditions are to be expected. This expectation is also related to
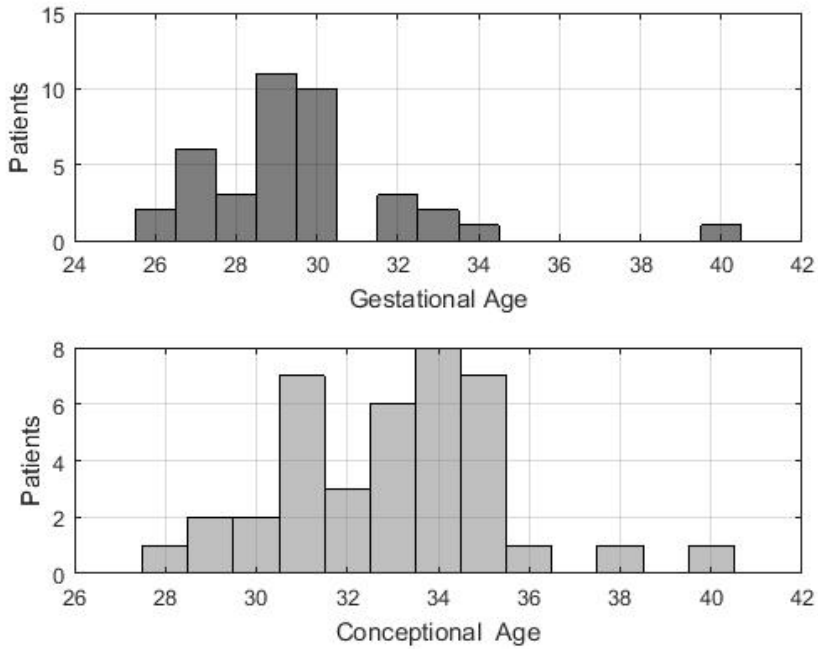
Figure 40 Number of patients sorted by gestational and conceptional age.

sleep staging because comorbidities, such as lung diseases, affect the vital sign parameters, which are used for classification. More cases that are unusual reduce the performance of the classifier, especially when only small datasets are at hand. In a large dataset, where sufficient "regular" cases are available, many outliers are welcomed as the trained classifier tends to generalize better. Generalization on outliers is especially important in the patient group of preterm infants as this patient group is predisposed for special cases and outliers. Interestingly, even though we consider our dataset as small, the results were similar between the different datasets, showing the good generalizability of the trained models. Alternatively, as mainly AS and QS were separated well, AS and QS are defined prominently despite the stability and overall health status of the patients. Which effect is more prominent should be further investigated.
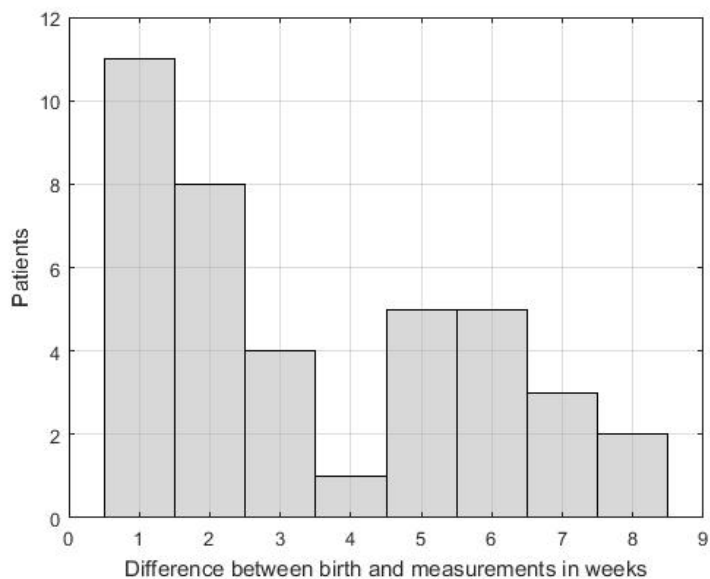
Figure 41 Difference between the time of birth and the time of recording in weeks.
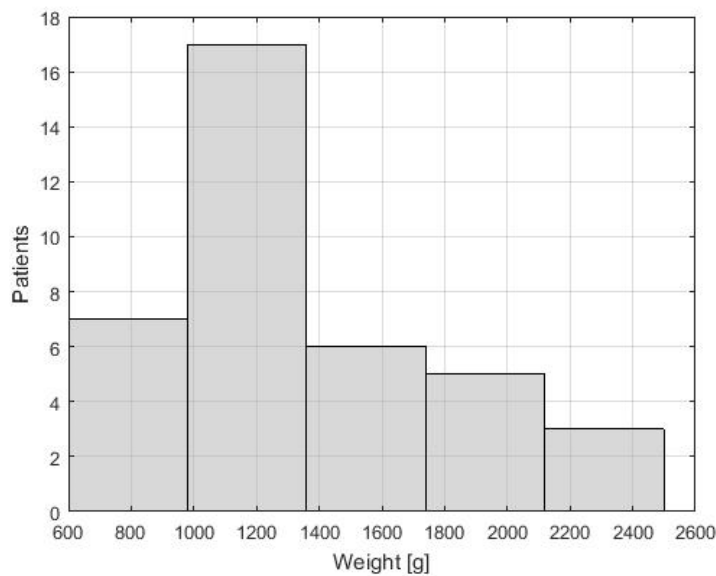
Figure 42 Weight at birth of patient group presented in this thesis.

6.4   Annotations

The datasets were all annotated by different annotators. For the datasets used mainly in Chapters 3 and 4, the annotators were trained by the same instructor who also was involved in the annotation oversight. They modified the behavioral annotation style of Prechtl [197] and adapted it to preterm infants. They watched mainly videos of the preterm infants, identifying subtle movements, changes in breathing, and the cycle of states. Moreover, the provided ECG signals were partly considered. Otte et al. [231] demonstrated that the annotations are on par with the gold standard when used on term infant sleep classifications. The publication is in progress at this point and should soon be announced on Researchgate (researchgate.net/profile/RA_Otte). For Chapter 4, the overall training time of the annotators was shorter than for Chapter 3. Whether this difference resulted in any major discrepancy between annotation qualities cannot be said as many other factors, such as quality of the data, the accuracy of the annotator, the provided working environment, the provided annotation tools/software, and others, come to play when comparing the quality of annotations. In general, no major difference regarding discrepancies in annotation decisions, training quality, annotator focus, general understanding of sleep state signs, and other indications of quality were observed. In all annotations, rare sleep cycles were annotated; in particular, "stress sleep," where QS follows directly after a wake period. As we had no direct insight into the annotation process, we could not determine whether this appeared more often than assumed or false annotations occurred in one or the other case. In both cases, these scenarios could lead to reducing the classification performance in chapter 5 when the LSTM trained on more often occurring outliers of long-term patterns. The SVM and random forest algorithms in Chapters 3 and 4 are not affected as long-term dependencies are not learned. In the second case, the sleep state was wrongly annotated regarding the underlying information in the vital signs, thereby falsely training the classifier and leading to reduced performance. With increasing data, the impact of those cases reduces.

6.5   The beneficiary of sleep monitoring

Information is the key to success. We argue that anyone who can derive information from continuous monitoring is the beneficiary of such a system. More specifically, pediatricians and nursing staff would derive the main, enabling benefit from such a system. However, a good pediatrician or nurse can most likely determine the developmental status of a baby by other parameters such as movement, response to stimuli, increase in size and weight, reduction of respiratory problems, vital sign stabilization, overall strength, and other indications. This might challenge the advantage of the technique described in the thesis. We observed nurses trained in the "newborn individualized developmental care and assessment program" (NIDCAP). Those nurses would sit next to a preterm infant and record its behavior on a prepared form to establish the state of the preterm infants' neuro and motor development. They would continuously fill in a prepared form throughout several hours. This procedure was considered by the NIDCAP nurses to be a very exerting process. This process gathers information on one particular preterm infant, at one particular point in time, and at the same time requiring the attention of one NIDCAP nurse for several hours. Replacing all of the mentioned observation skills by an automated system will not be realizable in the near future, but providing the caretakers with additional information of parameters, which are difficult to visualize, such as sleep state cycles and sleep state distribution are realistic goals. Adding this information automatically to the patient information will add tremendous aid to the daily routine of the caretakers. Long term ambitions should aspire to free up clinical staff by implementing a fully autonomous monitoring system. This will not replace clinical experts, but rather shift their valuable time to more pressing tasks and primary duties: the care of their patients.

Another interesting approach is the use of an automated monitoring system not only for the caretakers but also for steering other processes automatically in the NICU. Very interesting could be central apnea prevention. Internal temperature control due

to the process of controlling total body heat loss leads to instabilities and central apneas in sleeping preterm infants [286]. In general, central apnea is more frequent during AS [287]–[289]. An automatic adaption of the incubator temperature to the sleep states may reduce the control mechanisms, which try to reduce total body heat loss and consequentially central apnea.

More application might be to adapt the patient room lighting based on the sleep state. Furthermore, the alarms can be silenced at the bedside during QS. The nasogastric tube for feeding can be blocked during specific sleep states to avoid temperature drops and instability of the preterm infant. Alarms can be automatically turned off and also not be stored in the data warehouse during caretaking procedures as alarms are likely provoked due to the procedure. False alarm statistics due to caretaking can be reduced this way.

An alternative question is whether automated sleep monitoring can comfort parents in the first days after release from the hospital to home. In several discussions with parents and caretakers, most of the parents we encountered felt insecure and unfit to take over the full care for their preterm infant child during the first weeks at home. They reported that the insecurity was mainly due to an often harsh cut from a fully monitored environment with trained staff to an unmonitored home environment. Parents have an increased risk of postpartum depression, posttraumatic stress disorder (PTSD), and anxiety disorder after discharge from the NICU [5], [290], [291]. New concepts, regarding family inclusion are launched like the mentioned M-NICU [6] with the mother having her bed next to the incubator and being an active caregiver, and NIPU [5] which similarly centers the care not mainly around the preterm infant but around the family as a whole with additional mental health support and post-discharge support. An important part of the new concepts is the inter-professional collaboration with families involving them in the care plan and mentor them to establish knowledgeable caretakers [5], [292]. This mentoring should achieve parental decision-making, taking responsibility and achieving a realistic and informed view of the preterm infant's health condition. Mentored parents might have a lower risk of

PTSD by reducing the feeling of helplessness [293]. Part of these novel approaches could be an automated sleep monitoring system. As sleep is a major part of the preterm infant's early part of life, mentoring parents on their infant's sleep patterns and sleep cycles during their hospital stay, reinforced by an automated monitoring system, would help them understand their child further on a core level. This understanding can be taken home as one of the first parenting skills, strengthening the parents on their journey and easing the hospital to home transition.

As mentioned at the beginning of this chapter, we believe that an automated system can be implemented with camera solutions and latest ANNs, which then can be transferred into the home environment using existing camera/video solutions. If needed, video monitoring can extend hospital monitoring on a lower level into the home environment until monitoring is deemed unnecessary. From a clinical perspective, home monitoring may indicate ill development at home and potentially reduce critical re-admissions to the hospital by early recognition.

In discussing this hospital to home process, one question was prominent. Would any monitoring solution not hinder the natural adaption process of the parents to their child? Would the constant monitoring not delay them from relying on their parental instincts and the knowledge they gathered during the hospital stay? Would such a monitoring system not be more of a burden than an aid? We cannot answer this question in its entity as it might vary from parent to parent based on different personalities, cultural, and/or knowledge backgrounds. However, we believe that with the mentioned novel concepts, parents will become experts themselves using the monitoring rather as insurance than being fully depending on them. Informed and trained parents will not be hindered to adapt to their infant, as the hospital to home transition will not be an abrupt change in knowledge by missing experts, but only a change in location.

In general, the questions about home monitoring and hospital to home experience may yield different answers for different cultures with varying beliefs and traditions. This is meant with the clear distinction between stereotypeization and generalization.

At this point, we will not dive into depth about cultural differences on the view of home monitoring, but instead raise another critical question for the last part of this thesis. Do we need such elaborate monitoring solutions when the majority of preterm infants are born in low-income countries where that research has probably limited effects due to cost barriers?

Moreover, despite our effort to enable a low-income appropriate solution by focusing on an ECG solution, the answer would be yes. History has shown that inventions, innovations, and processes are trickling down from high-profit solutions to general, widespread use. An example close to the subject of this thesis would be the use of incubators for preterm babies and the formation of NICUs. The first incubator used on preterm infants was set up in 1903 by Martin Arthur Couney, and they were introduced in Europe at around 1933 [294]. From 1950 onwards, hospitals started to adopt special units for preterm infants, which evolved into today's highly specialized NICUs. After NICUs were established in the high-income countries, the knowledge was exported to developing and low-income countries. To date, NICUs are being deployed in rural areas all over the world, often by western health organizations. Whether the deployment of NICUs by external organizations is the best way to establish long-lasting positive effects in low-income countries is another interesting question, which would extend this thesis disproportional and will be left undiscussed. Another example of technology transfer from high- to low-income countries is the smartphone. Smartphones were limited in early 2000 to a specific user group in high-income countries. To date, low-income countries, such as in Africa, have a very high mobile penetration rate, which is also changing the health care sector and enables the use of mobile telecommunication to raise health awareness and provide telehealth to reduce neonatal transports in rural areas [295]. Smartphone sensor and computation power can be used for health care in rural areas such as the smartphone stethoscope Steth IO [296], or as ultrasound monitors in combination with a smartphone-compatible ultrasound head [297]. Finally, countless health apps can be used worldwide, benefitting personal health. With the rise of deep learning and

specialized hardware acceleration for AI, such as the Google tensor processing unit [298], with the next smartphone generation, we will see an increase in smartphone apps implementing AI to analyze and monitor patients.

Despite these practical applications for low-income regions, research should not have the limitations to be bound to direct use cases or benefits. This statement is also true for the application-oriented field of engineering. The path of research results might not always be foreseeable. Whenever we encounter a research topic or project that is based on fastidious conditions and with a limited focus, we should see it as a sign of peace. There might not be another time where "useless" projects can be realized. In blessed times and conditions, we have the responsibility to tackle those questions despite a direct or even indirect benefit.

Unnecessary research is a token of peace!

References

[1]   World Health Organization, "Preterm birth," 2018.

[2]   American Academy Of Pediatrics, "Age terminology during the perinatal period," *Pediatrics*, vol. 114, no. 5, pp. 1362–4, Nov. 2004.

[3]   World Health Organization, "ICD-11 for mortality and morbidity statistics," 2018.

[4]   J. P. Baker, "The incubator and the medical discovery of the premature infant," *J. Perinatol.*, vol. 20, no. 5, pp. 321–328, Jul. 2000.

[5]   S. L. Hall *et al.*, "The neonatal intensive parenting unit: An introduction," *J. Perinatol.*, vol. 37, no. 12, pp. 1259–1264, Dec. 2017.

[6]   H. Chellani, P. Mittal, and S. Arya, "Mother-Neonatal Intensive Care Unit (M-NICU): A novel concept in newborn care," *Indian Pediatr.*, vol. 55, no. 12, pp. 1035–1036, Dec. 2018.

[7]   D. Kim and A. Saada, "The social determinants of infant mortality and birth outcomes in western developed nations: A cross-country systematic review," *Int. J. Environ. Res. Public Health*, vol. 10, no. 6, pp. 2296–2335, Jun. 2013.

[8]   H. Xu *et al.*, "Time trends and risk factor associated with premature birth and infants deaths due to prematurity in Hubei Province, China from 2001 to 2012," *BMC Pregnancy Childbirth*, vol. 15, no. 1, p. 329, Dec. 2015.

[9]   J. E. Lawn *et al.*, "Born too soon: Care for the preterm baby," *Reprod. Health*, vol. 10, no. S1, p. S5, Nov. 2013.

[10]   E. C. Howson, M. Kinney, J. Lawn, and World Health Organization, "Born too soon. The global action report on preterm birth," 2012.

[11]   H. Blencowe *et al.*, "Born too soon: The global epidemiology of 15 million preterm births," *Reprod. Health*, vol. 10, no. Suppl 1, p. S2, 2013.

[12]   J. A. Martin *et al.*, "Births: final data for 2009," *Natl. Vital Stat. Rep.*, vol. 60, no. 1, pp. 1–70, Nov. 2011.

[13]   H.-X. Qi, P.-Y. Chen, and S.-L. Yuan, "[Diseases in the Neonatal period among preterm infants: An epidemiological investigation].," *Zhongguo Dang Dai Er Ke Za Zhi*, vol. 14, no. 1, pp. 11–4, Jan. 2012.

[14]   A. Greenough, E. Limb, L. Marston, N. Marlow, S. Calvert, and J. Peacock, "Risk factors for respiratory morbidity in infancy after very premature birth," *Arch. Dis. Child. Fetal Neonatal Ed.*, vol. 90, no. 4, pp. F320-3, Jul. 2005.

[15]   E. J. L. E. Vrijlandt, H. M. Boezen, J. Gerritsen, E. F. Stremmelaar, and E. J. Duiverman, "Respiratory health in prematurely born preschool children with and without bronchopulmonary dysplasia," *J. Pediatr.*, vol. 150, no. 3, pp. 256–61, Mar. 2007.

[16]   I. Pramana *et al.*, "Respiratory symptoms in preterm infants: burden of disease in the first year of life," *Eur. J. Med. Res.*, vol. 16, no. 5, p. 223, 2011.

[17]   B. Resch, A. Pasnocht, W. Gusenleitner, and W. Müller, "Rehospitalisations for respiratory disease and respiratory syncytial virus infection in preterm infants of 29-36 weeks gestational age," *J. Infect.*, vol. 50, no. 5, pp. 397–403, Jun. 2005.

[18]   B. Blondel, P. Truffert, A. Lamarche-Vadel, M. Dehan, and B. Larroque, "Use of medical services by very preterm children during the first year of life in the Epipage cohort," *Arch. Pediatr.*, vol. 10, no. 11, pp. 960–8, Nov. 2003.

[19]   H. S. Alan Siegel, *Essential Neuroscience*. Baltimore, MD: Wolters Kluwer Lippincott Williams &Wilkins, 2010.

[20]   W. McGuire, L. Clerihew, and P. W. Fowlie, "Infection in the preterm infant," *BMJ*, vol. 329, no. 7477, pp. 1277–1280, Nov. 2004.

[21]   R. Berger and S. Söder, "Neuroprotection in preterm infants," *Biomed Res. Int.*, vol. 2015, pp. 1–14, 2015.

[22]   R. Joshi, C. van Pul, L. Atallah, L. Feijs, S. Van Huffel, and P. Andriessen, "Pattern discovery in critical alarms originating from neonates under intensive care," *Physiol. Meas.*, vol. 37, no. 4, pp. 564–579, Apr. 2016.

[23]   R. Joshi *et al.*, "A strategy to reduce critical cardiorespiratory alarms due to intermittent enteral feeding of preterm neonates in intensive care," *Interact. J. Med. Res.*, vol. 6, no. 2, p. e20, Oct. 2017.

[24]   M. Hack, J. D. Horbar, M. H. Malloy, J. E. Tyson, E. Wright, and L. Wright, "Very low birth weight outcomes of the National Institute of Child Health and Human Development Neonatal Network.," *Pediatrics*, vol. 87, no. 5, pp. 587–97, May 1991.

[25]   E. C. Eichenwald and A. R. Stark, "Management and outcomes of very low birth weight.," *N. Engl. J. Med.*, vol. 358, no. 16, pp. 1700–11, Apr. 2008.

[26]   B. J. Stoll *et al.*, "Neonatal outcomes of extremely preterm infants from the NICHD Neonatal Research Network.," *Pediatrics*, vol. 126, no. 3, pp. 443–56, Sep. 2010.

[27]   H. C. Glass, A. T. Costarino, S. A. Stayer, C. M. Brett, F. Cladis, and P. J. Davis, "Outcomes for extremely premature infants," *Anesth. Analg.*, vol. 120, no. 6, pp. 1337–1351, Jun. 2015.

[28]   V. B. de Graaf-Peters and M. Hadders-Algra, "Ontogeny of the human central nervous system: What is happening when?," *Early Hum. Dev.*, vol. 82, no. 4, pp. 257–266, Apr. 2006.

[29] J. J. Volpe *et al.*, *Neurology of the newborn*, 5th ed. Philadelphia: Elsevier, 2008.

[30] G. Moscoso, "Early embryonic development of the brain," in *Fetal and Neonatal Neurology and Neurosurgery*, 4th Editio., C. F. Levene M, Ed. Churchill Livingstone Elsevier, 2008, pp. 12–28.

[31] P. S. Hüppi *et al.*, "Quantitative magnetic resonance imaging of brain development in premature and mature newborns," *Ann. Neurol.*, vol. 43, no. 2, pp. 224–235, Feb. 1998.

[32] J. L. Y. Cheong *et al.*, "Brain volumes at term-equivalent age are associated with 2-Year neurodevelopment in moderate and late preterm children," *J. Pediatr.*, vol. 174, pp. 91-97.e1, Jul. 2016.

[33] G. Poulsen *et al.*, "Gestational age and cognitive ability in early childhood: a population-based cohort study," *Paediatr. Perinat. Epidemiol.*, vol. 27, no. 4, pp. 371–379, Jul. 2013.

[34] M. L. McKee *et al.*, "Bayley scales of infant development: third edition," in *Encyclopedia of Child Behavior and Development*, Boston, MA: Springer US, 2011, pp. 215–215.

[35] S. Johnson and N. Marlow, "Preterm birth and childhood psychiatric disorders," *Pediatr. Res.*, vol. 69, no. 5 Pt 2, pp. 11R–8R, May 2011.

[36] S. M. Ludington-Hoe *et al.*, "Neurophysiologic assessment of neonatal sleep organization: preliminary results of a randomized, controlled trial of skin contact with preterm infants," *Pediatrics*, vol. 117, no. 5, pp. e909–e923, May 2006.

[37] A. H. Parmelee, W. H. Wenner, Y. Akiyama, M. Schultz, and E. Stern, "Sleep states in premature infants," *Dev. Med. Child Neurol.*, vol. 9, no. 1, pp. 70–77, Feb. 1967.

[38] T. Anders, R. Emde, and A. H. Parmelee, *A manual of standardized terminology, techniques, and criteria for scoring of states of sleep and wakefulness in newborn infants.* UCLA Brain Information Service /BRI Publications Office, NINDS Neurological Information Network, 1971.

[39] A. W. De Weerd and R. A. S. Van den Bossche, "The development of sleep during the first months of life," *Sleep Med. Rev.*, vol. 7, no. 2, pp. 179–191, Apr. 2003.

[40] L. Curzi-Dascalova, P. Peirano, and F. Morel-Kahn, "Development of sleep states in normal premature and full-term newborns," *Dev. Psychobiol.*, vol. 21, no. 5, pp. 431–44, Jul. 1988.

[41] L. Curzi-Dascalova *et al.*, "Sleep state organization in premature infants of less than 35 weeks' gestational age," *Pediatr. Res.*, vol. 34, no. 5, pp. 624–628, Nov. 1993.

[42] D. Holditch-Davis and E. B. Thoman, "Behavioral states of premature infants: implications for neural and behavioral development," *Dev. Psychobiol.*, vol. 20, no. 1, pp. 25–38, Jan. 1987.

[43] D. Holditch-Davis, M. S. Scher, T. Schwartz, and D. Hudson-Barr, "Sleeping and waking state development in preterm infants," *Early Hum. Dev.*, vol. 80, no. 1, pp. 43–64, Oct. 2004.

[44] E. W. Ingersoll and E. B. Thoman, "Sleep/wake states of preterm infants: stability, developmental change, diurnal variation, and relation with caregiving activity," *Child Dev.*, vol. 70, no. 1, pp. 1–10, 1999.

[45] Mirmiran and Van Someren E, "Symposium: Normal and abnormal REM sleep regulation: the importance of REM sleep for brain maturation," *J. Sleep Res.*, vol. 2, no. 4, pp. 188–192, Dec. 1993.

[46] C. L. Booth, H. L. Leonard, and E. B. Thoman, "Sleep states and behavior

patterns in preterm and fullterm infants," *Neuropediatrics*, vol. 11, no. 4, pp. 354–64, Nov. 1980.

[47]   J. Kohyama and Y. Iwakawa, "Developmental changes in phasic sleep parameters as reflections of the brain-stem maturation: polysomnographical examinations of infants, including premature neonates," *Electroencephalogr. Clin. Neurophysiol.*, vol. 76, no. 4, pp. 325–30, Oct. 1990.

[48]   M. S. Scher, D. A. Steppe, D. L. Banks, R. D. Guthrie, and R. J. Sclabassi, "Maturational trends of EEG-sleep measures in the healthy preterm neonate," *Pediatr. Neurol.*, vol. 12, no. 4, pp. 314–22, May 1995.

[49]   S. Gertner, C. W. Greenbaum, A. Sadeh, Z. Dolfin, L. Sirota, and Y. Ben-Nun, "Sleep–wake patterns in preterm infants and 6 month's home environment: implications for early cognitive development," *Early Hum. Dev.*, vol. 68, no. 2, pp. 93–102, Jul. 2002.

[50]   L. Curzi-Dascalova, "Between-sleep states transitions in premature babies," *J. Sleep Res.*, vol. 10, no. 2, pp. 153–158, Jun. 2001.

[51]   M. Grigg-Damberger *et al.*, "The visual scoring of sleep and arousal in infants and children," *J. Clin. Sleep Med.*, vol. 3, no. 2, pp. 201–40, Mar. 2007.

[52]   P. Peirano, C. Algarín, and R. Uauy, "Sleep-wake states and their regulatory mechanisms throughout early human development," *J. Pediatr.*, vol. 143, no. 4 Suppl, pp. S70-9, Oct. 2003.

[53]   A. Kahn, B. Dan, J. Groswasser, P. Franco, and M. Sottiaux, "Normal sleep architecture in infants and children.," *J. Clin. Neurophysiol.*, vol. 13, no. 3, pp. 184–97, May 1996.

[54]   K. Palmu, T. Kirjavainen, S. Stjerna, T. Salokivi, and S. Vanhatalo, "Sleep wake cycling in early preterm infants: comparison of polysomnographic recordings with a novel EEG-based index.," *Clin. Neurophysiol.*, vol. 124, no.

9, pp. 1807–14, Sep. 2013.

[55]   D. Jouvet-Mounier, L. Astic, and D. Lacote, "Ontogenesis of the states of sleep in rat, cat, and guinea pig during the first postnatal month," *Dev. Psychobiol.*, vol. 2, no. 4, pp. 216–39, Jan. 1970.

[56]   G. A. Marks, J. P. Shaffery, A. Oksenberg, S. G. Speciale, and H. P. Roffwarg, "A functional role for REM sleep in brain maturation," *Behav. Brain Res.*, vol. 69, no. 1–2, pp. 1–11, 1995.

[57]   J. Lopez, H. P. Roffwarg, A. Dreher, G. Bissette, B. Karolewicz, and J. P. Shaffery, "Rapid eye movement sleep deprivation decreases long term potentiation stability and affects some glutamatergic signaling proteins during hippocampal development," *Neuroscience*, vol. 153, no. 1, pp. 44–53, 2008.

[58]   J. P. Shaffery, J. S. Allard, K. F. Manaye, and H. P. Roffwarg, "Selective rapid eye movement sleep deprivation affects cell size and number in kitten locus coeruleus," *Front. Neurol.*, vol. 3, p. 69, Jan. 2012.

[59]   M. Mirmiran, N. E. Van De Poll, M. A. Corner, H. G. Van Oyen, and H. L. Bour, "Suppression of active sleep by chronic treatment with chlorimipramine during early postnatal development: effects upon adult sleep and behavior in the rat," *Brain Res.*, vol. 204, no. 1, pp. 129–146, Jan. 1981.

[60]   M. Mirmiran, J. Scholtens, N. E. van de Poll, H. B. Uylings, J. van der Gugten, and G. J. Boer, "Effects of experimental suppression of active (REM) sleep during early development upon adult brain and behavior in the rat," *Brain Res.*, vol. 283, no. 2–3, pp. 277–86, Apr. 1983.

[61]   J. P. Shaffery, C. M. Sinton, G. Bissette, H. P. Roffwarg, and G. A. Marks, "Rapid eye movement sleep deprivation modifies expression of long-term

potentiation in visual cortex of immature rats," *Neuroscience*, vol. 110, no. 3, pp. 431–443, Mar. 2002.

[62]   R. Ursin, "Serotonin and sleep," *Sleep Med. Rev.*, vol. 6, no. 1, pp. 55–67, Feb. 2002.

[63]   S. W. Foreman, K. A. Thomas, and S. T. Blackburn, "Individual and gender differences matter in preterm infant state development," *J. Obstet. Gynecol. Neonatal Nurs.*, vol. 37, no. 6, pp. 657–665, Nov. 2008.

[64]   J.-M. Jia *et al.*, "Brain-derived neurotrophic factor-tropomyosin-related kinase b signaling contributes to activity-dependent changes in synaptic proteins," *J. Biol. Chem.*, vol. 283, no. 30, pp. 21242–21250, Jul. 2008.

[65]   E. B. Thoman, "Sleeping and waking states in infants: a functional perspective," *Neurosci. Biobehav. Rev.*, vol. 14, no. 1, pp. 93–107, Jan. 1990.

[66]   M. S. Scher, "Ontogeny of EEG-sleep from neonatal through infancy periods," *Sleep Med.*, vol. 9, no. 6, pp. 615–36, Aug. 2008.

[67]   L. K. McCorry, "Physiology of the autonomic nervous system.," *Am. J. Pharm. Educ.*, vol. 71, no. 4, p. 78, Sep. 2007.

[68]   X. Long, P. Fonseca, R. M. Aarts, R. Haakma, J. Rolink, and S. Leonhardt, "Detection of nocturnal slow wave sleep based on cardiorespiratory activity in healthy adults," *IEEE J. Biomed. Heal. Informatics*, vol. 21, no. 1, pp. 123–133, Jan. 2017.

[69]   M. J. Hayes *et al.*, "Apneic preterms and methylxanthines: arousal deficits, sleep fragmentation and suppressed spontaneous movements," *J. Perinatol.*, vol. 27, no. 12, pp. 782–789, Dec. 2007.

[70]   C. P. Clark, "Sleep medicine: essentials and review," *J. Clin. Psychiatry*, vol. 71, no. 06, pp. 810–811, Jun. 2010.

[71]   L. Thorngate, S. W. Foreman, and K. A. Thomas, "Quantification of neonatal amplitude-integrated EEG patterns," *Early Hum. Dev.*, vol. 89, no. 12, pp. 931–937, 2013.

[72]   J. W. Britton, L. C. Frey, and J. L. Hopp, "Electroencephalography (EEG): an introductory text and atlas of normal and abnormal findings in adults, children, and infants," in *The Developmental EEG: Premature, Neonatal, Infant, and Children. Available*, E. St. Louis and L. Frey, Eds. Chicago: American Epilepsy Society, 2016, pp. 20–41.

[73]   R. J. Ellingson, "Development of sleep spindle bursts during the first year of life." *Sleep*, vol. 5, no. 1, pp. 39–46, 1982.

[74]   K. Whitehead, R. Pressler, and L. Fabrizi, "Characteristics and clinical significance of delta brushes in the EEG of premature infants," *Clin. Neurophysiol. Pract.*, vol. 2, pp. 12–18, 2017.

[75]   B. Baykan *et al.*, "Delta brush pattern is not unique to NMDAR encephalitis: evaluation of two independent long-term EEG cohorts," *Clin. EEG Neurosci.*, vol. 49, no. 4, pp. 278–284, Jul. 2018.

[76]   S. Diekelmann and J. Born, "The memory function of sleep," *Nat. Rev. Neurosci.*, vol. 11, no. 2, pp. 114–126, Feb. 2010.

[77]   M. M. Morawska, F. Buchele, C. G. Moreira, L. L. Imbach, D. Noain, and C. R. Baumann, "Sleep modulation alleviates axonal damage and cognitive decline after rodent traumatic brain injury," *J. Neurosci.*, vol. 36, no. 12, pp. 3422–3429, Mar. 2016.

[78]   E. E. Benarroch, "The central autonomic network: functional organization, dysfunction, and perspective," *Mayo Clin. Proc.*, vol. 68, no. 10, pp. 988–1001, Oct. 1993.

[79]   F. Riganello, "Responsiveness and the autonomic control--CNS two-way

interaction in disorders of consciousness," in *Brain Function and Responsiveness in Disorders of Consciousness*, M. M. Monti and W. G. Sannita, Eds. Cham: Springer International Publishing, 2016, pp. 145–155.

[80]   F. Riganello, V. Prada, A. Soddu, C. di Perri, and W. G. Sannita, "Circadian rhythms and measures of CNS/autonomic interaction," *Int. J. Environ. Res. Public Health*, vol. 16, no. 13, p. 2336, Jul. 2019.

[81]   M. Mirmiran, R. B. Baldwin, and R. L. Ariagno, "Circadian and sleep development in preterm infants occurs independently from the influences of environmental lighting," *Pediatr. Res.*, vol. 53, no. 6, pp. 933–8, Jun. 2003.

[82]   C. C. W. Hsia, D. M. Hyde, and E. R. Weibel, "Lung structure and the intrinsic challenges of gas exchange," in *Comprehensive Physiology*, Hoboken, NJ, USA: John Wiley & Sons, Inc., 2016, pp. 827–895.

[83]   R. Gordan, J. K. Gwathmey, and L.-H. Xie, "Autonomic and endocrine control of cardiovascular function," *World J. Cardiol.*, vol. 7, no. 4, pp. 204–14, Apr. 2015.

[84]   M. A. Russo, D. M. Santarelli, and D. O'Rourke, "The physiological effects of slow breathing in the healthy human," *Breathe*, vol. 13, no. 4, pp. 298–309, Dec. 2017.

[85]   S. Vandeput, G. Naulaers, H. Daniels, and S. Van Huffel, "Heart rate variability during REM and non-REM sleep in preterm neonates with and without abnormal cardiorespiratory events," *Early Hum. Dev.*, vol. 85, no. 10, pp. 665–671, Oct. 2009.

[86]   K. Verhoeff and J. R. Mitchell, "Cardiopulmonary physiology: why the heart and lungs are inextricably linked," *Adv. Physiol. Educ.*, vol. 41, no. 3, pp. 348–353, Sep. 2017.

[87]   N. Singh, K. J. Moneghetti, J. W. Christle, D. Hadley, D. Plews, and V.

190

Froelicher, "Heart rate variability: an old metric with new meaning in the era of using health technologies for health and exercise training guidance. Part one: physiology and methods," *Arrhythmia Electrophysiol. Rev.*, vol. 7, no. 3, pp. 193–198, Aug. 2018.

[88] O. Hamzaoui, X. Monnet, and J.-L. Teboul, "Pulsus paradoxus," *Eur. Respir. J.*, vol. 42, no. 6, pp. 1696–1705, Dec. 2013.

[89] H. Fukuta and W. C. Little, "The cardiac cycle and the physiologic basis of left ventricular contraction, ejection, relaxation, and filling," *Heart Fail. Clin.*, vol. 4, no. 1, pp. 1–11, Jan. 2008.

[90] W. L. Kenney, A. E. Stanhewicz, R. S. Bruning, and L. M. Alexander, "Blood pressure regulation III: what happens when one system must serve two masters: temperature and pressure regulation?," *Eur. J. Appl. Physiol.*, vol. 114, no. 3, pp. 467–479, Mar. 2014.

[91] M. Di Rienzo, G. Parati, A. Radaelli, and P. Castiglioni, "Baroreflex contribution to blood pressure and heart rate oscillations: time scales, time-variant characteristics and nonlinearities," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 367, no. 1892, pp. 1301–1318, Apr. 2009.

[92] V. Papaioannou, I. Pneumatikos, and N. Maglaveras, "Association of heart rate variability and inflammatory response in patients with cardiovascular diseases: current strengths and limitations," *Front. Physiol.*, vol. 4, p. 174, 2013.

[93] D. S. Goldstein, O. Bentho, M.-Y. Park, and Y. Sharabi, "Low-frequency power of heart rate variability is not a measure of cardiac sympathetic tone but may be a measure of modulation of cardiac autonomic outflows by baroreflexes," *Exp. Physiol.*, vol. 96, no. 12, pp. 1255–1261, Dec. 2011.

[94] J. E. Mazursky, C. L. Birkett, K. A. Bedell, S. A. Ben-Haim, and J. L. Segar,

"Development of baroreflex influences on heart rate variability in preterm infants," *Early Hum. Dev.*, vol. 53, no. 1, pp. 37–52, Nov. 1998.

[95]   J. F. Thayer, R. Nabors-Oberg, and J. J. Sollers, "Thermoregulation and cardiac variability: a time-frequency analysis," *Biomed. Sci. Instrum.*, vol. 34, pp. 252–6, 1997.

[96]   E. Nattie and A. Li, "Central Chemoreceptors: Locations and Functions," in *Comprehensive Physiology*, Hoboken, NJ, USA: John Wiley & Sons, Inc., 2012.

[97]   P. Indic, E. Bloch-Salisbury, F. Bednarek, E. N. Brown, D. Paydarfar, and R. Barbieri, "Assessment of cardio-respiratory interactions in preterm infants by bivariate autoregressive modeling and surrogate data analysis," *Early Hum. Dev.*, vol. 87, no. 7, pp. 477–487, Jul. 2011.

[98]   T. Kara, K. Narkiewicz, and V. K. Somers, "Chemoreflexes - physiology and clinical implications," *Acta Physiol. Scand.*, vol. 177, no. 3, pp. 377–384, Mar. 2003.

[99]   V. E. Papaioannou, N. Maglaveras, I. Houvarda, E. Antoniadou, and G. Vretzakis, "Investigation of altered heart rate variability, nonlinear properties of heart rate signals, and organ dysfunction longitudinally over time in intensive care unit patients," *J. Crit. Care*, vol. 21, no. 1, pp. 95–103, Mar. 2006.

[100] T. A. Day and R. J. A. Wilson, "A negative interaction between brainstem and peripheral respiratory chemoreceptors modulates peripheral chemoreflex magnitude," *J. Physiol.*, vol. 587, no. 4, pp. 883–896, Feb. 2009.

[101] N. Koolen *et al.*, "Automated classification of neonatal sleep states using EEG," *Clin. Neurophysiol.*, vol. 128, no. 6, pp. 1100–1108, Jun. 2017.

[102] A. Dereymaeker *et al.*, "An automated quiet sleep detection approach in preterm infants as a gateway to assess brain maturation," *Int. J. Neural Syst.*,

192

vol. 27, no. 06, p. 1750023, Sep. 2017.

[103] 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 Behav. Dev.*, vol. 18, no. 3, pp. 329–337, Jul. 1995.

[104] C. Acebo, A. Sadeh, R. Seifer, O. Tzischinsky, A. Hafer, and M. a Carskadon, "Sleep/Wake patterns derived from activity monitoring and maternal report for healthy 1- to 5-year-old children," *Sleep*, vol. 28, no. 12, pp. 1568–1577, Dec. 2005.

[105] G. G. Haddad, H. J. Jeng, T. L. Lai, and R. B. Mellins, "Determination of sleep state in infants using respiratory variability," *Pediatr. Res.*, vol. 21, no. 6, pp. 556–562, Jun. 1987.

[106] H. C. McLane *et al.*, "Availability, accessibility, and affordability of neurodiagnostic tests in 37 countries," *Neurology*, vol. 85, no. 18, pp. 1614–1622, Nov. 2015.

[107] L. Atallah *et al.*, "Unobtrusive ECG monitoring in the NICU using a capacitive sensing array," *Physiol. Meas.*, vol. 35, no. 5, pp. 895–913, May 2014.

[108] A. Gruetzmann, S. Hansen, and J. Müller, "Novel dry electrodes for ECG monitoring," *Physiol. Meas.*, vol. 28, no. 11, pp. 1375–90, Nov. 2007.

[109] K. Jones, "Alarm fatigue a top patient safety hazard," *Can. Med. Assoc. J.*, vol. 186, no. 3, pp. 178–178, Feb. 2014.

[110] M. S. Scher, J. Turnbull, K. Loparo, and M. W. Johnson, "Automated state analyses: proposed applications to neonatal neurointensive care," *J. Clin. Neurophysiol.*, vol. 22, no. 4, pp. 256–70, Aug. 2005.

[111] P. I. Terrill, S. J. Wilson, S. Suresh, D. M. Cooper, and C. Dakin,

"Application of recurrence quantification analysis to automatically estimate infant sleep states using a single channel of respiratory data," *Med. Biol. Eng. Comput.*, vol. 50, no. 8, pp. 851–865, Aug. 2012.

[112] V. Gerla, K. Paul, L. Lhotska, and V. Krajca, "Multivariate analysis of full-term neonatal polysomnographic data," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 1, pp. 104–110, Jan. 2009.

[113] T. F. Anders, M. A. Keener, and H. Kraemer, "Sleep-wake state organization, neonatal assessment and development in premature infants during the first year of life. II," *Sleep*, vol. 8, no. 3, pp. 193–206, Jan. 1985.

[114] R. Michaelis, A. H. Parmelee, E. Stern, and A. Haber, "Activity states in premature and term infants," *Dev. Psychobiol.*, vol. 6, no. 3, pp. 209–15, May 1973.

[115] E. Stéphan-Blanchard, K. Chardon, A. Léké, S. Delanaud, V. Bach, and F. Telliez, "Heart rate variability in sleeping preterm neonates exposed to cool and warm thermal conditions," *PLoS One*, vol. 8, no. 7, p. e68211, Jan. 2013.

[116] E. Stern, A. H. Parmelee, and M. A. Harris, "Sleep state periodicity in prematures and young infants," *Dev. Psychobiol.*, vol. 6, no. 4, pp. 357–365, Jul. 1973.

[117] H. J. Niemarkt, P. Andriessen, J. Pasman, J. S. Vles, L. J. Zimmermann, and S. B. Oetomo, "Analyzing EEG maturation in preterm infants: the value of a quantitative approach," *J. Neonatal. Perinatal. Med.*, vol. 1, no. 3, pp. 131–144, 2008.

[118] H. J. Niemarkt, P. Andriessen, C. H. L. Peters, J. W. Pasman, L. J. Zimmermann, and S. B. Oetomo, "Quantitative analysis of maturational changes in EEG background activity in very preterm infants with a normal neurodevelopment at 1 year of age," *Early Hum. Dev.*, vol. 86, no. 4, pp. 219–

24, Apr. 2010.

[119]   A. H. Parmelee and E. Stern, "Development of states in infants," in *Sleep and the maturing nervous system*, C. Clemente, D. Purpura, and F. Mayer, Eds. London: Academic Press, 1972, pp. 199–228.

[120]   S. Sheldon, *Sleep: a comprehensive handbook*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2005.

[121]   R. B. Berry *et al.*, "Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events. deliberations of the sleep apnea definitions task force of the american academy of sleep medicine," *J. Clin. Sleep Med.*, vol. 8, no. 5, pp. 597–619, Oct. 2012.

[122]   P. Andriessen *et al.*, "Cardiovascular fluctuations and transfer function analysis in stable preterm infants," *Pediatr. Res.*, vol. 53, no. 1, pp. 89–97, Jan. 2003.

[123]   M. Eiselt, L. Curzi-Dascalova, J. Clairambault, F. Kauffmann, C. Médigue, and P. Peirano, "Heart-rate variability in low-risk prematurely born infants reaching normal term: a comparison with full-term newborns," *Early Hum. Dev.*, vol. 32, no. 2–3, pp. 183–195, Mar. 1993.

[124]   K. Pfurtscheller, G. R. Müller-Putz, B. Urlesberger, W. Müller, and G. Pfurtscheller, "Relationship between slow-wave EEG bursts and heart rate changes in preterm infants," *Neurosci. Lett.*, vol. 385, no. 2, pp. 126–30, Sep. 2005.

[125]   P. Andriessen, S. B. Oetomo, C. Peters, B. Vermeulen, P. F. F. Wijn, and C. E. Blanco, "Baroreceptor reflex sensitivity in human neonates: the effect of postmenstrual age," *J. Physiol.*, vol. 568, no. 1, pp. 333–341, Oct. 2005.

[126]   A. Kotini, P. Anninos, A. Adamopoulos, "Linear analysis of fetal

magnetocardiogram recordings in normal pregnancies at various gestational ages," *J. Obstet. Gynaecol. (Lahore).*, vol. 21, no. 2, pp. 154–157, Jan. 2001.

[127]   C. Krueger, J. H. van Oostrom, and J. Shuster, "A longitudinal description of heart rate variability in 28—34-week-old preterm infants," *Biol. Res. Nurs.*, vol. 11, no. 3, pp. 261–268, Jan. 2010.

[128]   A. Z. Khattak, N. S. Padhye, A. L. Williams, R. E. Lasky, F. R. Moya, and M. T. Verklan, "Longitudinal assessment of heart rate variability in very low birth weight infants during their NICU stay," *Early Hum. Dev.*, vol. 83, no. 6, pp. 361–366, Jun. 2007.

[129]   C. E. Hunt, "Ontogeny of autonomic regulation in late preterm infants born at 34-37 weeks postmenstrual age," *Semin. Perinatol.*, vol. 30, no. 2, pp. 73–76, Apr. 2006.

[130]   E. Longin, T. Gerstner, T. Schaible, T. Lenz, and S. König, "Maturation of the autonomic nervous system: differences in heart rate variability in premature vs. term infants," *J. Perinat. Med.*, vol. 34, no. 4, pp. 303–8, Jan. 2006.

[131]   B. Aysin and E. Aysin, "Effect of respiration in heart rate variability (HRV) analysis," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, 2006, vol. 1, pp. 1776–1779.

[132]   M. A. Lopez-Gordo, D. Sanchez-Morillo, and F. Pelayo Valle, "Dry EEG electrodes," *Sensors*, vol. 14, no. 7, pp. 12847–70, Jan. 2014.

[133]   J. Lee, J. Heo, W. Lee, Y. Lim, Y. Kim, and K. Park, "Flexible capacitive electrodes for minimizing motion artifacts in ambulatory electrocardiograms," *Sensors*, vol. 14, no. 8, pp. 14732–14743, Aug. 2014.

[134]   G. Ruffini *et al.*, "First human trials of a dry electrophysiology sensor using a carbon nanotube array interface," *Sensors Actuators A Phys.*, vol. 144, no. 2,

pp. 275–279, Jun. 2008.

[135] M. Matteucci *et al.*, "Micropatterned dry electrodes for brain– computer interface," *Microelectron. Eng.*, vol. 84, no. 5–8, pp. 1737–1740, May 2007.

[136] C. O'Mahony, F. Pini, A. Blake, C. Webster, J. O'Brien, and K. G. McCarthy, "Microneedle-based electrodes with integrated through-silicon via for biopotential recording," *Sensors Actuators A Phys.*, vol. 186, pp. 130–136, Oct. 2012.

[137] R. Lloyd, R. Goulding, P. Filan, and G. Boylan, "Overcoming the practical challenges of electroencephalography for very preterm infants in the neonatal intensive care unit," *Acta Paediatr.*, vol. 104, no. 2, pp. 152–157, Feb. 2015.

[138] R. A. Shellhaas *et al.*, "The American Clinical Neurophysiology Society's guideline on continuous electroencephalography monitoring in neonates," *J. Clin. Neurophysiol.*, vol. 28, no. 6, pp. 611–617, Dec. 2011.

[139] E. Nemati, M. Deen, and T. Mondal, "A wireless wearable ECG sensor for long-term applications," *IEEE Commun. Mag.*, vol. 50, no. 1, pp. 36–43, Jan. 2012.

[140] Y. M. Chi, T. Jung, and G. Cauwenberghs, "Dry-contact and noncontact biopotential electrodes: methodological review," *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 106–19, Jan. 2010.

[141] T. J. Sullivan, S. R. Deiss, and G. Cauwenberghs, "A low-noise, non-contact EEG/ECG sensor," in *2007 IEEE Biomedical Circuits and Systems Conference*, 2007, pp. 154–157.

[142] Y. M. Chi and G. Cauwenberghs, "Wireless non-contact EEG/ECG electrodes for body sensor networks," in *2010 International Conference on Body Sensor Networks*, 2010, pp. 297–301.

[143] A. Serteyn, R. Vullings, M. Meftah, and J. Bergmans, "Using an injection signal to reduce motion artifacts in capacitive ECG measurements," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, vol. 2013, pp. 4795–4798.

[144] R. M. Aarts and S. B. Oetomo, "Heart rate measuring device," US9445750, 2016.

[145] J. Paalasmaa, "Monitoring sleep with force sensor measurement," University of Helsinki, 2014.

[146] F. Wang, M. Tanaka, and S. Chonan, "Development of a PVDF piezopolymer sensor for unconstrained in-sleep cardiorespiratory monitoring," *J. Intell. Mater. Syst. Struct.*, vol. 14, no. 3, pp. 185–190, Mar. 2003.

[147] M. Brink, C. H. Müller, and C. Schierz, "Contact-free measurement of heart rate, respiration rate, and body movements during sleep," *Behav. Res. Methods*, vol. 38, no. 3, pp. 511–21, Aug. 2006.

[148] 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 *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012, vol. 2012, no. 11, pp. 5254–5257.

[149] S. H. Woodward *et al.*, "Estimating heart rate and RSA from the mattress-recorded kinetocardiogram," *Psychophysiology*, vol. 44, no. 4, pp. 635–8, Jul. 2007.

[150] S. Nukaya, M. Sugie, Y. Kurihara, T. Hiroyasu, K. Watanabe, and H. Tanaka, "A noninvasive heartbeat, respiration, and body movement monitoring system for neonates," *Artif. Life Robot.*, vol. 19, no. 4, pp. 414–419, Nov. 2014.

[151] B. H. Choi, G. S. Chung, J. S. Lee, D. U. Jeong, and K. S. Park, "Slow-wave sleep estimation on a load-cell-installed bed: a non-constrained method," *Physiol. Meas.*, vol. 30, no. 11, pp. 1163–70, Nov. 2009.

[152] "Angle Care," 2019. [Online]. Available: https://web.archive.org/web/20190712214933/https://www.angelcare.de/.

[153] "NannyCare," 2019. [Online]. Available: https://web.archive.org/web/20190618210246/https://www.nannycare.nl/ .

[154] "Nanny," 2019. [Online]. Available: https://web.archive.org/web/20190209191759/http://www.monitornanny. com/en/.

[155] J. J. Liu, S. Member, M. Huang, S. Member, and W. Xu, "BreathSens: a continuous on-bed respiratory monitoring system with torso localization using an unobtrusive pressure sensing array," *Biomed. Heal. informatics*, vol. 2194, no. c, pp. 1–7, 2014.

[156] M. Sekine and K. Maeno, "Non-contact heart rate detection using periodic variation in Doppler frequency," in *2011 IEEE Sensors Applications Symposium*, 2011, pp. 318–322.

[157] Changzhan Gu, Changzhi Li, Jenshan Lin, Jiang Long, Jiangtao Huangfu, and Lixin Ran, "Instrument-based noncontact doppler radar vital sign detection system using heterodyne digital quadrature demodulation architecture," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 6, pp. 1580–1588, Jun. 2010.

[158] J. S. Burnham, "Remote cardiac monitoring using radar," Massachusetts Institute of Technology, 2009.

[159] P. Marchionni, L. Scalise, I. Ercoli, and E. P. Tomasini, "An optical

measurement method for the simultaneous assessment of respiration and heart rates in preterm infants," *Rev. Sci. Instrum.*, vol. 84, no. 12, p. 121705, Dec. 2013.

[160] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiol. Meas.*, vol. 28, no. 3, pp. R1-39, Mar. 2007.

[161] E. Olsson, H. Ugnell, P. A. Oberg, and G. Sedin, "Photoplethysmography for simultaneous recording of heart and respiratory rates in newborn infants," *Acta Paediatr.*, vol. 89, no. 7, pp. 853–61, Jul. 2000.

[162] A. Johansson, "Neural network for photoplethysmographic respiratory rate monitoring," *Med. Biol. Eng. Comput.*, vol. 41, no. 3, pp. 242–8, May 2003.

[163] J. Lázaro, E. Gil, R. Bailón, A. Mincholé, and P. Laguna, "Deriving respiration from photoplethysmographic pulse width," *Med. Biol. Eng. Comput.*, vol. 51, no. 1–2, pp. 233–42, Feb. 2013.

[164] P. S. Addison, J. N. Watson, M. L. Mestek, and R. S. Mecca, "Developing an algorithm for pulse oximetry derived respiratory rate (RR(oxi)): a healthy volunteer study," *J. Clin. Monit. Comput.*, vol. 26, no. 1, pp. 45–51, Feb. 2012.

[165] D. Wertheim, C. Olden, E. Savage, and P. Seddon, "Extracting respiratory data from pulse oximeter plethysmogram traces in newborn infants," *Arch. Dis. Child. - Fetal Neonatal Ed.*, vol. 94, no. 4, pp. F301–F303, Jul. 2009.

[166] C. Mann *et al.*, "Can we improve delivery room monitoring for newborns? A novel photoplethysmographic heart rate monitor evaluated among stable nicu infants," *Arch. Dis. Child. - Fetal Neonatal Ed.*, vol. 96, no. Supplement 1, pp. Fa2–Fa3, Jun. 2011.

[167] A. Johansson, P. A. Oberg, and G. Sedin, "Monitoring of heart and respiratory rates in newborn infants using a new photoplethysmographic technique," *J. Clin. Monit. Comput.*, vol. 15, no. 7–8, pp. 461–7, Dec. 1999.

[168] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt. Express*, vol. 16, no. 26, p. 21434, Dec. 2008.

[169] Y. Sun, C. Papin, V. Azorin-Peris, R. Kalawsky, S. Greenwald, and S. Hu, "Use of ambient light in remote photoplethysmographic systems: comparison between a high-performance camera and a low-cost webcam," *J. Biomed. Opt.*, vol. 17, no. 3, p. 037005, Mar. 2012.

[170] N. Koolen *et al.*, "Automated respiration detection from neonatal video data," in *ICPRAM 2015*, 2015.

[171] L. A. M. Aarts *et al.*, "Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit — A pilot study," *Early Hum. Dev.*, vol. 89, no. 12, pp. 943–948, Dec. 2013.

[172] D. McDuff, S. Gontarek, and R. W. Picard, "Improvements in Remote Cardiopulmonary Measurement Using a Five Band Digital Camera," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 10, pp. 2593–2601, Oct. 2014.

[173] M. van Gastel, S. Stuijk, and G. de Haan, "Motion robust remote-PPG in infrared," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 5, pp. 1425–1433, May 2015.

[174] "IEC60601-2-21: International standard, medical electrical equipment – Part2.21. Particular requirements for the basic safety and essential performance of infant radiant warmer."

[175] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. A. Clifton, and C. Pugh, "Non-contact video-based vital sign monitoring using ambient light and auto-regressive models," *Physiol. Meas.*, vol. 35, no. 5, pp. 807–31, May 2014.

[176] Y. Sun, S. Hu, V. Azorin-Peris, S. Greenwald, J. Chambers, and Y. Zhu, "Motion-compensated noncontact imaging photoplethysmography to

monitor cardiorespiratory status during exercise," *J. Biomed. Opt.*, vol. 16, no. 7, p. 077010, Jul. 2011.

[177]   M. Scatena *et al.*, "An integrated video-analysis software system designed for movement detection and sleep analysis. Validation of a tool for the behavioural study of sleep," *Clin. Neurophysiol.*, vol. 123, no. 2, pp. 318–23, Feb. 2012.

[178]   M. Mirmiran, Y. G. H. Maas, and R. L. Ariagno, "Development of fetal and neonatal sleep and circadian rhythms," *Sleep Med. Rev.*, vol. 7, no. 4, pp. 321–334, Aug. 2003.

[179]   A. Roebuck *et al.*, "A review of signals used in sleep analysis," *Physiol. Meas.*, vol. 35, no. 1, pp. R1-57, Jan. 2014.

[180]   G. Calciolari and R. Montirosso, "The sleep protection in the preterm infants," *J. Matern. Neonatal Med.*, vol. 24, no. sup1, pp. 12–14, Oct. 2011.

[181]   S. Graven, "Sleep and brain development," *Clin. Perinatol.*, vol. 33, no. 3, pp. 693–706, Sep. 2006.

[182]   P. D. Peirano and C. R. Algarín, "Sleep in brain development," *Biol. Res.*, vol. 40, no. 4, pp. 471–8, Jan. 2007.

[183]   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 Med. Rev.*, vol. 32, pp. 109–122, Apr. 2017.

[184]   W. F. Liu, "The impact of a noise reduction quality improvement project upon sound levels in the open-unit-design neonatal intensive care unit," *J. Perinatol.*, vol. 30, no. 7, pp. 489–496, Jul. 2010.

[185]   W. F. Liu, S. Laudert, B. Perkins, E. MacMillan-York, S. Martin, and S. Graven, "The development of potentially better practices to support the

neurodevelopment of infants in the NICU," *J. Perinatol.*, vol. 27, no. S2, pp. S48–S74, Dec. 2007.

[186]   R. M. Harper, V. L. Schechtman, and K. A. Kluge, "Machine classification of infant sleep state using cardiorespiratory measures," *Electroencephalogr. Clin. Neurophysiol.*, vol. 67, no. 4, pp. 379–387, Oct. 1987.

[187]   G. Nason, T. Sapatinas, and A. Sawczenko, "Wavelet packet modelling of infant sleep state using heart rate data," *Indian J. Stat. Ser. B*, vol. 63, no. 2, pp. 199–217, 2001.

[188]   A. T. Lewicke, E. S. Sazonov, and S. A. C. Schuckers, "Sleep-wake identification in infants: heart rate variability compared to actigraphy," in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2004, vol. 3, pp. 442–445.

[189]   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 Trans. Biomed. Eng.*, vol. 55, no. 1, pp. 108–118, Jan. 2008.

[190]   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," *J. Med. Syst.*, vol. 35, no. 4, pp. 693–702, Aug. 2011.

[191]   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," *Dev. Psychobiol.*, vol. 58, no. 8, pp. 1108–1115, Dec. 2016.

[192]   M. S. Scher and K. a Loparo, "Neonatal EEG/sleep state analyses: a complex phenotype of developmental neural plasticity," *Dev. Neurosci.*, vol. 31, no. 4, pp. 259–75, Jan. 2009.

[193]   F. Ebrahimi, S.-K. Setarehdan, and H. Nazeran, "Automatic sleep staging by simultaneous analysis of ECG and respiratory signals in long epochs," *Biomed. Signal Process. Control*, vol. 18, pp. 69–79, Apr. 2015.

[194]   P. Fonseca, X. Long, M. Radha, R. Haakma, R. M. Aarts, and J. Rolink, "Sleep stage classification with ECG and respiratory effort," *Physiol. Meas.*, vol. 36, no. 10, pp. 2027–2040, Oct. 2015.

[195]   P. Fonseca, N. den Teuling, X. Long, and R. M. Aarts, "Cardiorespiratory sleep stage detection using conditional random fields," *IEEE J. Biomed. Heal. Informatics*, vol. 21, no. 4, pp. 956–966, Jul. 2017.

[196]   A. Dereymaeker *et al.*, "Review of sleep-EEG in preterm and term neonates," *Early Hum. Dev.*, no. xxxx, 2017.

[197]   H. F. R. Prechtl, "The behavioural states of the newborn infant (a review)," *Brain Res.*, vol. 76, no. 10, pp. 185–212, 1974.

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

[199]   R. J. Ellis, B. Zhu, J. Koenig, J. F. Thayer, and Y. Wang, "A careful look at ECG sampling frequency and R-peak interpolation on short-term measures of heart rate variability," *Physiol. Meas.*, vol. 36, no. 9, pp. 1827–1852, Sep. 2015.

[200]   R. W. C. G. R. Wijshoff, M. Mischi, and R. M. Aarts, "Reduction of periodic motion artifacts in photoplethysmography," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 196–207, Jan. 2017.

[201]   M. Malik, "Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology," *Circulation*, vol. 93, no. 5, pp. 1043–65, Mar. 1996.

[202] P. Indic, E. B. Salisbury, D. Paydarfar, E. N. Brown, and R. Barbieri, "Interaction between heart rate variability and respiration in preterm infants," in *2008 Computers in Cardiology*, 2008, vol. 35, pp. 57–60.

[203] T. Ruf, "The Lomb-Scargle periodogram in biological rhythm research: analysis of incomplete and unequally spaced time-series," *Biol. Rhythm Res.*, vol. 30, no. 2, pp. 178–201, Apr. 1999.

[204] P. K. Stein and Y. Pu, "Heart rate variability, sleep and sleep disorders," *Sleep Med. Rev.*, vol. 16, no. 1, pp. 47–66, Feb. 2012.

[205] X. Long, R. Haakma, T. R. M. Leufkens, P. Fonseca, and R. M. Aarts, "Effects of between- and within-subject variability on autonomic cardiorespiratory activity during sleep and their limitations on sleep staging: a multilevel analysis," *Comput. Intell. Neurosci.*, vol. 2015, pp. 1–17, 2015.

[206] O. M. Doyle, I. Korotchikova, G. Lightbody, W. Marnane, D. Kerins, and G. B. Boylan, "Heart rate variability during sleep in healthy term newborns in the early postnatal period," *Physiol. Meas.*, vol. 30, no. 8, pp. 847–860, Aug. 2009.

[207] S. Reulecke, "Autonomic regulation during quiet and active sleep states in very preterm neonates," *Front. Physiol.*, vol. 3, no. 4, pp. 1–9, 2012.

[208] D. R. Kommers *et al.*, "Features of heart rate variability capture regulatory changes during kangaroo care in preterm infants," *J. Pediatr.*, vol. 182, pp. 92-98.e1, Mar. 2017.

[209] J. E. Mietus, "The pNNx files: re-examining a widely used heart rate variability measure," *Heart*, vol. 88, no. 4, pp. 378–380, Oct. 2002.

[210] H. Mark, "Correlation-based feature selection of discrete and numeric class machine learning," Waikato, 08, May 2000.

[211] G. Cuaya, A. Muñoz-Meléndez, and E. F. Morales, "A minority class feature selection method," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, Berlin: Springer Berlin Heidelberg, 2011, pp. 417–424.

[212] F. Pedregosa *et al.*, "Scikit-learn: machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[213] C. Bishop, *Pattern recognition and machine learning*. New York, NY: Springer-Verlag New York, 2006.

[214] S. Raschka, *Python maschine learning*. Birmingham: Packt Publishing Ltd., 2015.

[215] C. M. Huang, W. S. Tung, L. L. Kuo, and Y. J. Chang, "Comparison of pain responses of premature infants to the heelstick between containment and swaddling," *J. Nurs. Res.*, vol. 12, no. 1, pp. 31–35, Mar. 2004.

[216] J. O. E. H. van Laar, C. H. L. Peters, R. Vullings, S. Houterman, and S. Oei, "Power spectrum analysis of fetal heart rate variability at near term and post term gestation during active sleep and quiet sleep," *Early Hum. Dev.*, vol. 85, no. 12, pp. 795–798, Dec. 2009.

[217] U. Rajendra Acharya, K. Paul Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: a review," *Med. Biol. Eng. Comput.*, vol. 44, no. 12, pp. 1031–1051, Dec. 2006.

[218] X. Long, *On the analysis and classification of sleep stages from cardiorespiratory activity*. Eindhoven: Technische Universiteit Eindhoven, 2015.

[219] M. Lucchini, W. P. Fifer, A. Perez, and M. G. Signorini, "Influence of sleep state and position on cardio-respiratory regulation in newborn babies," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, no. January 2016, pp. 302–305.

[220] M. Lucchini, W. P. Fifer, R. Sahni, and M. G. Signorini, "Novel heart rate parameters for the assessment of autonomic nervous system function in premature infants," *Physiol. Meas.*, vol. 37, no. 9, pp. 1436–1446, Sep. 2016.

[221] S. Abe, *Support vector machines for pattern classification*. London: Springer Verlag, 2005.

[222] W. Chen, S. Bambang Oetomo, L. Feijs, S. Bouwstra, I. Ayoola, and S. Dols, "Design of an integrated sensor platform for vital sign monitoring of newborn infants at neonatal intensive care units," *J. Healthc. Eng.*, vol. 1, no. 4, pp. 535–553, Oct. 2010.

[223] D. Potuzakova, W. Chen, S. B. Oetomo, and L. Feijs, "Innovative design for monitoring of neonates using reflectance pulse oximeter," in *2011 Seventh International Conference on Intelligent Environments*, 2011, pp. 200–205.

[224] S. Rajala and J. Lekkala, "Film-type sensor materials PVDF and EMFi in measurement of cardiorespiratory signals: a review," *IEEE Sens. J.*, vol. 12, no. 3, pp. 439–446, Mar. 2012.

[225] J. Kranjec, S. Beguš, G. Geršak, and J. Drnovšek, "Non-contact heart rate and heart rate variability measurements: a review," *Biomed. Signal Process. Control*, vol. 13, pp. 102–112, Sep. 2014.

[226] S. Lange, P. Van Leeuwen, D. Geue, W. Hatzmann, and D. Grönemeyer, "Influence of gestational age, heart rate, gender and time of day on fetal heart rate variability," *Med. Biol. Eng. Comput.*, vol. 43, no. 4, pp. 481–486, Aug. 2005.

[227] E. Rosenstock, Y. Cassuto, and E. Zmora, "Heart rate variability in the neonate and infant: analytical methods, physiological and clinical observations," *Acta Paediatr.*, vol. 88, no. 5, pp. 477–482, Jan. 2007.

[228] J. O. E. H. Van Laar, M. Porath, C. H. L. Peters, and S. Oei, "Spectral

analysis of fetal heart rate variability for fetal surveillance: review of the literature," *Acta Obstet. Gynecol. Scand.*, vol. 87, no. 3, pp. 300–306, Jan. 2008.

[229] J. Werth *et al.*, "Unobtrusive assessment of neonatal sleep state based on heart rate variability retrieved from electrocardiography used for regular patient monitoring," *Early Hum. Dev.*, vol. 113, pp. 104–113, Oct. 2017.

[230] V. Gerla, M. Bursa, L. Lhotska, K. Paul, and V. Krajca, "Newborn sleep stage classification using hybrid evolutionary approach," *Int. J. Bioelectromagn.*, vol. 9, no. 1, pp. 25–26, 2007.

[231] R. Otte and X. Long, "A framework for observational coding of sleep in infants," *Prep.*

[232] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochem. medica*, vol. 22, no. 3, pp. 276–82, 2012.

[233] D. H. Gordon, "Triboelectric interference in the ECG," *IEEE Trans. Biomed. Eng.*, vol. BME-22, no. 3, pp. 252–255, May 1975.

[234] D. E. Becker, "Fundamentals of electrocardiography interpretation," *Anesth. Prog.*, vol. 53, no. 2, pp. 53-63;, 2006.

[235] J. Veen *et al.*, "Electro-physiological measurement with reduced motion artifacts," US9603542B2, 2011.

[236] A. Serteyn, R. Vullings, M. Meftah, and J. W. M. Bergmans, "Motion artifacts in capacitive ECG measurements: reducing the combined effect of DC voltages and capacitance changes using an injection signal," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 1, pp. 264–273, Jan. 2015.

[237] R. Vullings, C. H. L. Peters, M. J. M. Hermans, P. F. F. Wijn, S. G. Oei, and J. W. M. Bergmans, "A robust physiology-based source separation method for QRS detection in low amplitude fetal ECG recordings," *Physiol. Meas.*,

vol. 31, no. 7, pp. 935–951, Jul. 2010.

[238] M. J. Rooijakkers, C. Rabotti, S. G. Oei, and M. Mischi, "Low-complexity R-peak detection for ambulatory fetal monitoring," *Physiol. Meas.*, vol. 33, no. 7, pp. 1135–1150, Jul. 2012.

[239] J. M. Yentes, N. Hunt, K. K. Schmid, J. P. Kaipust, D. McGrath, and N. Stergiou, "The appropriate use of approximate entropy and sample entropy with short data sets," *Ann. Biomed. Eng.*, vol. 41, no. 2, pp. 349–365, Feb. 2013.

[240] D. E. Lake, "Improved entropy rate estimation in physiological data," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 1463–1466.

[241] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.

[242] Haibo He, Yang Bai, E. A. Garcia, and Shutao Li, "ADASYN: adaptive synthetic sampling approach for imbalanced learning," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008, pp. 1322–1328.

[243] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. NewYork: Wiley, 2000.

[244] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Front. Neurorobot.*, vol. 7, p. 21, 2013.

[245] P. Fonseca, R. M. Aarts, X. Long, J. Rolink, and S. Leonhardt, "Estimating actigraphy from motion artifacts in ECG and respiratory effort signals," *Physiol. Meas.*, vol. 37, no. 1, pp. 67–82, Jan. 2016.

[246]   L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[247]   A. H. Ansari *et al.*, "Quiet sleep detection in preterm infants using deep convolutional neural networks," *J. Neural Eng.*, vol. 15, no. 6, Dec. 2018.

[248]   V. Doria *et al.*, "Emergence of resting state networks in the preterm human brain," *Proc. Natl. Acad. Sci.*, vol. 107, no. 46, pp. 20015–20020, Nov. 2010.

[249]   J. Werth, A. Serteyn, P. Andriessen, R. M. Aarts, and X. Long, "Automated preterm infant sleep staging using capacitive electrocardiography," *Physiol. Meas.*, vol. 40, no. 5, p. 055003, Jun. 2019.

[250]   M. Radha, G. Garcia-Molina, M. Poel, and G. Tononi, "Comparison of feature and classifier algorithms for online automatic sleep staging based on a single EEG signal," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 1876–1880.

[251]   A. Procházka, J. Kuchyňka, O. Vyšata, P. Cejnar, M. Vališ, and V. Mařík, "Multi-class sleep stage analysis and adaptive pattern recognition," *Appl. Sci.*, vol. 8, no. 5, p. 697, May 2018.

[252]   X. Long, P. Fonseca, R. Haakma, and R. M. Aarts, "Actigraphy-based sleep/wake detection for insomniacs," in *2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 2017, pp. 1–4.

[253]   J. Schmidhuber, "Deep learning in neural networks: an overview," *Neural Networks*, vol. 61, pp. 85–117, Jan. 2015.

[254]   S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[255]   K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *Circuits Syst. 2006. APCCAS 2006. IEEE*

Asia Pacific Conf., Jun. 2014.

[256] S. Biswal *et al.*, "SLEEPNET: Automated sleep staging system via deep learning," *Computing Research Repository*, vol. eprint. 25-Jul-2017.

[257] S. Chambon, M. N. Galtier, P. J. Arnal, G. Wainrib, and A. Gramfort, "A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 4, pp. 758–769, Apr. 2018.

[258] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: a model for automatic sleep stage scoring based on raw single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 1998–2008, Nov. 2017.

[259] A. N. Olesen, P. E. Peppard, H. B. Sorensen, P. J. Jennum, and E. Mignot, "0316 end-to-end deep learning model for automatic sleep staging using raw PSG waveforms," *Sleep*, vol. 41, no. suppl_1, pp. A121–A121, Apr. 2018.

[260] A. Sano, W. Chen, D. Lopez-Martinez, S. Taylor, and R. W. Picard, "Multimodal ambulatory sleep detection using LSTM recurrent neural networks," *IEEE J. Biomed. Heal. Informatics*, vol. 23, no. 4, pp. 1607–1617, Jul. 2019.

[261] M. Radha, P. Fonseca, M. Ross, A. Cerny, P. Anderer, and R. M. Aarts, "LSTM knowledge transfer for HRV-based sleep staging," *Quant. Biol.*, vol. eprint, no. 9, pp. 1–11, Sep. 2018.

[262] R. Joshi *et al.*, "Cardiorespiratory coupling in preterm infants," *J. Appl. Physiol.*, p. japplphysiol.00722.2018, Nov. 2018.

[263] K. Cross and T. Oppe, "The respiratory rate and volume in the premature infant." *J. Physiol.*, vol. 116, no. 2, pp. 168–74, Feb. 1952.

[264] F. Chollet and others, "Keras." 2015.

[265] M. Abadi *et al.*, "TensorFlow: large-scale machine learning on heterogeneous distributed systems," Mar. 2016.

[266] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Front. Psychol.*, vol. 4, Dec. 2015.

[267] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, vol. 1, pp. 5987–5995.

[268] D. Balduzzi, M. Frean, L. Leary, J. Lewis, K. W.-D. Ma, and B. McWilliams, "The shattered gradients problem: if resnets are the answer, then what is the question?," *Comput. Res. Repos.*, vol. eprint, Feb. 2017.

[269] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[270] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," *Comput. Res. Repos.*, vol. eprint, 2014.

[271] R. Joshi *et al.*, "A ballistographic approach for continuous and non-obtrusive monitoring of movement in neonates," *IEEE J. Transl. Eng. Heal. Med.*, vol. 6, pp. 1–10, 2018.

[272] D. H. Crowell *et al.*, "Infant polysomnography: reliability. Collaborative Home Infant Monitoring Evaluation (CHIME) Steering Committee," *Sleep*, 1997.

[273] H. Danker-Hopfe *et al.*, "Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard.," *J. Sleep Res.*, vol. 18, no. 1, pp. 74–84, Mar. 2009.

[274] K. Woods and K. W. Bowyer, "Generating ROC curves for artificial neural

networks," *IEEE Trans. Med. Imaging*, vol. 16, no. 3, pp. 329–337, Jun. 1997.

[275] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and f-score, with implication for evaluation," in *ECIR 2005: Advances in Information Retrieval*, 2005, pp. 345–359.

[276] M. Zihlmann, D. Perekrestenko, and M. Tschannen, "Convolutional recurrent neural networks for electrocardiogram classification," in *Computing in Cardiology*, 2017.

[277] J. Zhang, J. Tian, Y. Cao, Y. Yang, X. Xu, and C. Wen, "Fine-grained ECG classification based on deep CNN and online decision fusion," *Comput. Res. Repos.*, vol. abs/1901.0, Jan. 2019.

[278] J. T. Ruiz, J. D. B. Pérez, and J. R. B. Blázquez, "Arrhythmia detection using convolutional neural models," in *Distributed Computing and Artificial Intelligence*, 2019, 16th ed., pp. 120–127.

[279] "Collaborative Home Infant Monitoring Evaluation," 1998. [Online]. Available: https://web.archive.org/web/20151028224033/http://slone-web2.bu.edu/ChimeNisp/Main_Chime.asp.

[280] H. Jin, Q. Song, and X. Hu, "Auto-Keras: efficient neural architecture search with network morphism," Jun. 2018.

[281] C. Cortes, X. Gonzalvo, V. Kuznetsov, M. Mohri, and S. Yang, "AdaNet: adaptive structural learning of artificial neural networks," Jul. 2016.

[282] A. Krizhevsky, V. Nair, and G. Hinton, "CIFAR-10 and CIFAR-100 datasets," 2009. [Online]. Available: https://www.cs.toronto.edu/~kriz/cifar.html.

[283] Y. LeCun and C. Cortes, "MNIST handwritten digit database," 2010. [Online]. Available: http://yann.lecun.com/exdb/mnist/.

[284] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms," Aug. 2017.

[285] V. A. Harpin and N. Rutter, "Sweating in preterm babies," *J. Pediatr.*, vol. 100, no. 4, pp. 614–9, Apr. 1982.

[286] P. Tourneux *et al.*, "Influence of thermal drive on central sleep apnea in the preterm neonate," *Sleep*, vol. 31, no. 4, pp. 549–56, Apr. 2008.

[287] L. Curzi-Dascalova, L. Christova-Guéorguiéva, F. Lebrun, and G. Firtion, "Respiratory pauses in very low risk prematurely born infants reaching normal term. A comparison to full-term newborns," *Neuropediatrics*, vol. 15, no. 1, pp. 13–7, Feb. 1984.

[288] M. Gabriel, M. Albani, and F. J. Schulte, "Apneic spells and sleep states in preterm infants," *Pediatrics*, vol. 57, no. 1, pp. 142–7, Jan. 1976.

[289] G. W. Don and K. A. Waters, "Influence of sleep state on frequency of swallowing, apnea, and arousal in human infants," *J. Appl. Physiol.*, vol. 94, no. 6, pp. 2456–64, Jun. 2003.

[290] N. Feeley, P. Zelkowitz, C. Cormier, L. Charbonneau, A. Lacroix, and A. Papageorgiou, "Posttraumatic stress among mothers of very low birthweight infants at 6 months after discharge from the neonatal intensive care unit," *Appl. Nurs. Res.*, vol. 24, no. 2, pp. 114–117, May 2011.

[291] S. Gangi, D. Dente, E. Bacchio, S. Giampietro, G. Terrin, and M. De Curtis, "Posttraumatic stress disorder in parents of premature birth neonates," *Procedia - Soc. Behav. Sci.*, vol. 82, pp. 882–885, Jul. 2013.

[292] T. Ramezani, Z. Hadian Shirazi, R. Sabet Sarvestani, and M. Moattari, "Family-centered care in neonatal intensive care unit: a concept analysis," *Int. J. community based Nurs. midwifery*, vol. 2, no. 4, pp. 268–78, Oct. 2014.

[293]   C. R. Brewin, B. Andrews, and S. Rose, "Fear, helplessness, and horror in posttraumatic stress disorder: investigating DSM-IV Criterion A2 in victims of violent crime," *J. Trauma. Stress*, vol. 13, no. 3, pp. 499–509, Jul. 2000.

[294]   W. A. Silverman, "Incubator-baby side shows (Dr. Martin A. Couney)," *Pediatrics*, vol. 64, no. 2, pp. 127–41, Aug. 1979.

[295]   J. Albritton, L. Maddox, J. Dalto, E. Ridout, and S. Minton, "The effect of a newborn telehealth program on transfers avoided: a multiple-baseline study," *Health Aff.*, vol. 37, no. 12, pp. 1990–1996, Dec. 2018.

[296]   "Steth IO," 2019. [Online]. Available: https://web.archive.org/web/20190712214157/https://stethio.com/stethoscope_features/.

[297]   C. D. Gerardo, E. Cretu, and R. Rohling, "Fabrication and testing of polymer-based capacitive micromachined ultrasound transducers for medical imaging," *Microsystems Nanoeng.*, vol. 4, no. 1, p. 19, Dec. 2018.

[298]   Google, "Google tensor processing unit," 2018. .

[299]   D. Karch, K. Rohmer, and P. Lemburg, "Prognostic significance of polygraphic recordings in newborn infants on ventilation," *Dev. Med. Child Neurol.*, vol. 26, no. 3, pp. 358–68, Jun. 1984.

[300]   C. Dreyfus-Brisac, *Human Growth 3, Neurobiology and Nutrition*. New York: Plenum Press, 1979.

[301]   "Tekscan," 2019. [Online]. Available: https://web.archive.org/web/20190712214157/https://stethio.com/stethoscope_features/.

List of authors publications

Peer reviewed journal publications

J. Werth, M. Radha, P. Andriessen, R. M. Aarts, X. Long, "Deep learning approach for ECG-based automatic sleep state classification in preterm infants," *Biomedical Biomed. Signal Process. Control*, 56:101663, Feb. 2020

J. Werth, A. Serteyn, P. Andriessen, R.M. Aarts, X. Long, "Automated preterm infant sleep staging using capacitive electrocardiography," *Physiol. Meas.*, 40(5):055003, Jun. 2019

X. Long., R. Otte, E. van Sanden, J. Werth, T. Tan, "Video-Based Actigraphy for Monitoring Wake and Sleep in Healthy Infants: A Laboratory Study," *Sensors,* 19(5):1075, Mar 2019

J. Werth, X. Long, E. Zwartkruis-Pelgrim, H. J. Niemarkt, W. Chen, R. M. Aarts, P. Andriessen, "Unobtrusive assessment of neonatal sleep state based on heart rate variability retrieved from electrocardiography used for regular patient monitoring*," Early Hum. Dev.,* 113:104–113, Oct. 2017

J. Werth, L. Atallah, P. Andriessen, X. Long, E. Zwartkruis-Pelgrim, R.M. Aarts, "Unobtrusive sleep state measurements in preterm infants - A review," *Sleep Med Rev.,* 32:109-122, Apr 2016

Conference papers

X. Long, J. Werth, J. E. Perez, "Comparing video actigraphy across premature infant sleep states," *IPSA 2018*, Apr 2018

J. Werth, L. Atallah, X. Long, P. Andriessen, E. Zwartkruis-Pelgrim, R. M. Aarts, "Preterm infant heart rate variability feature selection for automated sleep state classification," *Biomedica Summit 2016*, May 2016

M. Nano, X. Long, J. Werth, R. M. Aarts, R. Heusdens, "Sleep apnea detection using time-delayed heart rate variability," *Conf Proc IEEE Eng Med Biol Soc. 2015*:7679-82, 2015

Published patents

J. Kahlman, J. Werth, R. Bezemer, "Device, system and method for detecting a cardiac and/or respiratory disease of a subject", European Patent Office, EP3282950A1

L. Atallah, J. Werth, "Method and apparatus for determining a health status of an infant," European Patent Office, EP3378386A1

Acknowledgments

This chapter is taking quite some space in this thesis, as I spend several wonderful years at Philips, where I met many interesting and magnificent people. All of you helped me directly or indirectly to make it a very pleasant and salutary memory. I will not name all of you, but I will focus on the persons I met most frequently. Please feel thanked and thought of even if you might not appear in name.

Dear Ronald Aarts, I want to thank you most sincerely for taking me under your wing and share your boundless experience with me. You had the intuition to ask the right questions at the right time, driving me towards publications. You gave me the guidance and support I needed at the time I needed it. If I ever can be just half as influential on technology as you are, I will be more than satisfied with my professional life. As a realistic person, I should reduce that number to even have a chance to come close. Ph.D. hope you can enjoy your retirement to the fullest and continue to support the people around you with your sharp and analytic mind. Hartelijk bedankt!

Dear Xi Long, I am struggling to find the right words to thank you. You are a great supervisor and friend. At the right times, you would have comforting, critical, or supporting advice. Always on point. Without discrediting the great supervision of your precursor, I noticed that my Ph.D. got a different momentum after you stepped into the ring. Every student you supervise can be called lucky. Additionally, to have you as a researcher at my site was a great experience, as I could see your brilliant analytic mind directly at work. I am looking forward to hearing your inaugural lecture. Finally, I wish you all the best with your new family, that you might have plenty of healthy kids and enjoy a good ratio between exciting research at home and at work. 願你好運永遠與你同在.

Dear Peter Andriessen, first, I want to thank you for the time and effort you put into correcting my manuscripts from top to bottom. As we grow over our development as researchers, you helped me tremendously to take the first steps correctly. Second,

I want to thank you for the time you made available for me while you had limited time due to your daily heroic work of saving preterm infants lives. I hope my research can help you, in the long run, to make up for the invested energy. I wish you all the best and that you will experience automated preterm infant sleep measures at one point in your career in your NICU.

Dear Sabine van Huffel, I want to thank you that you took the time and effort to be on my defense committee. I always enjoyed talking to you, and I am again looking forward to a vivid discussion during and after the ceremony. Too bad that the joint PhD did not work out. However, you being in my committee makes up twice for this. I hope that we will meet again, maybe in the beautiful city of Leuven.

prof.dr.ir. Peter Baltus, thank you for chairing my Ph.D. ceremony. I appreciate your time and promise to represent the TU/e in the best way possible, adding to the good reputation of the electrical engineering department.

Dear Lou Feijs, it was always very stimulating to meet you at the IMPULS meetings and to get good ideas for further research thanks to your thoughtful and precise comments. Thank you for taking the time to be part of my defense committee.

prof.dr. Sebastiaan Overeem, thank you for taking the time to be part of my defense committee. Unfortunately, we have not met yet, but I am looking forward to meeting you and having interesting discussions during and after the defense ceremony.

Dear Hans Weda, thank you for taking over as a supervisor even though it was only for a short duration. If you would have been there from the beginning, I cannot imagine what I would have learned in the fields of data analytics, statistics, and machine learning. Luckily, I had also other highly capable colleagues of yours supervising me. I enjoyed exchanging thoughts in all areas and discussing the "la distinction" a la Bourdieu between the Dutch and Germans with you. To be continued. I will always look up to you.

Dear Jan Bergmanns, thank you for being a great IMPULS coordinator backing us with support and advice whenever needed. I remember amused when you reassured me after I found out that the pizza night was paid by my Ph.D. budget. As has been proven, you were right; the budged lasted to finalize my Ph.D.

Dear Carola van Pul, thank you for looking after me at the MMC. I felt warmly welcomed and supported. Even though you had always an agenda fuller than the day has hours, you made time for me to go through any question in detail. Thank you for that. I also want to thank you for the inspiring discussions we had and your ideas on the topic of preterm infant sleep. Too bad that we never converted your idea to compare preterm infant sleep and actual brain growth. I hope that a capable Ph.D. candidate will read this and contact you.

Dear Elly Zwartkruis-Pelgrim, thank you for being the lovely shark that the department needs. You combine a kind and radiant nature with the bite of a great white. If that is not the perfect combination for a leading position, I do not know what is. While as a supervisor, you offered me help wherever you could and time allowed. As my department head, you would listen to my contributions and suggestions, and you jumped through hoops to secure my future at Philips. It was an honor to research with you.

Marjolein van Lieshout, Louis Atallah, and Richard Kemkers thank you for trusting in me and my ability to finish this assignment by choosing me for the Ph.D. position at Philips Research. Sorry that I got the corporate haircut.

Liebster Paul, großartig, dass aus unserer ersten Begegnung, bei der wir beide noch nicht viel miteinander anfangen konnten, eine solch innige Beziehung geworden ist. Auf das Menschliche kommt es halt an. Es hat schon Spaß gemacht. Ich denke gerne zurück an die gemeinsamen Kaffeepausen, das Jammern, das Lachen, und Erfinden. Ich glaube Philips war für die meisten unserer Ideen einfach noch nicht bereit. Ich hoffe wir verlieren uns nicht aus den Augen. Wenigstens werde ich immer bei Herrn Herbst an dich denken, was ja zum Glück regelmäßig vorkommt. Ich wünsch dir eine

steile Karriere (positive slope) und Familienglück. Bis auf bald mein Bester, auf dass der bessere Matt setze.

Dearest Rick, the RAM of my PC is too small to write everything up I want to thank you for. You know that you gave me great advice at every step of the way, from small pointers to directional once. You also know your crisp, clear analysis of scientific and interpersonal questions, which you shared with me. However, I would rather point out how I enjoyed our great friendship and hilarious conversations. I still remember our bus rides to Aachen vividly, where we discussed our common passions. I wish you all the best in all categories where there are wishes to grant. Amsterdam is lucky to have you. I hope we find us again through the haze of time.

Dear Bishal, thank you for our fun get-togethers on the third-floor coffee corner. I know now that a person can learn a lot in 10min. It was great to have a knowledgeable friend who would take the time out of his day at any moment to brainstorm with me on solutions for my problems or simply have a fun discussion about abstruse philosophical questions. I read that people are working on our idea of a crypto dl-layer. Nice that our idea will produce some good. Have a great time and success in all parts of your life. You are a kind and gentle person. P.s., your countryman cheated on the game; you should have a word with him.

Dear Toine, thanks for making Amsterdam unforgettable for us. It would be only half the place without you. You will be my main Dutch archetype whenever I will talk about the Netherlands, if that is not something. My chest is still sore from our training sessions, and I remember now to focus more on keeping lunch in during training. I wish you all the best on your path; may it cross with mine again.

Dear Rohan, I very much enjoyed sharing our passion for the Appalachian Trail. I hope we meet someday accidentally on the way. After Werth's law: "The lower the probability, the more likely it will happen". It was sometimes hard to be your co-PhD with your pace of output, but as a fine person you are, you always offered your help and collaboration. Too bad, we never published together. One day we might share a

patent. Let's focus on that. You are an outstanding scientist. People will notice by themselves.

Dear Renée Otte, thanks for keeping the girls in line and being supportive all the way. I wish you and your family well. Best of luck.

Dear Mustafah Radah, thank you for answering the phone weekend or nighttime to discuss matters of ANNs with me. It is always good to have a capable sparring partner, and I could not have wished for a better one. Thank you for teaching me the fine art of tuning neural networks. I am looking forward to our next call.

Dear Ayline Sertein, it was great to have such a joyful and uplifting college. I will miss your critical thinking and precise questions. I wish you all the best with your home. If you put in laminaat, I will personally file a complaint with the bouwautoriteit.

Bester Niels, ich schreibe diesen Text gerade um 23:00 am Samstag, kurz nachdem du für mich noch den Druck organisiert hast. Vielen Dank für deine spontane Hilfe egal zu welcher Uhrzeit. Ohne dich würde das Buch mit Sicherheit aus zusammengetackerten Karopapierblättern bestehen. Ich denke wir sehen uns die Tage im Kindergarten.

Liebe Kristina und Marlene, vielen Dank, dass ich dank euch jetzt mein eigenes Gehirn besser verstehe, auch wenn es sich bei dem Thema verknotet hat.

Dear Richard Lewis, thank you for the time and effort you put into the community. Your contribution helped me through some slow and dragging times during my Ph.D. Please keep up the great work despite the loud voice trying to baulk you. I wish you further success on your path and that you will be seen for who you are.

Dear Joanneke Groen, Jorn op den Buijs, Ton Ackermanns, Steffen Pauws, Michel van Bruggen, and Jörg Habetha, thank you for giving me such a wonderful starting experience at Philips. You created my desire to stay at Philips. I wish you all success and peace in life.

Thanks to the creators and maintainers of scikit-learn, Keras, and Tensorflow as it made it viable to operate with machine- and deep learning for newcomers such as me. Please keep up the great work. I, for my part, will try my best to support this community.

I want to thank the Museo Nacional del Prado for allowing me to use the masterful painting of Jheronimus Bosch "De keisnijding" for this thesis print cover.

To the team of Draw.io, thank you for offering this great, free tool, which I used plenty of times during the last couple of years and will continue to do so.

Thank you L. Gaben for CS

Liebster Beni, hier fängt der Familienabschnitt der Danksagungen an. Schön, dass wir so eng verbunden sind und ich mich jedes Mal freue dich zu sehen oder hören, auch wenn ich danach zum Spray greifen muss. Unsere Gespräche zum Ende meines Masters haben mich dazu geführt mich auf die Doktorstelle zu bewerben. Auch haben mich deine wertvollen Tipps durch die letzten Jahre getragen. Ich danke dir für deine innige Freundschaft und freue mich auf alle noch kommenden Jahre mit dir.

Liebste Nina, allerbesten Dank, dass du uns immer wieder selbstlos unterstützt. Das spontane Übernehmen der Kinder hat mich mental entlastet und mir benötigte Zeiträume geschaffen. Ich höre schon das Wort „selbstverständlich" aus deinem Mund, es ist aber eher hochherzig und bewunderungswürdig. Toll, dass du da bist. Bis nachher.

Liebster Peter, Paul, und Franz, danke für Eure mitleidigen Blicke in der Bretagne, wenn ich mich hinter den Laptop gezwungen habe. Ebenso für die Streitgespräche über Formen der Intelligenz, in denen ich schlussendlich immer recht behalten habe ;-). Ich habe jetzt auch wieder Zeit und freue mich schon auf unsere gemeinsame Firma.

Liebste Petra, ich habe mich immer sehr gefreut wie du Anteil an meiner Arbeit und den Publikationen genommen hast. In meiner Wichtigsten darfst du nicht fehlen. Bis nachher.

Liebste Mutter, liebster Vater, danke für eure Erziehung. Ein wissenschaftlicher Geist ist nicht angeboren. Das Interesse an Neuem ist nicht angeboren. Der Durchhalte-willen ist nicht angeboren. Urvertrauen ist nicht angeboren. Meinen Weg habe ich euch zu verdanken und ich bin sehr Glücklich über alle genommenen Abzweigungen. Danke, für das gute Schuhwerk das Ihr mir mitgegeben habt.

Meine Jungs, gibt es was Schöneres als nach einem langen Tag zu euch zu kommen und sich mit dem Wesentlichen zu befassen? Ihr habt mir jeden Abend den Kopf geerdet. Danke, dass Ihr bei uns seid. Ich liebe Euch.

Allerliebste Maria, wenn ich anfange alles in Worte zu fassen für das ich dir dankbar bin, kann ich dir nur Unrecht tun. Es könnte nur ein winziger Bruchteil sein. Danke für Alles. Ich liebe Dich.

"De keisnijding" by Jheronimus Bosch

ca. 1494; Oil on Oak. Museo Nacional del Prado.

De keisnijding, the extraction of the stone of madness, or the cure of folly are different names for the painting by Jheronimus (or later Hieronymus) Bosch, showing a doctor performing a maniacal practice of removing part of the brain to cure the patient's madness. The procedure is observed by a priest and nun, symbolizing knowledge at the time of Bosch. The closed book on the head of a nun and the inverted funnel on the head of a doctor symbolize that their knowledge is meaningless when combined with absurdity and that the art of healing is quackery.

I chose this painting for the thesis book cover, as I see some resemblance and connections with my Ph.D. At first, Hieronymus Bosch is one of my favorite painters, who happened to be born in the Netherlands, where my Ph.D. took place. S-Hertogenbosch, his place of birth is even quite close to Eindhoven. During my Ph.D. his 500 birthday was celebrated in 2016. However, I also see a resemblance to the here presented work, as I tried to gain knowledge of the preterm infant brain and development with comparably rudimentary methods of ECG analysis compared to CT scans, EEG analysis, and other more elaborate methods. However, my hope was to deliver an instrument for the physicians and caretakers that they might gain inside to the preterm infant brain without using manual or obtrusive methods like most of the more elaborate methods; rendering antiquated manual and "opening the skull" methods obsolete. Further, the idea of development monitoring aims towards curing or preventing madness at an early stage, resembling the extraction of the stone of madness, turning this ancient superstition into reality. Lastly, it could also be interpreted that I needed to cure my own folly with gaining knowledge during my Ph.D. and preventing myself from madness by finalizing this thesis.

About the author
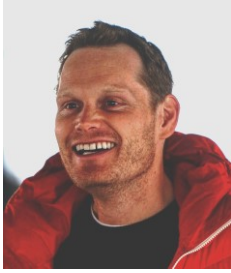
Jan Werth was born 1982 in Mainz, Germany. After his civil service, he moved to the city of Stuttgart in 2003 to become a journeyman in the trade of carpeting. His journeyman's piece was added to the BM-archive for excellent design. After successfully working as a carpenter in the city of Trier, he started his studies of electrical engineering at the University of Applied Science in Trier. He finished his studies in 2014 with a Master of Science degree in electrical engineering. The topics he was working on were mainly in the health care domain, for example, a biofeedback system based on EEG, or an impedance measurement chamber modeling for in vitro tissue measurements in the laboratory of Prof. Dr. Klaus Peter Koch. He received the Deutschland stipendium for excellent grades and engagement. During his master studies, he started working at Philips Research in Eindhoven, where he successfully worked on an algorithm for pneumonia detection. Subsequent, he got the Ph.D. position at the Technical University Eindhoven, Philips Research, and the Maxima Medical Center Veldhoven, working on the here presented topic under the supervision of Prof. Dr. Ronald Aarts, Dr. Md. Peter Andriessen, Dr. Louis Atallah, Dr. Marjolein van Lieshout, Dr. Nicol Heijnen, Dr. Johannes Weda, and Dr. Xi Long. In his time at Philips, he worked in three different departments of Philips Research for the department-heads Dr. Joerg Habetha, Richard Kemkes, Dr. Rogier Niessen, and PDEng. Elly Zwartkruis-Pelgrim. Jan holds two first US patent application filings and 10 Philips invention disclosures. During the Ph.D., he moved with his wife to Aachen and finally back to Mainz am Rhein, where he still lives with his wife and now three children. At the time, he is working at Phytec Messtechnik GmbH as a data scientist. Actual career information can be found at https://www.linkedin.com/in/jan-werth/