KU LEUVEN

ARENBERG DOCTORAL SCHOOL
Faculty of Engineering Science

Advanced solutions for neonatal sleep analysis and the effects of maturation

Ofelie De Wel

Supervisor:
Prof. dr. ir. S. Van Huffel
Co-supervisor:
Prof. dr. G. Naulaers

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

February 2020

Advanced solutions for neonatal sleep analysis and the effects of maturation

Ofelie DE WEL

Examination committee:
Prof. dr. ir. H. Hens, chair
Prof. dr. ir. S. Van Huffel, supervisor
Prof. dr. G. Naulaers, co-supervisor
Prof. dr. ir. L. De Lathauwer
Prof. dr. ir. M. De Vos
Prof. dr. K. Jansen
Prof. dr. ir. B. Puers
Prof. dr. ir. J. Suykens
Prof. dr. J. Dudink
(University Medical Center Utrecht)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

February 2020

© 2020 KU Leuven – Faculty of Engineering Science
Uitgegeven in eigen beheer, Ofelie De Wel, Kasteelpark Arenberg 10 box 2446, B-3001 Leuven (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm, electronic or any other means without written permission from the publisher.

Abstract

Worldwide approximately 11% of the babies are born before 37 weeks of gestation. The survival rates of these prematurely born infants have steadily increased during the last decades as a result of the technical and medical progress in the neonatal intensive care units (NICUs). The focus of the NICUs has therefore gradually evolved from increasing life chances to improving quality of life. In this respect, promoting and supporting optimal brain development is crucial. Because these neonates are born during a period of rapid growth and development of the brain, they are susceptible to brain damage and therefore vulnerable to adverse neurodevelopmental outcome. In order to identify patients at risk of long-term disabilities, close monitoring of the neurological function during the first critical weeks is a primary concern in the current NICUs.
Electroencephalography (EEG) is a valuable tool for continuous noninvasive brain monitoring at the bedside. The brain waves and patterns in the neonatal EEG provide interesting information about the newborn brain function. However, visual interpretation is a time-consuming and tedious task requiring expert knowledge. This indicates a need for automated analysis of the neonatal EEG characteristics. The work presented in this thesis aims at contributing to this.

The first part of this thesis focuses on the development of algorithms to automatically classify sleep stages in preterm babies. In total three different strategies are proposed. In the first method, the problem is traditionally approached and a new set of EEG complexity features is combined with a classification algorithm. This analysis demonstrates that the complexity of the EEG signal is fundamentally different dependent on the vigilance state of the infant. Building on this finding, a novel tensor-based approach that detects quiet sleep in an unsupervised manner is presented. Finally, a deep convolutional neural network to classify neonatal sleep stages is implemented. This end-to-end model optimizes the feature extraction and classification model simultaneously, avoiding the challenging task of feature engineering.

The second part concentrates on the quantification of functional brain maturation in preterm infants. We establish that the complexity of the EEG time series is significantly positively correlated with the postmenstrual age of the neonate. Moreover, these promising biomarkers of brain maturity are used to develop a brain-age model. This model can accurately estimate the infant's age and thereby assess the functional brain maturation. In addition, the relationship between the early functional and structural brain development is investigated based on two complementary neuromonitoring modalities, EEG and MRI. Regression models show that the brain activity during the first postnatal days is related to the size and growth of the cerebellum in the subsequent weeks. At last, the influence of the thyroid function on the developing brain is examined in extremely premature infants. No significant association was observed between the change in free thyroxine concentrations during the first week of life and maturational features extracted from the EEG at term equivalent age. To shed more light on the precise relationship between thyroid function and brain maturation, prospective studies with a more homogeneous dataset are needed in the future.

Beknopte samenvatting

Wereldwijd wordt ongeveer 11% van de baby's vóór 37 weken zwangerschap geboren. De overlevingskansen van deze premature baby's zijn de laatste decennia gestaag toegenomen als gevolg van de technische en medische vooruitgang op de dienst neonatale intensieve zorgen (NIC). De focus van de medische zorg bij vroeggeboren kinderen is daarom geleidelijk geëvolueerd van het verhogen van levenskansen naar het verbeteren van de levenskwaliteit. In dit opzicht is het bevorderen en ondersteunen van optimale hersenontwikkeling van cruciaal belang. Omdat deze baby's geboren worden tijdens een periode van snelle groei en ontwikkeling van de hersenen, zijn ze gevoelig voor hersenletsels en bijgevolg kwetsbaar voor neurologische ontwikkelingsachterstand. Om patiënten met verhoogd risico op beperkingen te identificeren, is nauwgezet toezicht op de neurologische functie tijdens de eerste kritieke weken van primair belang in de huidige NICs.
Elektro-encefalografie (EEG) is een nuttig instrument voor continue niet-invasieve hersenmonitoring tijdens het verblijf in de couveuse. De hersengolven en patronen in het neonatale EEG verschaffen interessante informatie over de hersenfunctie van pasgeborenen. Visuele interpretatie is echter een tijdrovende en eentonige taak die bovendien kennis van experts vereist. Er is dus behoefte aan geautomatiseerde analyse van de neonatale EEG kenmerken. Het werk dat in dit proefschrift wordt gepresenteerd wil hiertoe bijdragen.

Het eerste deel van deze thesis richt zich op de ontwikkeling van algoritmes om slaapstadia bij premature baby's automatisch te classificeren. Drie verschillende methodes worden uitgewerkt. In de eerste methode wordt het probleem traditioneel benaderd en wordt een reeks complexiteitskenmerken van het EEG gecombineerd met een classificatiemodel. Deze analyse toont aan dat de complexiteit van het EEG-signaal fundamenteel verschillend is afhankelijk van de slaapfase waarin de baby zich bevindt. Voortbouwend op deze vaststelling, wordt een nieuwe, tensor-gebaseerde aanpak voorgesteld die diepe slaap op een ongesuperviseerde manier detecteert. Tot slot wordt een diep convolutioneel

neuraal netwerk geïmplementeerd om neonatale slaapstadia van elkaar te onderscheiden. Dit model optimaliseert tegelijkertijd de extractie van de attributen en het classificatiemodel, waardoor de moeilijke taak van feature engineering overbodig wordt.

Het kwantificeren van de hersenontwikkeling bij premature baby's maakt het onderwerp uit van het tweede deel. We stellen vast dat de complexiteit van de EEG signalen significant positief gecorreleerd is met de postmenstruele leeftijd van de baby. Bovendien worden deze veelbelovende kenmerken gebruikt om een hersenleeftijdsmodel te ontwikkelen. Hiermee kan de leeftijd van de baby nauwkeurig geschat worden en is het dus mogelijk om de functionele hersenrijping te beoordelen. Daarnaast wordt de relatie tussen vroege functionele en structurele hersenontwikkeling onderzocht aan de hand van twee complementaire beeldvormingstechnieken, EEG en MRI. Regressiemodellen tonen aan dat de hersenactiviteit tijdens de eerste postnatale dagen verband houdt met de grootte en groei van het cerebellum in de daaropvolgende weken. Ten slotte wordt de invloed van de schildklierfunctie op de ontwikkelende hersenen bij extreem premature baby's onderzocht. Er werd geen significant verband waargenomen tussen de verandering in vrije thyroxine concentraties tijdens de eerste levensweek en de maturiteit van corticale activiteit op a terme leeftijd. Om meer licht te werpen op de precieze relatie tussen schildklierwerking en de hersenmaturatie, zijn er in de toekomst prospectieve studies nodig met een meer homogene dataset.

List of Abbreviations

ACF	Autocorrelation function
aEEG	Amplitude-integrated electroencephalography
ALS	Alternating least squares
ApEn	Approximate entropy
AS	Active sleep
AUC	Area under the curve
CFM	Cerebral function monitor
CLASS	Cluster-based Adaptive Sleep Staging
CNN	Convolutional neural network
Conv	Convolutional
CORCONDIA	Core consistency diagnostic
CPD	Canonical polyadic decomposition
CSF	Cerebrospinal fluid
cUS	Cranial ultrasound
DIFFIT	Difference of fit
DIO2	Type 2 deiodinase
DTI	Diffusion tensor imaging
DWI	Diffusion weighted imaging
ECG	Electrocardiogram
ECL	ElectroChemiLuminescence
EEG	Electroencephalography
ELGAN	Extremely low gestational age neonate
EMG	Electromyogram
FIR	Finite impulse response
fT4	Free thyroxine

GA	Gestational age
GM	Grey matter
HIE	Hypoxic ischaemic encephalopathy
HVS	High-voltage slow pattern
IBI	Interburst interval
ICA	Independent component analysis
IHF	Interhemispheric fissure
IS	Indeterminate sleep
ISI	Inter-SAT interval
LASSO	Least absolute shrinkage and selection operator
LME	Linear mixed-effects
LS-SVM	Least squares support vector machines
LSTM	Long-short term memory
MR	Magnetic resonance
MRI	Magnetic resonance imaging
MSE	Multiscale entropy
My	Myelinated white matter
NIC	Dienst neonatale intensieve zorgen
NICU	Neonatal intensive care unit
NIRS	Near-infrared spectroscopy
NLEO	Nonlinear energy operator
NLS	Nonlinear least squares
NQS	Non-quiet sleep
NREM	Non-rapid eye movement
PARAFAC	Parallel factor analysis
PCA	Principal component analysis
PD	Polyadic decomposition
PMA	Postmenstrual age
PSD	Power spectral density
PVL	Periventricular leukomalacia
QS	Quiet sleep
RBF	Radial basis function
ReLU	Rectified linear unit
REM	Rapid eye movement

RMSE	Root mean square error
ROC	Receiver operating characteristic
SAT	Spontaneous activity transient
SEF	Spectral edge frequency
SGD	Stochastic gradient descent
SGM	Subcortical grey matter
SVD	Singular value decomposition
SVM	Support vector machine
T3	Triiodothyronine
T4	Thyroxine
TA	Tracé alternant
TBV	Total brain volume
TD	Tracé discontinu
TEA	Term equivalent age
TH	Thyroid hormone
THOP	Transient hypothyroxinemia of prematurity
THRA	Thyroid hormone receptor alpha
THRB	Thyroid hormone receptor beta
TS	Transitional sleep
TSH	Thyroid-stimulating hormone
UWM	Unmyelinated white matter
WKZ	Wilhelmina Children's Hospital

List of Symbols

α	EEG frequency band: 8-12 Hz
β	EEG frequency band: 12-30 Hz
δ	EEG frequency band: 0.5-4 Hz
κ	Cohen's Kappa score
$\|.\|_F$	Frobenius norm
ρ	Pearson's correlation coefficient
τ	Scale factor
θ	EEG frequency band: 4-8 Hz
a,b,...	Scalar
m	Embedding dimension
r	Tolerance for sample entropy computation
$\mathbf{A}$,$\mathbf{B}$,...	Matrix
$\mathcal{A}$,$\mathcal{B}$,...	Tensor
$\mathbf{a}$,$\mathbf{b}$,...	Vector
R^2	Coefficient of determination

Contents

Abstract i

Beknopte samenvatting iii

List of Abbreviations vii

List of Symbols ix

Contents xi

List of Figures xix

List of Tables xxv

I Introduction 1

1 Introduction 3

1.1 Problem statement 3

1.2 Outline of the thesis 4

1.2.1 Part I: Introduction 5

1.2.2 Part II: Automated neonatal EEG sleep staging 5

1.2.3 Part III: Automated brain maturation quantification . . 6

1.2.4 Part IV: Conclusion 7
1.3 Collaborations 7
1.4 Conclusion 9

2 Physiological interpretation of the neonatal EEG 11
2.1 Preterm birth 11
2.1.1 Motivation 11
2.1.2 Age terminology 12
2.2 The neonatal brain 13
2.2.1 Building blocks of the brain 13
2.2.2 Early structural brain development 13
2.2.3 Monitoring of the developing brain in the NICU 16
2.3 The electroencephalogram of the newborn 18
2.3.1 Recording technique 18
2.3.2 aEEG 20
2.3.3 Patterns in the neonatal EEG 20
2.3.4 Artifacts 29
2.4 Conclusion 31

3 Mathematical background 33
3.1 EEG features 33
3.1.1 EEG continuity 33
3.1.2 Entropy of the EEG 34
3.1.3 Spectral features 39
3.2 Tensors 40
3.2.1 Multiway data 40
3.2.2 Notations and definitions 41
3.2.3 Canonical polyadic decomposition 42

3.3 Supervised learning 46
3.3.1 Support vector machines 47
3.3.2 Least squares support vector machines 51
3.3.3 Deep learning 52
3.3.4 Linear regression 55
3.4 Performance metrics 57
3.4.1 Classification 57
3.4.2 Regression 60
3.5 Conclusion 60

II Automated neonatal EEG sleep staging 61

4 Neonatal sleep stage classification based on EEG complexity features 63
4.1 Introduction 64
4.2 Materials and methods 65
4.2.1 Database 65
4.2.2 Preprocessing 67
4.2.3 Multiscale entropy computation 67
4.2.4 Feature extraction 68
4.2.5 Classification model and training procedure 68
4.3 Results 70
4.4 Discussion 70
4.5 Conclusion 72

5 CPD of a multiscale tensor for neonatal sleep stage identification 73
5.1 Introduction 74
5.2 Database 75

5.3 The proposed tensor-based sleep stage identification method . . 75
5.3.1 EEG preprocessing 76
5.3.2 Multiscale entropy computation 76
5.3.3 Tensorization 76
5.3.4 Tensor decomposition 77
5.3.5 Selection of the component of interest 80
5.3.6 Postprocessing and clustering 80
5.3.7 Classification Performance 82
5.3.8 Statistical Analysis 82
5.4 Results 84
5.5 Discussion 87
5.6 Conclusion 90

6 Quiet sleep detection in preterm infants using deep CNN 91
6.1 Introduction 92
6.2 Materials and methods 94
6.2.1 Database 94
6.2.2 The proposed CNN for sleep stage classification 94
6.2.3 Spectral feature based neonatal sleep stage classifier . . 96
6.2.4 Cluster-based Adaptive Sleep Staging (CLASS) 97
6.2.5 Classification performance 97
6.2.6 Error correlation 98
6.2.7 Computational time 98
6.3 Results 98
6.3.1 Feature evolution during sleep-wake cycling 98
6.3.2 Classification performance 99
6.3.3 Error correlation 100

6.3.4 Computational Time 102
6.4 Discussion 103
6.5 Conclusion 105

7 Comparison of neonatal sleep stage classification algorithms 107
7.1 Performance comparison 108
7.1.1 Database and performance evaluation 108
7.1.2 Results 109
7.1.3 Statistical analysis 109
7.1.4 Discussion 110
7.2 Maturational effect 110
7.3 Generalizability 112
7.4 Computational time 114
7.5 Conclusion 115

III Automated brain maturation quantification 117

8 Assessing brain maturation in preterm infants using EEG complexity features 119
8.1 Introduction 120
8.2 Materials and methods 121
8.2.1 Database 121
8.2.2 EEG preprocessing 121
8.2.3 Multiscale entropy 122
8.2.4 Feature extraction 122
8.2.5 Correlation and linear regression analysis 123
8.2.6 Topological analysis 124
8.3 Results 126

8.3.1 Correlation and linear regression analysis 126
8.3.2 Topological analysis 126
8.4 Discussion 128
8.5 Conclusion 130

9 Relationship between early functional and structural brain development in preterm infants 131

9.1 Introduction 132
9.2 Materials and methods 134
9.2.1 Database 134
9.2.2 Preprocessing and feature extraction 135
9.2.3 Brain maturation quantification using EEG features . . 140
9.2.4 Relationship between early brain function and structure 141
9.3 Results 143
9.3.1 Brain maturation quantification using EEG features . . 143
9.3.2 Relationship between early brain function and structure 145
9.4 Discussion 148
9.5 Conclusion 151

10 Measurement of thyroid hormone action in the preterm infants' brain using EEG 153

10.1 Introduction 154
10.2 Materials and Methods 155
10.2.1 Database 155
10.2.2 Thyroid hormone function 157
10.2.3 Automated EEG analysis 158
10.3 Results 160
10.3.1 Thyroid hormone function 160

10.3.2 Automated EEG analysis 161
10.4 Discussion 162
10.5 Conclusion 166

IV Conclusion 167

11 Conclusions and Future directions 169
11.1 Conclusions 170
11.1.1 Automated EEG sleep staging 170
11.1.2 Automated brain maturation quantification 171
11.2 Future directions 172
11.2.1 Automated EEG sleep staging 172
11.2.2 Automated brain maturation quantification 178

A Performance of algorithms for automated EEG sleep staging in preterm infants 181

Bibliography 189

Curriculum vitae 211

List of publications 213

List of Figures

1.1 Schematic overview of the structure of the thesis. 8

2.1 Age terminology in preterm infants. 12

2.2 Schematic diagram of a neuron and synapse [145]. 14

2.3 Early structural brain development [145]. 15

2.4 Electrode placement according to the international 10-20 system [126]. 19

2.5 Schematic overview of maturational EEG features. 22

2.6 Example of the EEG, ECG, respiration (Resp) and electrooculogram (EOG) during quiet sleep and active sleep in a preterm infant (PMA: 32 weeks). 24

2.7 Examples of the EEG, ECG, respiration (Resp) and electrooculogram (EOG) during quiet sleep in a term infant. 25

2.8 Examples of the EEG, ECG, respiration (Resp) and electrooculogram (EOG) during active sleep in a term infant. 26

2.9 Illustration of the most common EEG graphoelements [183]. . . 28

3.1 Example of a discontinuous EEG segment with indicated bursts and interburst intervals. 34

3.2 (a) Entropy and complexity versus randomness of the time series [216]. (b) Multiscale entropy curves for white gaussian noise, pink noise, a sine wave and a neonatal EEG segment. 37

3.3 Illustration of the computation of multiscale entropy. (a) Coarse-graining procedure. (b) Sample entropy computation. 38

3.4 Power spectral density of a 30 s EEG segment with indicated delta, theta, alpha and beta frequency band. 40

3.5 A scalar x, vector $\mathbf{x}$, matrix $\mathbf{X}$ and tensor $\mathcal{X}$. 40

3.6 The canonical polyadic decomposition of a third-order tensor. . 43

3.7 Illustration of (a) underfitting, (b) appropriate fitting and (c) overfitting. 47

3.8 Optimal hyperplane of a support vector machine. 48

3.9 Schematic illustration of the convolution operation in CNN. . . 54

3.10 Illustration of a simple linear regression. 57

4.1 Example of a labelled non-quiet sleep and quiet sleep segment. 66

4.2 (a) Window length optimization for multiscale entropy computation. (b) Multiscale entropy of original EEG time series and its randomly shuffled surrogate. 69

4.3 (a) The ROC curve of the neonatal sleep stage classifier based on complexity features assessed on the complete test set. (b) The ROC curves showing the performance for the three age groups separately. 71

5.1 Multiscale entropy of quiet sleep versus non-quiet sleep. 77

5.2 The rank-R polyadic decomposition of a multiscale entropy tensor $\mathcal{T}$. 78

5.3 Illustration of the automated selection of the temporal signature of interest. 81

5.4 Illustration of the postprocessing and clustering of the temporal signature for a rank-1 (a) and rank-2 (b) decomposition of the multiscale entropy tensor. 83

5.5 The area under the ROC curve as a function of the postmenstrual age at the moment of the recording. 86

5.6 Boxplots of the area under absolute value of autocorrelation of the temporal signatures after sorting them in descending order according to their Kappa score. 88

6.1 Architecture of the convolutional neural network. 94

6.2 Features derived by the CNN during sleep cycling. 99

6.3 ROC curves for the CNN sleep stage classifier. 100

6.4 The histogram of the training data segments is displayed in light grey (left y-axis). The blue circles show the AUC for each recording from the test set (right y-axis). 101

6.5 The error correlation between the CNN and two existing sleep stage classifiers. 102

6.6 The average computational time for 2 h multichannel EEG segments for the CNN, the CLASS algorithm and the feature-based approach. 103

7.1 The AUC for each test recording using the three proposed algorithms for neonatal sleep stage classification. 111

7.2 Performance of the CNN sleep stage classifier on the complete test set for one missing electrode. 113

7.3 Performance of the CNN sleep stage classifier on the complete test set for a reduced montage. 113

8.1 Multiscale entropy curves of EEG recordings measured between 29 and 39 weeks postmenstrual age. The multiscale entropy curve shifts upwards with increasing PMA. 123

8.2 (a) The relationship between the complexity index of channel T3 and the postmenstrual age (PMA) fitted by simple linear regression. (b) Boxplots of the complexity index averaged over all channels for both quiet sleep and non-quiet sleep. A clear increase of electroencephalogram (EEG) complexity can be observed in both sleep stages. 127

8.3 The topoplot of the grand average of the complexity index during (a) quiet sleep (QS) and (b) non-quiet sleep (NQS). 129

9.1 Visualization of the dataset used to investigate the relationship between early functional and structural brain development in preterm infants. 134

9.2 Illustration of detection of spontaneous activity transient and interburst intervals in the preterm EEG. 139

9.3 Examples of an automatically segmented MRI at (a) 30 weeks PMA and (b) 40 weeks PMA. 140

9.4 The fixed effects of the regression model fitting the relationship between SAT% and PMA are shown on the left. The regression model on the right shows the association between the complexity index and the PMA. 143

10.1 Fetal brain development in relation to maternal thyroid hormone supply and fetal thyroid hormone metabolism. 156

10.2 (a) Boxplot of the gestational age of infants with positive ΔfT4 and negative ΔfT4. (b) Comparison of fT4 levels at day of birth, at the end of the first week of life and the difference between the two measurements for the patients with positive ΔfT4 versus patients with negative ΔfT4. 162

10.3 Postmenstrual age at which EEG time series were measured versus the change in free thyroxine concentration during the first week of life. 165

11.1 Temporal signature obtained by the rank-1 CPD brain connectivity tensor showing increased functional connectivity during quiet sleep. 175

11.2 Temporal signature obtained by the rank-1 CPD brain connectivity tensor not showing a clear relation to the neonatal sleep stages. 176

11.3 Regularity of the respiration signal versus clinically labelled quiet sleep segments. 177

11.4 The area under the ROC curves constructed based on the respiration regularity and the clinical sleep labels as a function of PMA. 177

11.5 (a) The boxplot shows the average complexity index among the complete recording length and all channels for HIE neonates with good and poor outcome. (b) The boxplots illustrate the difference in complexity index across different severity grades of background activity. 180

List of Tables

5.1 The classification performance of the proposed tensor-based method for different values of the rank R. 85

6.1 Layers of the designed network. 95

6.2 Overview of the classification performance of the proposed CNN, the CLASS algorithm and the feature-based approach with and without postprocessing step. 101

7.1 Performance comparison of various algorithms for automated sleep stage classification in preterm infants. 109

7.2 Overview of the pros and cons of the proposed preterm sleep stage classification algorithms. 115

8.1 Results of correlation and regression analysis investigating the association between EEG complexity features and postmenstrual age. 125

9.1 Overview of maturational features extracted from EEG recordings and developmental and injury measures extracted from magnetic resonance images. 136

9.2 Results of the correlation and regression analysis performed to assess the relationship between the maturational EEG features and the postmenstrual age at the moment of the recording. . . 144

9.3 The relationship between the change in maturational EEG feature and the significantly correlated MRI metrics at term equivalent age. 146

9.4 The relationship between early brain activity and features extracted from subsequent MRI recordings. 147

10.1 Patient characteristics 161

10.2 Results of the mixed-effects model. For each maturational EEG feature, the regression coefficient indicating the association between the EEG feature and delta fT4 is set out. Moreover, the confidence interval of the regression coefficient and its p-value is presented. 163

A.1 Performance of the complexity feature-based algorithm on each of the test recordings. 182

A.2 Performance of the tensor-based algorithm on each of the test recordings. 183

A.3 Performance of the CNN sleep stage classifier on each of the test recordings. 184

A.4 Performance of the cluster-based adaptive sleep staging (CLASS) algorithm on each of the test recordings. 185

A.5 Performance of the spectral feature-based algorithm without postprocessing on each of the test recordings. 186

A.6 Performance of the spectral feature-based algorithm with postprocessing on each of the test recordings. 187

Part I

Introduction

Chapter 1

Introduction

1.1 Problem statement

The human brain is a complex system composed of 86 billion interacting neurons [81]. A vast amount of its growth and development takes place before birth. However, when a baby is born too early, this natural process of rapid fetal brain development is suddenly interrupted. Hence, these neonates are born with an immature central nervous system. Therefore, an important part of their brain development has to take place in the noisy neonatal intensive care unit instead of the safe environment of the mother's womb. As a consequence, it comes as no surprise that early birth can result in far-reaching consequences. Preterm infants often face serious neurodevelopmental challenges and their long-term health prospects are strongly dependent on the perinatal care provided in neonatal intensive care unit (NICU). Therefore, close monitoring of these vulnerable newborns, especially their brains, during the first critical weeks is of utmost importance.

The electroencephalogram (EEG) is a non-invasive and cheap tool to monitor the electrical activity of the brain. Continuous electroencephalography is commonly used to monitor the brain function of newborns in need of intensive care at the cot-side. It provides valuable information about the neonate's brain development and can assist in early detection and assessment of cerebral abnormalities. Moreover, continuous neuromonitoring can guide optimal neurological care and can be used to predict the infant's prognosis.

Ideally, the electrocortical activity of these vulnerable neonates is continuously monitored. However, visual interpretation of the complex patterns in the

neonatal EEG is challenging and requires expertise. Moreover, reading long-term EEG recordings is a tedious and time-consuming task for the neonatal electrophysiologist, who is not always available in the NICU. Therefore, there is an urgent need for reliable automated analysis of the EEG to monitor the developing preterm brain.

A considerable amount of literature has focused on the development of algorithms for automated neonatal seizure detection or assessment of the background patterns in patients with hypoxic ischaemic encephalopathy (HIE). However, as mentioned above, also in patients without severe pathologies continuous brain monitoring is of added value. During the last decades, there is an increased interest in objective methods to quantify brain maturation in preterm infants [54, 142]. Similar to the existing growth charts used to track the weight and height of newborns, models to monitor the brain maturation of preterm neonates have been developed [150, 179]. In addition to giving an estimation of the maturational age of the infant, these models attempt to gain insight in the physiological processes taking place during early brain maturation. Previous studies performing automated quantification of brain maturation have mainly focussed on the increase of EEG continuity during ageing [54, 150, 179]. Moreover, features characterizing the spectral content of the EEG time series are often assessed to evaluate the early brain development. In addition to characteristics of specific waves and patterns observed in the EEG, the organization of the sleep-wake cycling of preterm infants also carries important information about the functional brain integrity [56].

In this dissertation, fully automated algorithms to monitor the early brain development in preterm infants are proposed. Briefly, the aim of this thesis is two-fold. On the one hand, different methodologies to classify EEG sleep stages in preterm infants have been investigated. On the other hand, we have looked at other aspects of the EEG, more specifically the complexity of the electrocortical recordings, to quantify the maturity of the neonatal brain. It is hoped that these automated algorithms will further improve the monitoring of these vulnerable infants, allowing early therapeutic intervention and in this way improve their clinical outcome.

1.2 Outline of the thesis

The overall structure of this manuscript takes the form of four main parts. The first part is composed of two introductory chapters, dealing respectively with the physiological and mathematical background necessary to understand the following chapters. The second part of the thesis comprises four chapters, each of

the first three propose a different approach to perform automated sleep staging in preterm infants, while the fourth one compares and discusses the presented algorithms. The third part focuses on quantification of brain maturation in neonates. At last, the findings of the research are presented and suggestions for further studies are given. A more detailed description of each chapter can be found below. Moreover, Figure 1.1 presents a graphical representation of the structure of the thesis.

1.2.1 Part I: Introduction

Chapter 1 is the current chapter, introducing the problem statement and the research objectives of the thesis. Moreover, the collaborations that led to the findings presented in this thesis are described.

Chapter 2 provides an overview of the neurophysiological background of the neonatal EEG. First, the building blocks and the early structural development of the brain is set out. Next, the most commonly used neuroimaging and neuromonitoring techniques are described. Finally, the recording technique and the maturational patterns observed in the neonatal EEG are thoroughly discussed.

Chapter 3 is concerned with a comprehensive background of the mathematical methodologies and signal processing tools used in the remainder of the thesis. First, the maturational features typically extracted from the EEG are described. Second, tensors and their decompositions are briefly introduced. Third, different strategies for supervised learning, including support vector machines, convolutional neural networks and regression models, are described. At last, metrics to assess the algorithm's performance are listed.

1.2.2 Part II: Automated neonatal EEG sleep staging

Chapter 4 presents an automated algorithm for sleep stage classification in a preterm cohort based on EEG complexity. A fixed-size LS-SVM classifier trained using features reflecting the complexity of the EEG signal is able to discriminate quiet sleep from non-quiet sleep in a wide PMA range.

Chapter 5 exploits tensor algebra to identify neonatal EEG sleep stages. Based on the findings of Chapter 4, the EEG complexity is used to tensorize the EEG

time series. By decomposing a multiscale entropy tensor, a reliable estimate of the sleep cycling can be obtained. This data-driven approach avoids the need to train a machine learning algorithm. Hence, this unsupervised approach can be easily applied in other clinical centers.

Chapter 6 adopts deep learning to construct a robust automated algorithm to detect quiet sleep in preterm infants. This end-to-end learning approach reaches a high performance and has a low computational time, making it suitable for real-time sleep staging in clinical practice.

Chaper 7 compares the performance and properties of the three aforementioned algorithms for neonatal sleep stage identification on the same test set. Moreover, the advantages and disadvantages of each of the methods are set out.

1.2.3 Part III: Automated brain maturation quantification

Chapter 8 proposes a brain-age model, estimating the age of the neonate based on complexity features extracted from the EEG signal. This model can be used to assess the brain maturation of the infant by computing the deviation from the real postmenstrual age.

Chapter 9 presents an exploratory study investigating the relationship between the function and structure of the developing premature brain. In order to examine this association, maturational features are extracted from EEG recordings measured immediately after birth and during the first postnatal weeks. A correlation and regression analysis is then used to evaluate the relationship between EEG features assessing the brain function and structural metrics extracted from MRI recordings measured around 30 weeks and 40 weeks postmenstrual age.

Chapter 10 investigates whether poor thyroid function in preterm infants is reflected by abnormal brain maturation in EEG recordings measured at term equivalent age. The change in the free thyroxine concentration during the first week of life is measured to assess the thyroid function of the neonate, while the brain function is assessed via a set of maturational features extracted from the EEG.

1.2.4 Part IV: Conclusion

Chapter 11 summarizes the main findings of the research presented in this thesis. Moreover, it includes a discussion of the implication of the findings to future research into this area.

1.3 Collaborations

The research presented in this thesis was carried out in the Biomed research group[1] under the supervision of Prof. Sabine Van Huffel. However, collaborations with several people played an important role in the development of the algorithms proposed in the following chapters. An overview of the different collaborations set up during my PhD will be given below.

First of all, the idea of the complexity analysis of the neonatal EEG with applications in both sleep stage classification (Chapter 4) and brain maturation quantification (Chapter 8) resulted from discussions with Mario Lavanga. In general, together with Prof. Alexander Caicedo, he provided essential suggestions and feedback during numerous brainstorming sessions.

The work presented in Chapter 5 contributes to the ERC Advanced Grant BIOTENSORS (no. 339804). During the last decade, several studies have proven that tensor decompositions are useful in neonatal brain function monitoring. A novel tensor-based approach to localize neonatal seizures was proposed by Deburchgraeve et al. [53]. Moreover, an objective algorithm to automatically assess the EEG background pattern in neonates with hypoxic ischaemic encephalopathy based on higher order discriminant analysis has been presented in [129]. Our group has also adopted a canonical polyadic decomposition (CPD) updating algorithm for the monitoring of brain haemodynamics in neonates [26]. This thesis continues this promising research line and illustrates how the polyadic decomposition of a multiscale entropy tensor can be used to identify neonatal sleep stages (Chapter 5).

The deep learning approach for neonatal sleep stage classification presented in Chapter 6 is the result of a close collaboration with Dr. Amir Hossein Ansari. We thoroughly discussed the decisions on the network architecture and the setup of the study. Moreover, we both contributed to the implementation of the algorithms, the data analysis and the interpretation of the results.

[1] Biomedical data processing research team, division of STADIUS, center of dynamical systems, signal processing and data analytics, Department of Electrical Engineering (ESAT), KU Leuven, Belgium

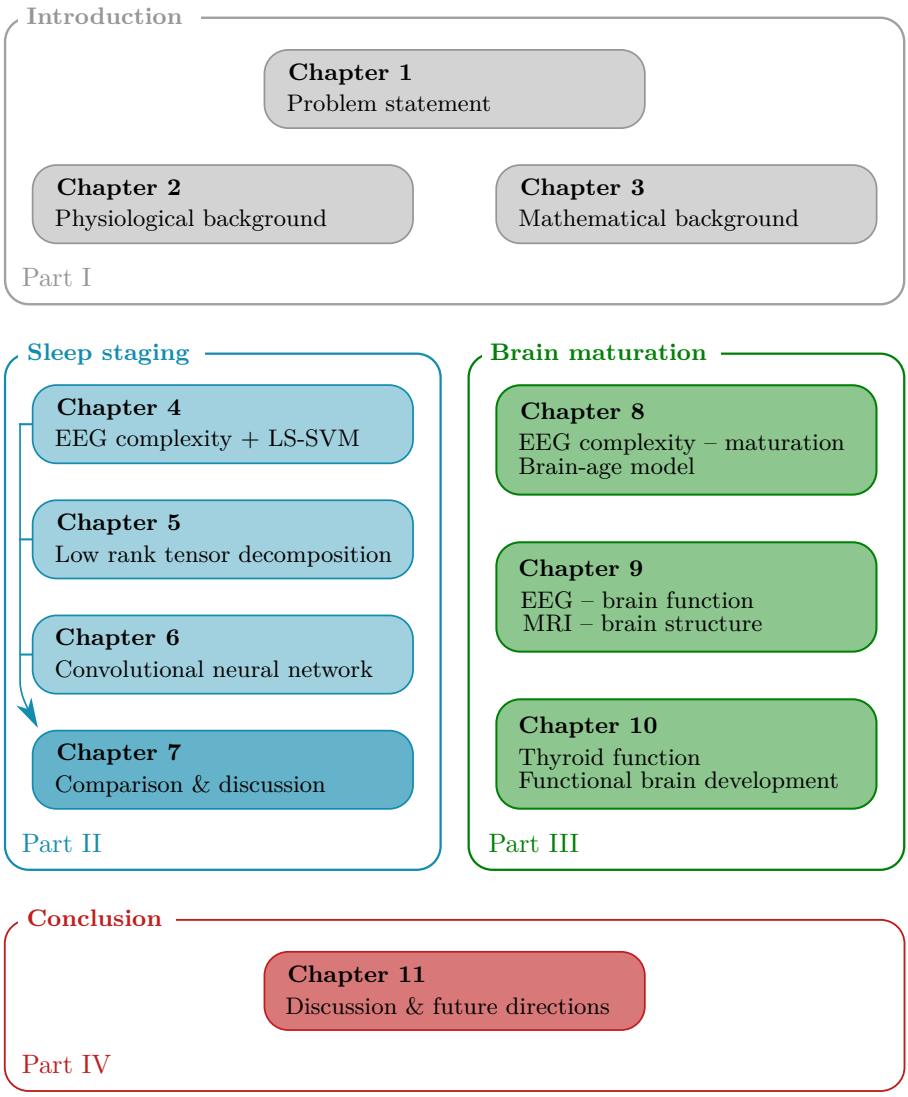


Figure 1.1: Chapter-by-chapter overview of the structure of the thesis. The blue and green boxes represent the contributions of this thesis in the domain of automated EEG sleep staging and brain maturation quantification, respectively.

The research was conducted in strong collaboration with the Department of Neonatalogy from the University Hospitals Leuven (Prof. Gunnar Naulaers, Prof. Katrien Jansen, Dr. Anneleen Dereymaeker, Dr. An Eerdekens and Jan Vervisch). The database of high quality multichannel EEG used to develop the sleep staging algorithms (Chapter 4 – 6) and to build the regression model in Chapter 8 was recorded in UZ Leuven. Prof. Katrien Jansen and Dr. Anneleen Dereymaeker visually annotated the quiet sleep periods in this database. Dr. An Eerdekens provided the patient characteristics, change in free thyroxine levels during the first week of life and the EEG data of the database used in Chapter 10. Moreover, all clinicians helped with the clinical interpretation of the results and gave important feedback during the biweekly neonatal meetings.

Finally, an international collaboration was established with Dr. Maria Luisa Tataranno associated to the Department of Neonatalogy, UMC Utrecht/Wilhelmina Children's Hospital, the Netherlands. She provided the database of EEG recordings and the MRI metrics used to study the relationship between early functional and structural brain development as described in Chapter 9.

1.4 Conclusion

This chapter introduced the problem statement and presented the two main research objectives of the thesis: automated neonatal EEG sleep staging and assessment of brain maturation. Moreover, a chapter-by-chapter overview of the structure of the dissertation is presented. At last, the collaborations leading to the results presented in this thesis are described.

Chapter 2

Physiological interpretation of the neonatal EEG

This chapter provides a brief overview of the physiological background and terminology required to follow the remainder of the thesis. It begins by explaining the physiology and early structural development of the neonatal brain. It will then go on to the most common neuroimaging and neuromonitoring techniques used in the current neonatal intensive care units. At last, EEG, the modality analysed in this thesis, will be described in more detail. The acquisition and interpretation of neonatal EEG will be thoroughly discussed.

2.1 Preterm birth

2.1.1 Motivation

Preterm birth is defined as a delivery before 37 completed weeks of gestation. Worldwide approximately 11% of all babies are born too early, accounting for around 15 million babies every year [20]. The rates of prematurity are rising and its complications are among the leading causes of mortality in children below the age of 5 [122]. Although the remarkable medical advancements and the sophisticated neonatal intensive care units (NICUs) have lead to increased survival rates of these neonates, they are still at an increased risk of brain injury and neurodevelopmental problems [36, 166]. The most common long-term neurodisabilities associated with preterm birth are cerebral palsy, hearing and

visual impairments, learning difficulties and psychological problems. These neurobehavioural sequelae of premature birth have a major impact on the lives of these patients and their families, and are one of the major public health issues in the current society. Therefore, during the last decades the focus has shifted from improving survival towards cot-side brain monitoring and neuroprotection of these vulnerable infants [146].

2.1.2 Age terminology

Specific terminology is used to refer to the age of preterm infants as shown in Figure 2.1 [64]. Gestational age (GA) is defined as the age from the first day of the last menstrual period of the mother until the delivery, while the chronological age refers to the time elapsed from birth onwards. The postmenstrual age (PMA) is the sum of the gestational and chronological age. The corrected age is the chronological age taken into account the time period born before 40 weeks gestation [64].
Based on the gestational age of the neonate, they can be classified into different groups. This is also indicated by the colours in Figure 2.1. Extremely preterm infants are born before 28 weeks of gestation, very preterm between 28 to 32 weeks of gestation and moderate to late preterm neonates from 32 to 37 weeks of pregnancy. The gestational age is a major determinant of the clinical outcome. The earlier the baby is born, the higher the risk for severe disabilities.

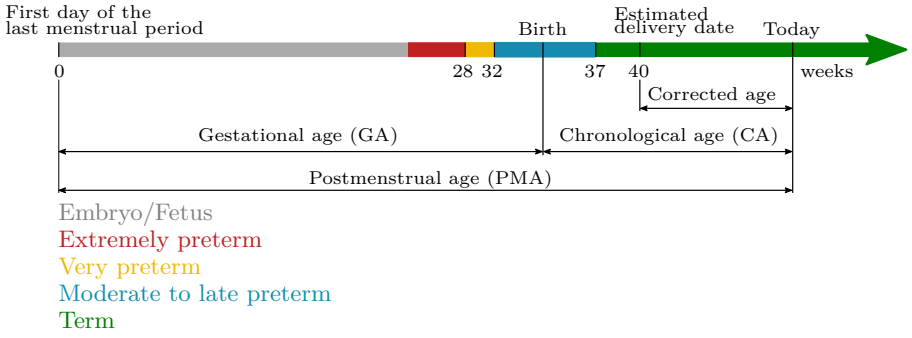


Figure 2.1: Age terminology in preterm infants.

2.2 The neonatal brain

2.2.1 Building blocks of the brain

Neurons are the information processing units of the brain. They consist of a cell body and are able to connect with each other via dendrites and axons. The former are short tree-like structures, which receive information from other neurons. While the latter is a longer nerve fiber conducting electrical impulses, called action potentials, away from the cell body. Axons are covered in a fatty substance called myelin, which enhances its conduction velocity. The axon of a neuron can connect to the dendrites of another neuron via a specialized connection known as the synapse. At the synapse, an action potential will trigger the presynaptic neuron to release chemical neurotransmitters which will bind to the receptors of the target postsynaptic neuron [145]. During development, a vast amount of interconnections between neurons are made resulting in extensive, well-connected neural networks [182]. A schematic diagram of a neuron and its synapse is shown in Figure 2.2.

2.2.2 Early structural brain development

The human brain is the most complex organ and grows enormously during gestation. The development of the human brain is a highly ordered and protracted process starting shortly after conception and continuing well into adolescence [82].
The first step of brain development is taken, when gastrulation occurs and epiblast cells differentiate into different type of stem cells which play a major role in embryonic development. Among these stems cells are the neural stem cells, also called the neural progenitor cells, which are able to produce all cells part of the central nervous system [182]. The region of the embryo containing the neural progenitor cells, the neural plate, is the basis of the nervous system. During a process called neurulation, this flat plate will fold and close to form a cylindrical neural tube (formed during the third week of gestation). This hollow tube is the first neural structure and the precursor of the central nervous system [182]. The neural stem cells in the anterior part of the neural tube will give rise to the brain, while the posterior part of this tube will later-on form the spinal cord and the hindbrain. As the neural tube closes, it will develop bulges and bends and will undergo segmentation. By the fourth week of gestation, the neural tube will have formed three primary brain vesicles: the prosencephalon (forebrain), mesencephalon (midbrain) and rhombencephalon (hindbrain). As the brain develops, two regions will further subdivide, finally resulting in five secondary brain vesicles which will develop into specialized brain structures

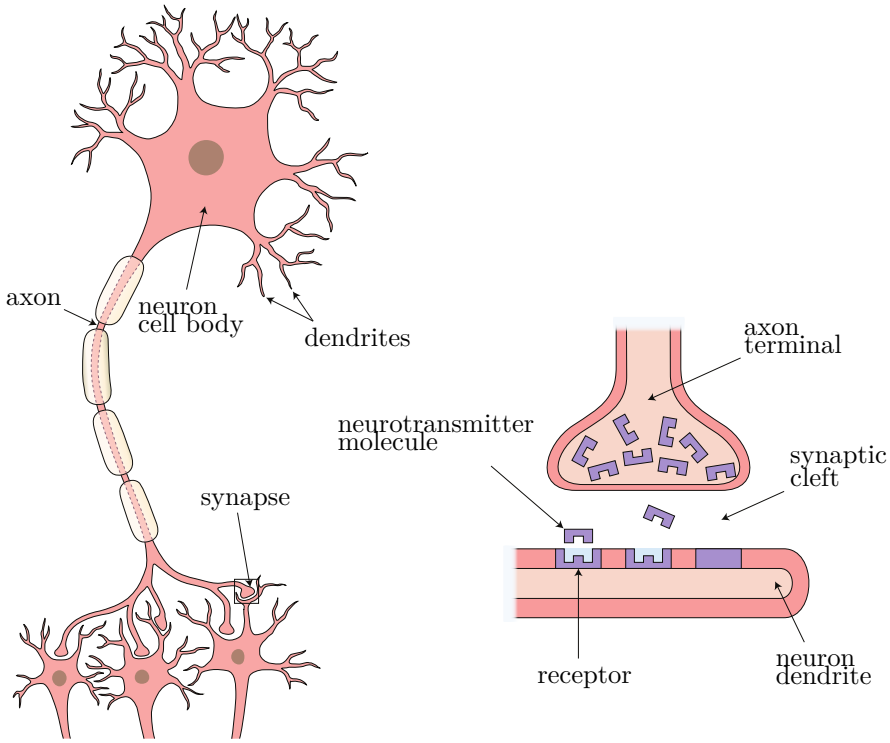


Figure 2.2: Schematic diagram of a neuron and synapse. Adapted from [145].

[182, 188]. By the end of the embryonic period, gradients of signalling molecules along the anterior-posterior and dorso-ventral axes of the neural tube initiate neural patterning of the neocortex. As a consequence, a primitive organisation of sensorimotor regions in the neocortex is already established after 8 weeks of gestation [182].

During the fetal period, so from 9 weeks of gestation up to birth, four important processes occur in the brain development: neural proliferation, migration and differentiation and cell death. The mature brain consists of around 86 billion neurons and the majority of them are produced prenatally. After neurulation the number of neurons starts to increase drastically. Up to day 42 of gestation, the neural progenitor cells divide symmetrically resulting in two neural progenitor cells. From day 42 until midgestation, the division of the neural stem cell becomes asymmetric and results in one neural progenitor cell and one neuron.

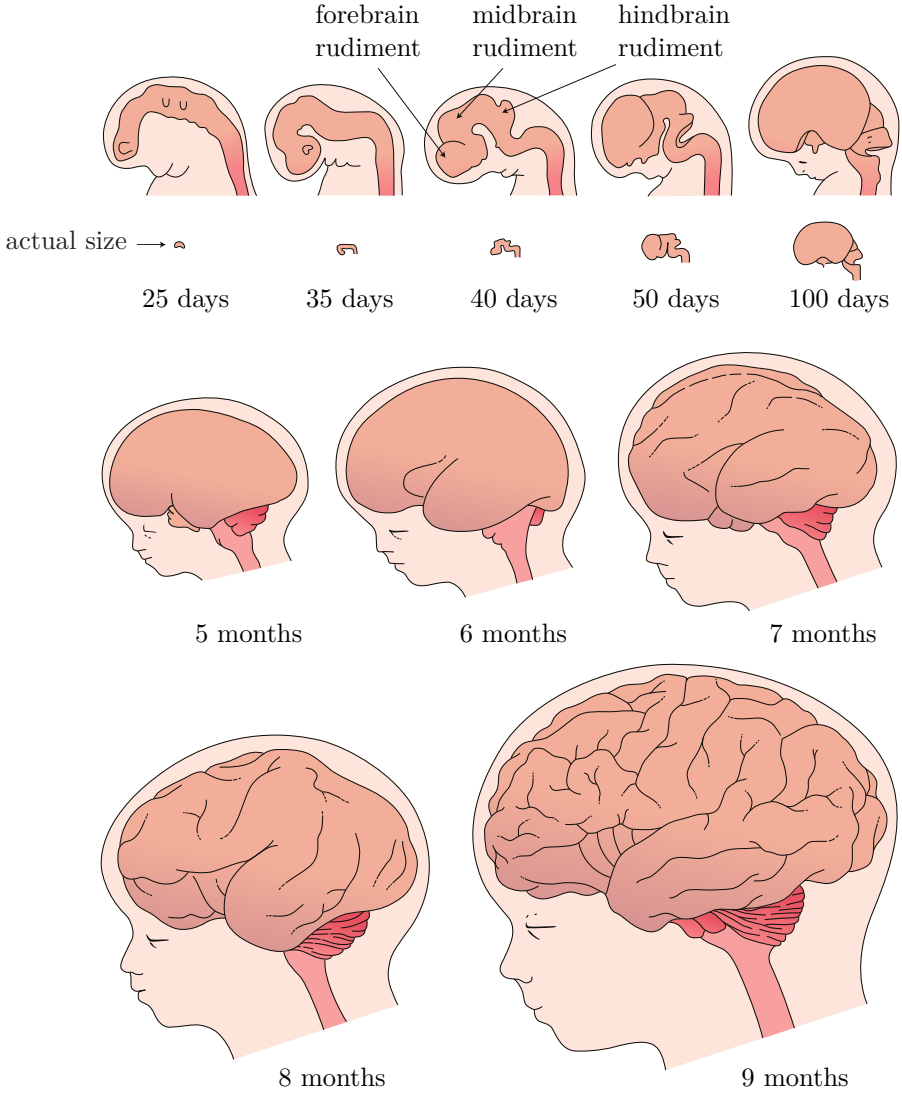


Figure 2.3: Early structural brain development. Adapted from [145].

At the peak of neurogenesis, neurons are produced at a rate of 250,000 neurons per minute [145]. These newly produced neurons will migrate outwardly to their final positions in the cortex and differentiate into a specific type of neuron. Once the neurons have arrived at their target position, they will extend dendrites and axons and form new synaptic connections with other neurons (synaptogenesis). In this way, information-processing neural networks are formed [182]. Next to neural proliferation, part of the neuron population will also be eliminated due to (programmed) cell death.

In addition to the microscopic changes in the fetal brain, the morphology also alters quickly. The most striking macroscopic change is the rapid increase of the brain's volume and mass. MRI studies have demonstrated that the volume of the premature brain even triples during the last trimester of the pregnancy [34]. Concurrently, the smooth cerebral surface progressively changes into a highly convoluted structure consisting of gyri and sulci. This cortical folding is an ordered and complex process, leading to a drastic increase of the cortical surface.

As mentioned before, brain development is not finished by the time of birth and continues postnatally. Myelination and synaptic pruning (selective elimination of neural connections) are the two main events that continue in the postnatal period.

2.2.3 Monitoring of the developing brain in the NICU

From the section above, it is clear that the early development of the neonatal brain is a highly regulated sequence of events. It is obvious that a disturbance of this process can have catastrophic consequences. Depending on the timing, different brain dysfunctions can arise. Therefore, the measurement of vital signs, such as heart rate, breathing pattern, and blood pressure, is often accompanied by brain monitoring tools in the current NICU setting. Various techniques to monitor the neonatal brain have been developed. Dependent on the condition of the patient or suspected injury another technique might be recommended. In the next sections, the brain imaging modalities most commonly used in the NICU will be briefly explained.

Structural neuroimaging

Cranial ultrasound (cUS) is part of the routine neurological monitoring in the NICU due to the fact that it is a safe, portable, relatively low-cost and fast procedure which can be performed at the bedside [77, 146, 199]. As a result, it is

the initial diagnostic modality used in the NICU. This tool is traditionally used to detect and track the evolution of peri- and intraventricular haemorrhages, hydrocephalus and periventricular leukomalacia (PVL, a common white matter brain injury in preterm infants caused by a shortage of oxygen and blood supply) [199].
Magnetic resonance imaging (MRI) can also be used to visualize structural lesions and has a higher spatial resolution, and thus greater sensitivity for injury detection compared to cUS. Therefore, this complementary imaging technique is often used to confirm the presence, exact location and extent of the lesions found using cUS [199]. Even though MRI also does not employ ionising radiation, it is a more expensive and challenging imaging technique often requiring transportation and sedation of the neonate. Due to these safety issues, it is only suitable for medically stable infants. Recently, more advanced, diffusion-based MRI techniques have emerged in the NICU. Diffusion-weighted imaging (DWI), including diffusion tensor imaging (DTI), can be used to evaluate the microstructure, integrity and fiber orientation of the white matter tracts [163]. Nevertheless, these advanced tools are not widely adopted in clinical practice and remain mainly experimental in the preterm population.

Functional neuromonitoring

While cUS and MRI are used to quantify the structure of the developing brain, near-infrared spectroscopy (NIRS) and the electroencephalogram (EEG) are functional neuromonitoring techniques. In contrast to the structural imaging techniques explained before, these tools also allow continuous monitoring of the brain function.
Near-infrared spectroscopy can noninvasively measure the cerebral tissue oxygenation. The technique relies on two basic physical principles: 1) the relative transparency of biological tissue (especially neonatal brain tissue) to light in the near-infrared range, and 2) the oxygen-dependent light absorption properties of haemoglobin [58, 215]. This harmless and painless procedure is of great clinical value since many neonatal brain pathologies are associated with poor cerebral oxygenation or haemodynamics. NIRS monitoring is indicated in patients with hypoxic ischaemic encephalopathy (HIE), haemodynamically relevant patent ductus arteriosis, unstable or low blood pressure, neonates receiving respiratory support and infants at risk of impaired autoregulation [58, 215].
At last, the electroencephalogram (EEG) provides a multichannel recording of the electrocortical activity. Due to the ease of electrode application and interpretation, amplitude-integrated electroencephalography (aEEG), rather than conventional EEG, is utilized in many centers. aEEG is a filtered and time-compressed version of the EEG measured by only two up to four scalp

electrodes. Continuous registration of the EEG or aEEG is the gold standard in diagnosis of (subclinical) seizure activity, the assessment of the background activity (e.g. after a hypoxic ischaemic insult) and evaluation of the sleep-wake cycling [30, 79, 80]. Like NIRS, EEG/aEEG is an affordable noninvasive neuromonitoring tool, feasible at the bedside of the patient.

2.3 The electroencephalogram of the newborn

The electroencephalogram provides a measurement of the electrical activity of the cerebral cortex via electrodes attached to the scalp. More specifically, it records postsynaptic potentials generated by large populations of similarly oriented active cortical pyramidal neurons close to the scalp electrode [21]. The signal is severely attenuated because it has to propagate through different brain tissue layers, such as the cerebrospinal fluid, the scalp and the skull. As a consequence, it has to be strongly amplified for display purposes. Moreover, this volume conduction together with the limited number of scalp electrodes deteriorates the spatial resolution of the EEG. In contrast to the poor spatial resolution, the temporal resolution of the EEG is excellent and even rapidly changing patterns can be captured.

2.3.1 Recording technique

Electrode setup

The conventional EEG is a multichannel recording, where each EEG channel represents the potential difference between two electrode recording sites. In general, two distinct types of montages can be distinguished. On the one hand, a bipolar montage where the voltage difference between two scalp electrodes is measured. On the other hand, a referential montage where one common reference is used. This reference can be either a scalp electrode, typically the vertex electrode (Cz), nose tip, linked mastoids or ears, or the average activity among all leads [21]. The scalp electrodes are placed according to the standard international 10-20 system, which is illustrated in Figure 2.4. The "10-20" refers to distance between neighbouring electrodes which equals either 10% or 20% of the total distance from nasion (bridge of the nose) to inion (bump at the back of the skull). This system based on anatomical landmarks leads to a consistent placement of the electrodes, ensures that all brain regions are covered and allows comparison of different EEG measurements (e.g. from different subjects or at different recording times). Besides, every electrode position is

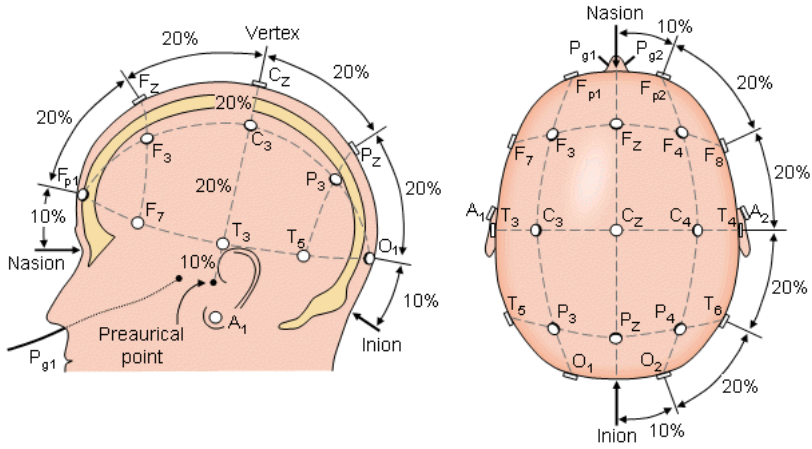


Figure 2.4: Electrode placement according to the international 10-20 system. Reprinted from [126].

represented by a letter and a number. The letter refers to the brain area it overlies, i.e. F is frontal, C is central, T is temporal, P is parietal and O is occipital. Odd-numbered electrodes correspond to the left hemisphere, whereas even numbers record from the right hemisphere and "z" refers to the midline [21].

Because of the smaller head size and vulnerable skin of the newborn, often fewer electrodes are used compared to adult EEG [118]. A restricted 10-20 electrode system consisting of nine electrodes: $Fp_{1,2}$, $C_{3,4}$, Cz, $T_{3,4}$ and $O_{1,2}$ is used in many clinical centers. Since most clinical indications for EEG at this age do not require an excellent spatial resolution, this reduced montage does not compromise the diagnostic capabilities [30].
The electrodes used to record the EEG in these compromised neonates should be sterile. Moreover, the change of the head size with gestational age complicates the use of predesigned electrode caps in the neonatal population. Therefore, adhesive disposable electrodes are commonly used in the NICU. The skin of neonates has a high electrical impedance, so adequate preparation of the scalp is required to obtain good quality tracings. The scalp is typically cleaned using an abrasive gel and a conducting paste is used to lower the impedance. In order to acquire good quality EEG, the impedance should not exceed 10kΩ during the recording [7, 30, 123].

Clinical information

When recording and interpreting the EEG of a newborn, important patient information has to be taken into account. First, the age of the baby should be considered in order to properly interpret the signal. Second, the clinician assessing the EEG should be familiar with the clinical state of the patient and potential medication use. Furthermore, the recording time should be long enough to cover the complete sleep-wake cycling and the vigilance states have to be taken into account during interpretation [30]. At last, the electrode placement and acquisition of the EEG should disturb the neonate as little as possible [7] .

2.3.2 aEEG

As briefly mentioned before, in many clinical centers aEEG is monitored. At the end of the 1960s, Maynard et al. introduced the cerebral function monitor (CFM) with the aim of quickly scanning the brain function of adults at the intensive care unit [130]. Nowadays, it is commonly used in the NICU and is often called amplitude-integrated EEG (aEEG). It is based on an EEG recording with limited electrodes, typically the central electrodes C_3, C_4 and/or the parietal electrodes P_3, P_4. The signals measured from these electrodes are then passed through a bandpass filter enhancing frequencies from 2 to 15 Hz. After filtering, a semi-logarithmic amplitude compression, rectification and smoothing is performed. Moreover, the recording is time-compressed such that each 6 cm on display corresponds to a recording of 1 hour. The key benefits of aEEG are the simple electrode application and the relatively easy interpretation requiring little training. As a result, it is a feasible tool for continuous long-term brain monitoring in the NICU. However, due to the limited spatial coverage and compression important information might be missed [79, 80].

2.3.3 Patterns in the neonatal EEG

Not only the brain, but also the electroencephalogram of a newborn, especially a preterm newborn, is drastically different compared to that of an older child or adult. Moreover, the appearance of the preterm EEG changes rapidly in parallel with the fast physiological maturation of the central nervous system as described in section 2.2.2 [118]. As a result, a pattern that is common at a certain postmenstrual age, might indicate brain abnormalities at another stage in development. For this reason, the most important characteristics of the neonatal EEG for specific ages will be presented in the following sections.

The EEG of a neonate is usually assessed in terms of background continuity, interhemispheric synchrony, the appearance of specific waveforms and the organization of behavioural states [7, 207]. In the next paragraphs, these different maturational effects will be explained in more detail. The most important developmental changes in the EEG are visualized in Figure 2.5.

Organization of behavioural states

In preterm infants only two main sleep stages, active sleep (AS or rapid-eye movement (REM)) and quiet sleep (QS or non REM sleep (NREM)), and wakefulness can be distinguished [56, 118]. Periods with discordant characteristics, which cannot explicitly be assigned to either quiet sleep or non-quiet sleep, are labelled as indeterminate sleep (IS). These often occur at the transition between two well-defined sleep stages and are then labelled as transitional sleep (TS) [7].

Appearance In neonates, the first signs of sleep staging can be observed at around 28 weeks PMA in the EEG. However, it is only at about 30 weeks PMA that the differentiation of sleep states is well established [30, 56]. The EEG pattern typically observed in very young infants, before 30 weeks PMA, is tracé discontinu (TD). This is a highly discontinuous pattern with bursts of high-voltage mixed activity (50 – 300 μV) alternated by long periods of electrographic quiescence ($<$ 25 μV) [56, 118]. The duration of these flat periods, also called interburst intervals (IBI), progressively decreases during maturation while its amplitude increases. Simultaneously, the duration of the bursts increases and their voltage decreases. From 32 weeks onwards, the EEG trace during wakefulness and active sleep becomes more continuous and evolves gradually in a tracé continu [30]. Quiet sleep is consistently more discontinuous compared to active sleep and evolves slower towards a continuous trace. An example of an EEG segment during quiet sleep and active sleep in a preterm infant at 32 weeks PMA is shown in Figure 2.6.
Around 36 weeks PMA, a more complex sleep state organization consisting of four sleep stages and wakefulness is established [56]. Mixed frequency pattern (M) and low voltage irregular pattern (LVI) mainly occur during active sleep, while tracé alternant (TA) and high voltage slow wave (HVS) are most often seen during quiet sleep. An example of the EEG during quiet sleep stages in a term infant can be seen in Figure 2.7, while Figure 2.8 shows an example of the EEG in the two term active sleep stages. As the name suggests, the low voltage irregular pattern is composed of low-voltage (15 – 35 μV) irregular waves of mixed frequencies with dominance of delta and theta activity. The mixed frequency activity M has a similar appearance compared to the LVI,

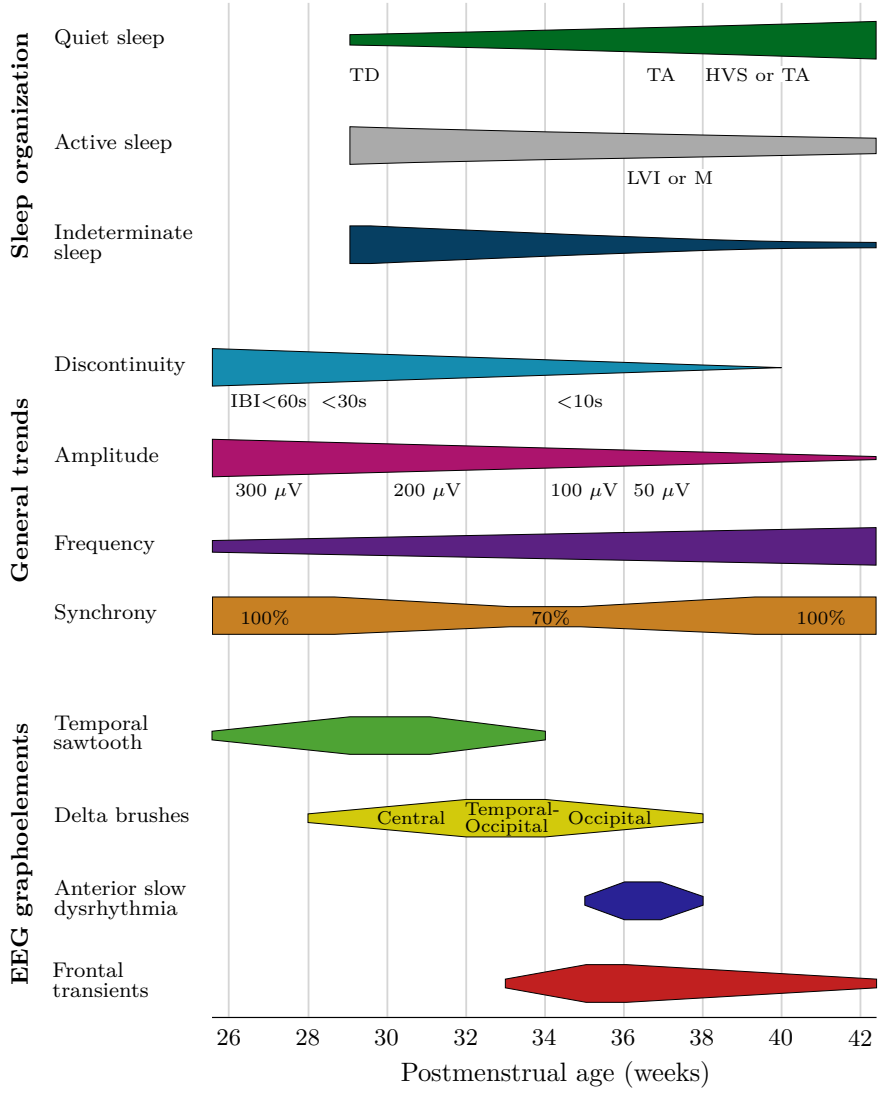
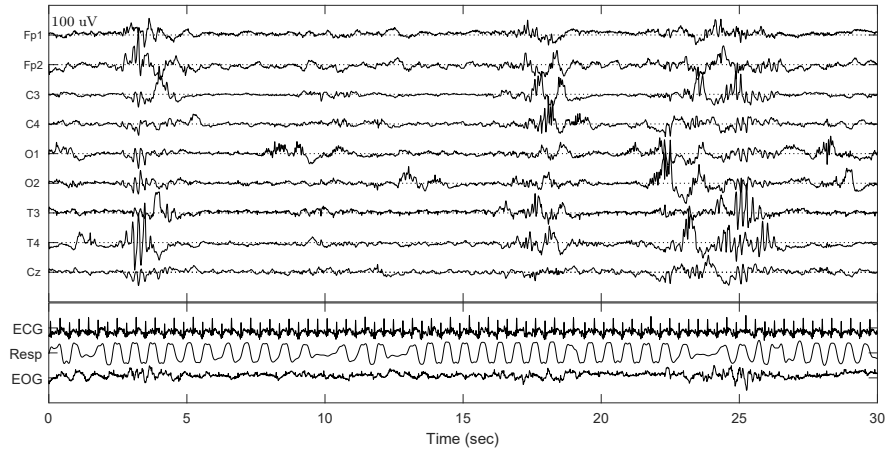


Figure 2.5: Schematic overview of maturational EEG features based on findings by [7, 21, 30, 118, 155, 208]. The top part visualizes how the proportion of the sleep stages evolves during maturation. The middle part illustrates the general changes, related to continuity, amplitude, frequency and synchrony of the signal. The bottom part shows when specific EEG graphoelements appear on the neonatal EEG.

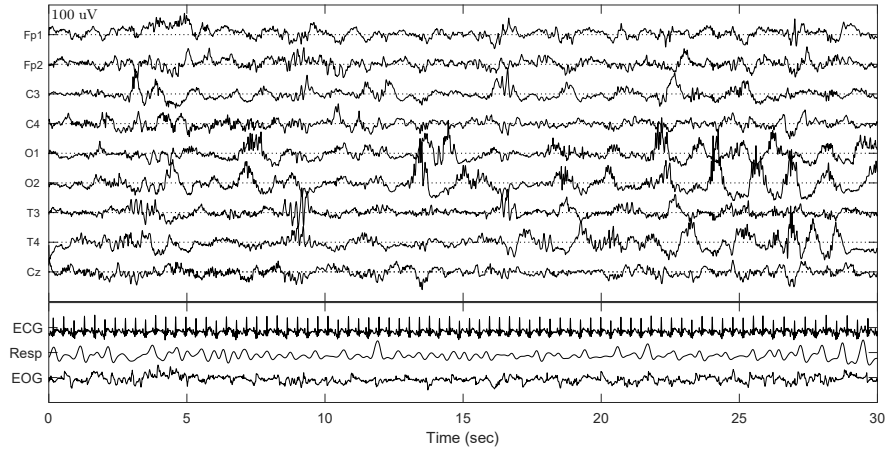
despite the fact that it has higher amplitudes and more slow waves [118]. The tracé discontinu pattern typically observed during quiet sleep changes gradually into the tracé alternant (TA) pattern [56]. This pattern consists of high voltage bursts with interspersed flatter periods. Although this is still a discontinuous pattern, the bursts and IBIs now have approximately the same length (4 – 8 s), while during tracé discontinu the bursts are much shorter compared to the IBIs [96]. Moreover, the bursts are less pronounced, whereas the flat periods have a higher amplitude. Besides, high voltage slow wave (HVS) emerges during quiet sleep and will gradually replace tracé alternant. The HVS pattern consists of continuous diffuse slow wave activity at high voltage (50 – 150 μV) [96, 118]. With the advent of HVS, the EEG trace is becoming more continuous during quiet sleep as well. As a result, there is only a slight difference in discontinuity between quiet sleep and active sleep [56].

Since the electroencephalographic features alone are not enough to distinguish the sleep stages at every postmenstrual age, the golden standard for sleep stage identification is visual analysis of the EEG in combination with noncerebral physiological criteria (e.g. cardiorespiratory patterns, limb movements, electrooculogram). Quiet sleep is characterized by a more deep and regular breathing, and less body and eye movements compared to active sleep or awake. Because the non-cerebral physiological parameters (more body movements, irregular breathing) are similar during wakefulness and active sleep, information about the eye closure is crucial to differentiate these two states.

Sleep organization Not only the appearance of the EEG patterns during sleep, but also the proportion of the different sleep stages evolves at a fast rate during development. Initially, the preterm infant spends up to 90% of the time asleep. During maturation, the percentage of time spent asleep gradually decreases and is around 70% at term equivalent age [12]. In the young preterm infant active sleep is the predominant sleep stage, taking up to even 70% of the total sleep time before 30 weeks PMA. The sleep development is characterized by an increase of relative proportion of quiet sleep, while the time spent in active sleep decreases. At term age, the neonate spends approximately half of the sleep time in quiet sleep and half in active sleep. In addition, with maturation more distinct sleep stages occur and the sleep time labelled as indeterminate sleep reduces progressively [69, 74].

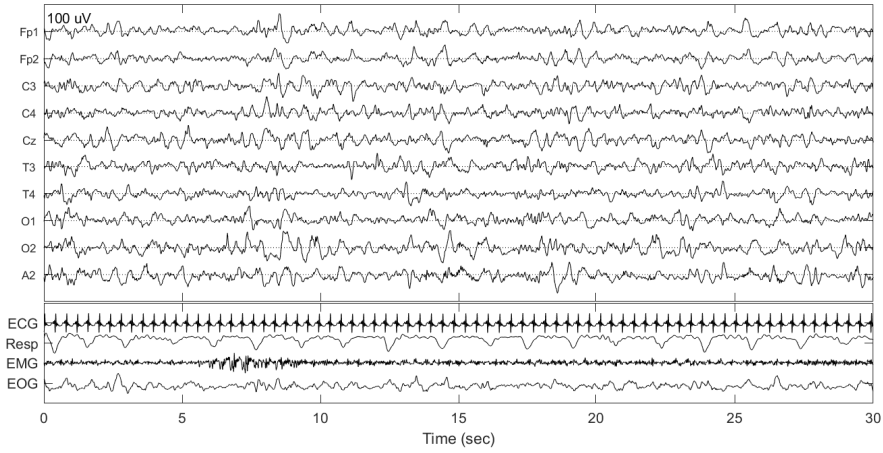


(a) Quiet sleep

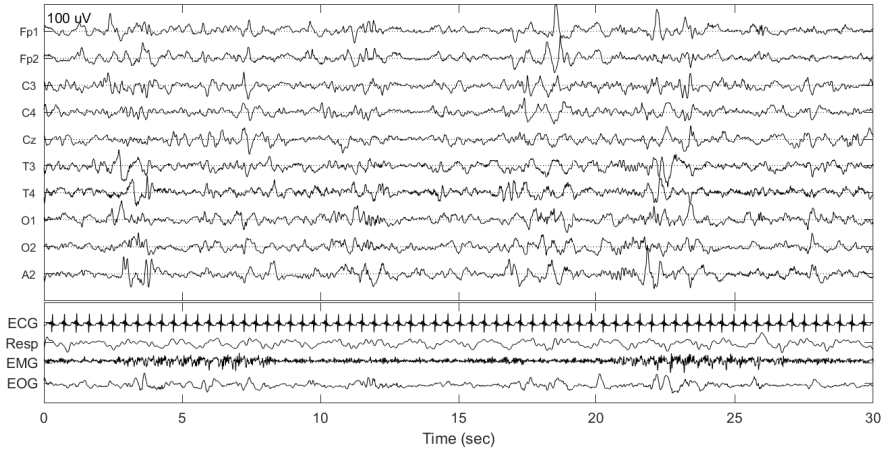


(b) Active sleep

Figure 2.6: Example of the EEG, ECG, respiration (Resp) and electrooculogram (EOG) during (a) quiet sleep and (b) active sleep in a preterm infant (PMA: 32 weeks). The sensitivity of 100 μV corresponds to the distance between the dashed lines.

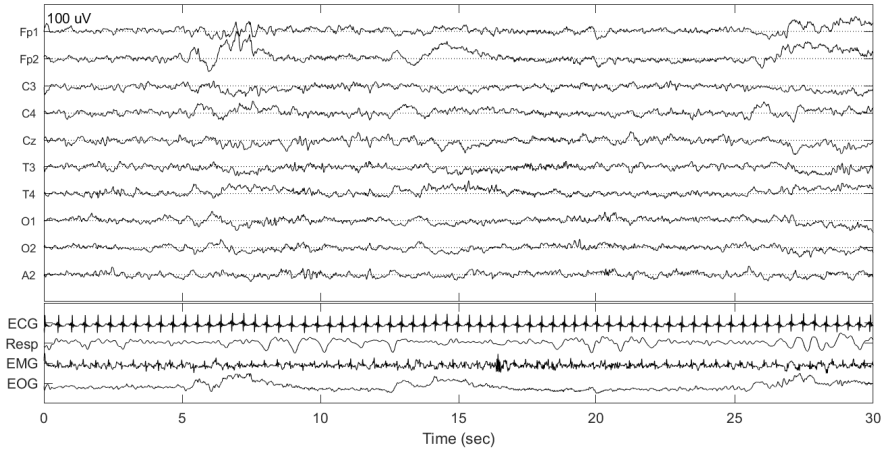


(a) High voltage slow wave

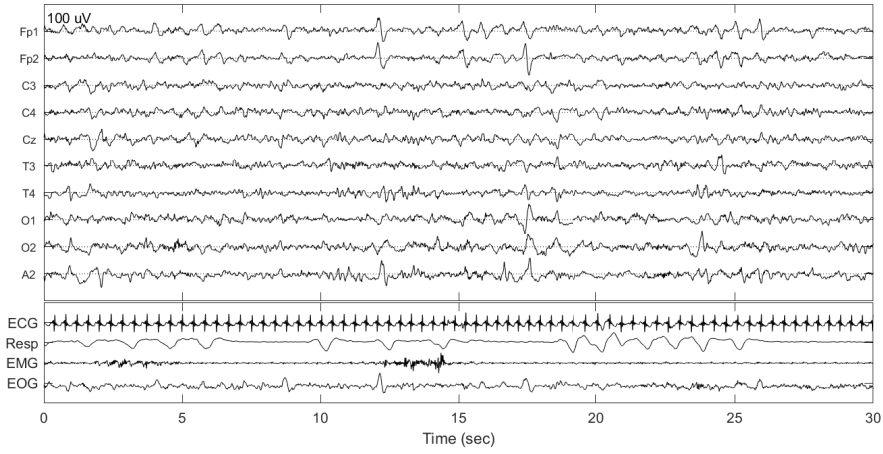


(b) Tracé alternant

Figure 2.7: Example of the EEG, ECG, respiration (Resp) and electrooculogram (EOG) during (a) high voltage slow wave and (b) tracé alternant in a term infant.



(a) Low voltage irregular pattern



(b) Mixed frequency pattern

Figure 2.8: Example of the EEG, ECG, respiration (Resp) and electrooculogram (EOG) during (a) low voltage irregular pattern and (b) mixed frequency pattern in a term infant.

Continuity

As was mentioned in the previous section, the key maturational feature of neonatal EEG is the increase of continuity. The background pattern gradually changes from a discontinuous pattern, consisting of high bursts of activity that alternate with periods of electrographic quiescence, towards a continuous trace with a relatively steady amplitude [21]. A pattern is considered discontinuous if more than 50% of a one minute analysis window is taken up by interburst intervals [7]. Since the continuity increase is one of the prominent electrographic features of brain maturation in preterm infants, the duration of these interburst intervals are often one of the first parameters to assess. The length of the suppressed EEG segments can go up to even 60 s in the very young infants and is less than 10 s long in neonates at 36 weeks [155]. Next to this shortening of the interburst intervals, their amplitude also increases with PMA. In addition, the amplitude of the delta-theta bursts decreases while their length and complexity increases.
The continuous pattern appears first during active sleep, then during the awake state and at last during quiet sleep. The discontinuity is consistently more present during quiet sleep, but near term age the EEG will have continuous activity in all vigilance states.

Synchrony

The synchrony of the EEG refers to the timing of background waves during discontinuous periods at homologous regions of the two hemispheres. It provides information about the development of the corpus callosum and the formation of interhemispheric connections. The EEG is labelled as asynchronous if the onset of the burst is more than 1.5 s apart between the right and left hemisphere [7, 21]. The initial severely discontinuous tracé discontinu pattern is accompanied by hypersynchrony of the EEG. This high degree of synchronization persists up to 30 weeks PMA, from then on interhemispheric asynchrony appears. This asynchrony is physiological and lasts up to 36 weeks, after which the EEG gradually evolves into a synchronous signal again [155]. The degree of synchrony is dependent on the sleep state, a higher degree of synchrony is observed during quiet sleep compared to non-quiet sleep.

EEG grahpoelements

In addition to these general EEG maturational trends described above, the appearance of EEG features at a specific postmenstrual age and with a particular spatial organization is of interest as well. These are called EEG graphoelements

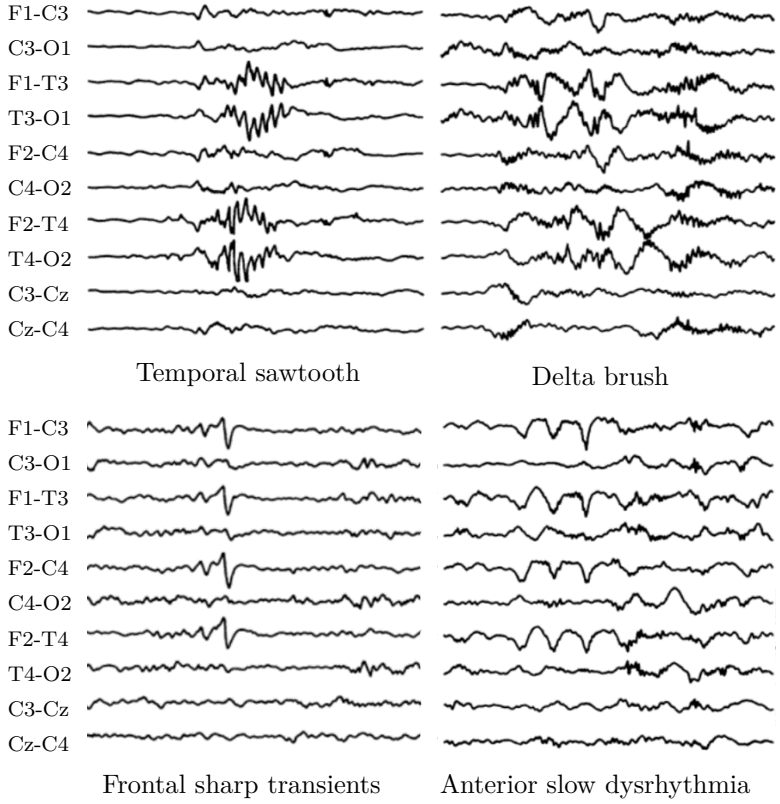


Figure 2.9: Illustration of the most common EEG graphoelements. Adapted from [183].

and the most common ones are delta brushes, temporal sawtooth waves, frontal sharp transients and anterior slow dysrhythmia. Figure 2.9 provides an example of each EEG graphoelement.

Delta brushes Delta brushes are the hallmark of premature EEG. These slow waves (0.3 – 1.5 Hz) with superimposed fast activity (> 8 Hz) can be seen for the first time at around 28 weeks. From then on their number increases and they reach their peak expression at 32 to 34 weeks. From then on their incidence decreases and they disappear around 38 weeks [7, 155]. These delta brushes have a strong spatial organization and are initially diffuse, then they are predominant in the central regions, subsequently they start to occur in

the temporal-occipital regions and near term age they are mainly observed occipitally [7]. The amplitude of the brushes peaks between 30 and 31 weeks at around 300 μV and decreases with age up to 50 to 100 μV at term equivalent age [155, 211]. Aside from the amplitude, the frequency also changes with maturation. The frequency of the slow component shifts from the lower (0.3 – 1 Hz) to the higher (2 – 3.5 Hz) border of the delta frequency range. The frequency of the superimposed fast activity increases up to 36 weeks, but reduces from then onwards [211].

Temporal sawtooth waves Temporal sawtooth waves, also called premature temporal theta, consists of bursts of theta activity (4 to 6 Hz) in the temporal areas [118]. They can be seen for the first time at 24 weeks PMA. They are expressed maximally between 29 to 31 weeks after which its occurrence declines. Whereas the length and number of theta bursts increases up to 31 weeks, the amplitude decreases. Temporal sawtooth waves can typically be observed on homologous EEG channels, but are not synchronous. Although the sleep states are not that well established yet at this age, they are more frequent during active sleep [7].

Frontal sharp transients or encoches frontales As the name implies these are high-voltage sharp waves that can be observed on the frontal channels of the EEG. Immature frontal transients can already be seen at 33 weeks PMA and are high-voltage, polymorphic, asymmetric discharges [7, 155]. The mature encoches frontales emerge at 35 weeks and have a diphasic morphology. Typically, a small negative deflection is followed by a wider, and sometimes larger, positive deflection. These EEG transients can be unilateral, but when bilateral they are symmetrical and synchronous among the two hemispheres. Their amplitude and duration is in the range of 50 – 200 μV and 0.5 – 0.75 s, respectively [7]. They persist up to term age and should disappear completely by 48 weeks PMA [155].

Anterior slow dysrhythmia These short mono- or polymorphic delta waves (1 to 3 Hz) appear in the frontal regions. They have an amplitude in the range of 50 to 100 μV and appear between 36 and 37 weeks during active sleep [7].

2.3.4 Artifacts

In addition to the typical preterm EEG patterns described above, artifacts are omnipresent in neonatal EEG recordings. Based on their source they can

be divided into two types: physiological or extra-physiological. Since artifacts can heavily distort the analysis, artifact detection and removal is a crucial and necessary step prior to performing automated EEG analysis. The most common artifacts in neonatal EEG will be briefly discussed below [7, 109].

Extra-physiological artifacts

- Powerline interference: this is the most common nonbiological source of artifacts and is characterized by interference at 50 or 60 Hz (with possibly additional harmonics). The presence of the artifact can be easily detected by investigating the power spectrum of the signal. Since it is a narrowband artifact, it can be removed using a notch filter [7].
- Electrode artifacts: these can be different types of artifacts related to the scalp electrodes (e.g. poor contact, lead movement, salt bridge artifact, etc.). However, the most common electrode artifacts on the EEG are electrode "pops". These are spontaneous discharges that appear as a sudden spike discharge followed by a more gradual return to baseline. These are seen focally in a single or a few electrodes, which typically have a poor contact and high impedance [7, 109]. They can be easily visually detected due to their specific morphology.
- High frequency ventilators: these ventilators induce a diffuse, rhythmic artifact on all electrodes at a constant frequency between 10 and 15 Hz [7, 109].
- Rocking or patting artifact: these are a diffuse, repetitive artifacts. These artifacts can have a varying frequency over time and can be confused with ictal activity [21, 109].

Physiological artifacts

- Muscular artifacts: different types of artifacts caused by muscle activity exist. EMG is an important source of noise in neonatal EEG and can be recognized by its higher frequency content (>50 Hz) [7]. In addition to EMG artifacts, glossokinetic artifacts, such as chewing and sucking, can also appear. Logically, these muscular artifacts are more frequently observed during wake periods [30].
- Ocular artifacts: these slow waves are due to the movement of the eye ball, which acts as a dipole. These can be observed in the frontal regions, but are less common in neonates compared to adults [7].

- Cardiac artifacts: these artifacts can be divided into electrical (ECG) and mechanical (pulse) artifacts. The electrical artifact is due to the ECG signal picked up by the scalp electrodes. This interference is observed as sharp peaks coinciding with the QRS complex. Hence, the artifact can be identified by comparing the EEG tracing with the ECG signal. Independent component analysis (ICA) is typically used to discard this source of interference. The mechanical artifact is caused by the movement of EEG electrodes on top of a pulsating blood vessel. This artifact can be recognized as a slow rhythmic wave. This occurs mainly in the frontal and temporal leads [109].
- Respiratory artifacts: these mechanical artifacts are caused by respiratory movement and appear as slow activity synchronous to the breathing pattern of the neonate.

2.4 Conclusion

The human brain undergoes a series of complex developmental processes during gestation. Various techniques have been proposed to monitor the structure and function of the developing brain. The EEG measures the electrocortical activity and is a useful modality to assess the brain function in prematurely born neonates. Brain maturation is characterized by a gradual increase of the EEG continuity, the appearance of specific EEG graphoelements and changes in the sleep state organization.

Chapter 3

Mathematical background

This chapter provides background information on the mathematical techniques used in this thesis. First, commonly used maturational or sleep-state related features will be described. Afterwards, we will give a brief introduction to tensors and the canonical polyadic decomposition. Then, the supervised learning techniques adopted in this thesis, namely support vector machines, deep learning and regression, will be explained in more detail. In the last section, the metrics to assess the performance of classification or regression models will be presented.

3.1 EEG features

This section provides an overview of the features extracted from the EEG in this research. These features characterize different aspects of the EEG signal and are used as a biomarker for brain maturation or can discriminate neonatal sleep stages. Some of these features are generally used in the analysis of EEG, while some features are tailored to the specific properties of the preterm EEG. First, features related to the continuity of the EEG time series will be discussed. Then measures of the signal entropy are explained. At last, features capturing the spectral content of the EEG signal are presented.

3.1.1 EEG continuity

As described in the previous chapter, the neonatal EEG evolves from a discontinuous towards a continuous tracing during maturation. As a

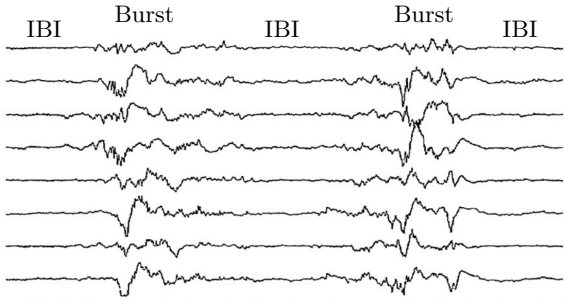


Figure 3.1: Example of a discontinuous EEG segment with indicated bursts and interburst intervals.

consequence, the continuity of the preterm EEG is commonly assessed in order to evaluate the background pattern. A straightforward way to quantify the continuity of the signal is to investigate the occurrence of high-activity bursts, also called spontaneous activity transients (SAT), or measures related to the duration of interburst intervals (IBIs). Therefore, various methods to detect these SATs have been developed [102, 150, 153]. In Figure 3.1 an example of a discontinuous EEG segment with labelled bursts and interburst intervals is provided. After automated identification of the bursts, the most commonly derived features are the burst percentage or SAT%, computed as the proportion of recording time labelled as burst, and the median or maximum length of the interburst intervals.

3.1.2 Entropy of the EEG

Approximate entropy

The term "entropy" is generally used to assess the order or disorder of a system. In signal processing, entropy is adopted to measure the irregularity or unpredictability of a time series.
Approximate entropy (ApEn) was introduced by Pincus et al. [160, 161] as a statistic to robustly quantify the regularity of a time series. In order to compute the approximate entropy of the time series $\{x(1), x(2), \ldots, x(N)\}$ of length N, the signal is divided into segments of m consecutive data points. The distance d from the ith segment or template $u(i) = [x(i) \ldots x(i+m-1)]$ to all other segments is then computed to determine the number of segments close to the template. A segment u(j) is similar to the template u(i) if the Chebyshev

distance between the two segments d[u(i),u(j)]= $\{\max[x(i+k)-x(j+k)]$, $0 \leqslant k \leqslant m-1\}$ is smaller than the defined tolerance r. The probability that any segment u(j) is similar to the template u(i) is computed as:

$$C_i^m(r) = \frac{\text{number of u(j) such that } d[u(i),u(j)] \leqslant r}{N-m+1}. \tag{3.1}$$

The approximate entropy is then defined as:

$$ApEn(m,r,N) = \phi^m(r) - \phi^{m+1}(r) \tag{3.2}$$

$$\tag{3.3}$$

where

$$\phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln C_i^m(r). \tag{3.4}$$

In layman's terms, approximate entropy measures the likelihood that sequences of m consecutive data points close to each other will still be close to each other when an extra data point is added to the sequence. A regular time series will be characterized by a lower ApEn compared to an irregular, less predictable time series [39].

Sample entropy

Richmann and Moorman proposed sample entropy as an improvement to approximate entropy, being less sensitive to the signal length N [165]. In contrast to approximate entropy, self-matches are not taken into account in sample entropy resulting in a reduction of the bias. The second difference is that only the first $N-m$ segments of length m are selected to ensure that the corresponding segment of length m+1 is also defined. Thus, the probability of two matching sequences of length m, $B^m(r)$, is then equal to

$$B^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r) \tag{3.5}$$

where $B_i^m(r)$ is $(N-m-1)^{-1}$ times the number of segments u(j) within a distance r from u(i) with $1 \leqslant j \leqslant N-m$ and $j \neq i$. Similarly for the segments of length m +1, $A^m(r)$ is the probability of two matching sequences of length m +1:

$$A^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} A_i^m(r) \tag{3.6}$$

where $A_i^m(r)$ is $(N-m-1)^{-1}$ times the number of segments u(j) within a distance r from u(i) with j going from 1 to $N-m$ and excluding self-matches. Sample entropy is then computed as:

$$SampEn(m, r, N) = -\ln \frac{A^m(r)}{B^m(r)}, \tag{3.7}$$

which can be reduced to

$$SampEn(m, r, N) = -\ln \frac{A}{B}, \tag{3.8}$$

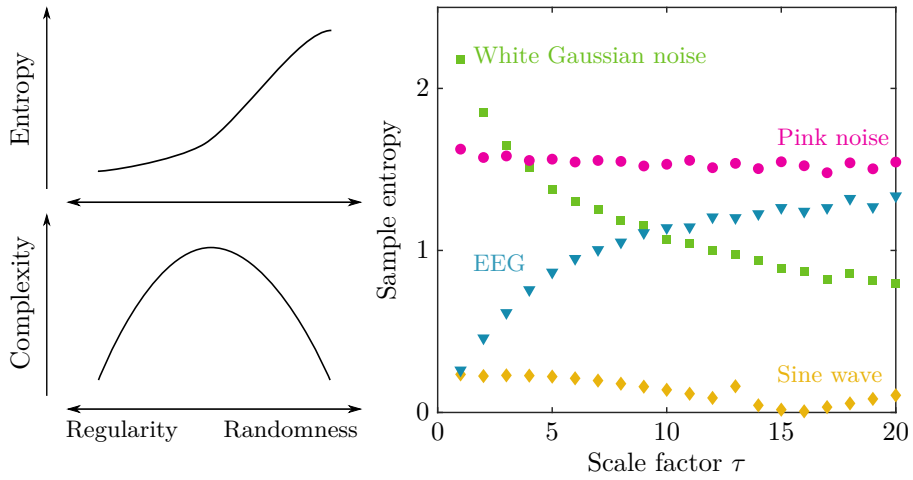
where B and A represents the number of template matches at length m and $m+1$, respectively. This ratio of $\frac{A}{B}$ is the conditional probability that two sequences of length m matching within a tolerance r will still be matching for length $m+1$. The tolerance is typically set as a percentage of the standard deviation of the original time series. This normalization prevents the result from being affected by the variance of the signal [39, 165].

Multiscale entropy

The conventional entropy measures explained in the previous paragraphs assess the regularity of a signal. A regular signal will be assigned a low entropy value, and the entropy will increase monotonically with the degree of randomness [38]. Consequently, entropy reaches its maximum for completely random processes [216]. This is illustrated in the left plot of Figure 3.2. However, complex signals contain "meaningful structural richness", hence they incorporate non-random fluctuations at multiple scales [38, 127]. So, neither a completely random signal (large entropy value) nor a completely regular signal (small entropy value) are truly complex. Therefore, entropy is not a suitable statistic to quantify the complexity of a time series. To address this problem, Costa et al. have proposed multiscale entropy (MSE) [38, 39]. This quantitative measurement of complexity measures the sample entropy over a range of time scales. As a consequence, it can distinguish truly complex dynamics from uncorrelated random or deterministic processes [38].

The procedure to compute the multiscale entropy of a signal consists of two steps (Figure 3.3). The signal is first coarse-grained according to the scale factor τ. In order to do this, the time series $\{x(1), x(2), \ldots, x(N)\}$ is segmented into nonoverlapping windows of length τ and the mean of the data points in each of these windows is computed. Thus, each element of the coarse-grained time series is computed as:

$$y_j^\tau = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad 1 \leqslant j \leqslant \frac{N}{\tau} \tag{3.9}$$



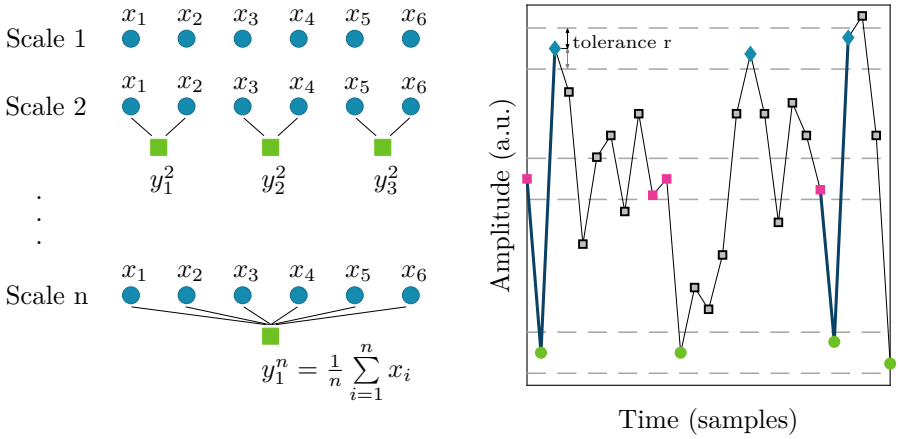
(a) (b)

Figure 3.2: (a) The convential entropy metrics increase monotonically with increasing randomness of the signal. However, the complexity of both a completely regular and a completely random signal is low [216]. (b) The multiscale entropy curves of white Gaussian noise (green squares) decreases with increasing scale factor. The sample entropy of pink noise (yellow circles) is constant and for most scale factors higher than for white Gaussian noise. The sample entropy of a regular sine wave (pink diamonds) is low across all scales. The sample entropy of a neonatal EEG segment (blue triangles) increases for low scales after which it levels off.

resulting in a new time series $y^\tau = \{y_j^\tau, j = 1, ..., \frac{N}{\tau}\}$. This coarse-graining procedure corresponds to applying a moving average filter

$$y_j^\tau = \frac{1}{\tau} \sum_{k=0}^{\tau-1} x(j-k), \quad 1 \leqslant j \leqslant N \tag{3.10}$$

followed by downsampling with a factor τ [83, 195]. The length of the coarse-grained time series is equal to the length of the original signal divided by the scale factor τ. At scale 1, the original time series will be obtained. The next step is to calculate the sample entropy for each coarse-grained time series. Finally, these sample entropy values can be plotted as a function of the scale τ which will be referred to as the multiscale entropy curve. This curve can then be used to assess and compare the complexity of different time series [39].



(a) Coarse-graining (b) Sample entropy computation

Figure 3.3: Illustration of the computation of the multiscale entropy. (a) The coarse-graining procedure consists of averaging consecutive samples in nonoverlapping windows of length τ. (b) Next, the sample entropy of each coarse-grained time series is computed.

The multiscale entropy of a regular signal is expected to be low at all temporal scales, while an uncorrelated random signal will have a high entropy at low scale factors but it declines for increasing values of the scale factor τ [39, 216]. At last, a complex signal typically has consistent entropy value over multiple scales. In order to illustrate this, we have computed the multiscale entropy of white Gaussian noise, pink noise, a sine wave and a segment of neonatal EEG. All signals were equally long and normalized to unit standard deviation. The resulting MSE curves are presented in Figure 3.2. From the graphs in Figure 3.2 it can be seen that the sample entropy of the white Gaussian noise monotonically decreases for increasing scale. The 1/f noise contains long-range correlations and is thus more complex than white Gaussian noise. This is confirmed by the fact that the MSE curve of the pink noise is higher compared to the one of Gaussian noise for most of the scales. Note that the sample entropy at scale one is higher for white Gaussian noise compared to the pink noise. This illustrates that the conventional entropy metric can be misleading and is not capable of quantifying the real complexity of the data. Moreover, the regular sine wave has a low entropy value across all scales. Finally, the MSE curve of the EEG segment increases first in the small scales after which it levels off, representing healthy complexity.

3.1.3 Spectral features

Relative power in frequency subbands

The EEG signal is often subdivided and analysed in different frequency bands: delta (0.5 – 4 Hz), theta (4 – 8 Hz), alpha (8 – 12 Hz) and beta (12 – 30 Hz) band. In preterm infants, the delta band is sometimes subdivided in two bands: delta 1 (0.5 – 1 Hz) and delta 2 (1 – 4 Hz) [142]. A commonly used feature in automated EEG analysis is the (relative) power in these different frequency subbands. In order to compute these features, the power spectral density is first estimated using Welch's method. The relative spectral power is then calculated as the proportion of power in the specific frequency band compared to the power in the complete frequency band of interest (delta - beta). So, the relative power in the frequency band from f_1 to f_2 can be computed as:

$$\text{relative power}(f_1, f_2) = \frac{\int_{f_1}^{f_2} P(f)df}{\int_{f_{min}}^{f_{max}} P(f)df}, \tag{3.11}$$

where P(f) represents the power of the signal as a function of the frequency and f_{min} and f_{max} denote the lower and upper boundary of the frequency range of interest, 0.5 and 30 Hz, respectively.
An example of the power spectral density of a 30 s EEG segment is plotted in Figure 3.4. The delta, theta, alpha and beta band are indicated in blue, red, yellow and purple, respectively. Computation of the relative power in a specific subband corresponds to the ratio of the area under the power spectral density (PSD) in that frequency band to the area under the PSD in the complete frequency band of interest.

Spectral edge frequency

Another feature extracted from the power spectral density is the spectral edge frequency. The spectral edge frequency x or SEF x is the frequency below which x% of the power of the signal is located and is typically computed for x equal to 75% or 90%.

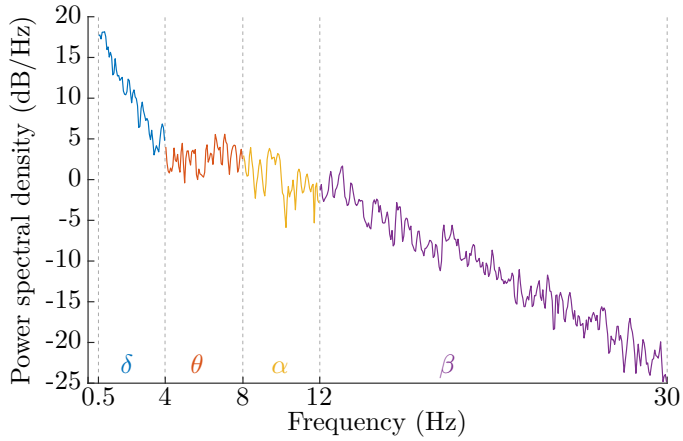


Figure 3.4: Power spectral density of a 30 s EEG segment with delta, theta, alpha and beta frequency band indicated in blue, red, yellow and purple, respectively.

3.2 Tensors

3.2.1 Multiway data

While vectors and matrices are respectively one-way and two-way arrays of numerical values, three-way or generally higher-way arrays are called tensors [174]. Hence, tensors are multidimensional arrays and can be seen as a generalization of vectors and matrices to a higher dimension. This is illustrated in Figure 3.5 where a scalar, vector, matrix and third-order tensor are represented. These higher-way arrays do not only appear in theoretical textbooks, but can actually be seen anywhere around you in daily life. For example, an RGB image is naturally represented by a third-order tensor where every slice contains

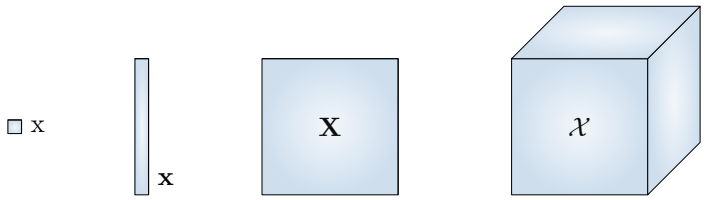


Figure 3.5: A scalar x, vector $\mathbf{x}$, matrix $\mathbf{X}$ and tensor $\mathcal{X}$.

the intensity of each pixel for one of the three specific colours. Video is even a fourth-order tensor due to the additional mode for temporal information. Furthermore, the recommender system of Netflix is based on a third-order tensor containing the ratings of customers for different movies dependent on the time they watched them. Or a company can store their sales in a three-way array with a mode for the products, the time and the location, etc.

Many physiological signals, including the EEG, are recorded using multiple leads. These multichannel data are typically represented as two-way arrays showing the amplitude of the signal as a function of time for each spatial location. However, in many (biomedical) applications more than two modes (space and time) naturally arise. Data can for example be collected from multiple subjects or for multiple trials of a specific experiment. As a result, the data will fit in a multi-way array with modes $channels \times time \times subjects$ or $channels \times time \times trials$, respectively. In this way, common patterns among different subjects or trials can be detected [87]. Even when the collected data is not naturally represented as a higher order tensor, several expansions can be performed on the time series. These transformations allow to capture additional information about the signal and will also yield a multi-way array [87]. Time-frequency analysis of the EEG is commonly used, because it is generally known that the frequency of neural oscillations carries important information for various applications. Performing a frequency transformation of the EEG time series, such as a wavelet transform or short-time Fourier transform, will result in a third-order tensor with modes *channels, time and frequency*. In addition, other tensorization techniques such as hankelization, segmentation or löwnerization can be useful in biomedical applications [85, 151]. The appropriate tensorization technique depends on the application and available data.

In order to analyse these higher order tensors, the multi-way array could be unfolded into a matrix, which can then be investigated using well-established matrix tools. However, it is obvious that reshaping the tensor into a matrix destroys the interactions among the different modes. By using tensor algebra rather than reshaping the tensor, the data structure can be optimally exploited [37]. In addition to the fact that the structural information is better captured, there are also mathematical uniqueness properties which make tensor decompositions attractive. The fact is, in contrast with matrix operations, tensor decompositions are easily unique under mild conditions.

3.2.2 Notations and definitions

In the remainder of this thesis, scalars will be denoted by lowercase letters (x), vectors by bold lowercase letters ($\mathbf{x}$), matrices by bold uppercase letters ($\mathbf{X}$)

and higher-order tensors by calligraphic letters ($\mathcal{X}$).
The *order* of a tensor $\mathcal{X}$ refers to the number of dimensions, and is also known as modes or ways [32]. So, a scalar is a tensor of order zero, a vector is a tensor of order one, a matrix is a tensor of order two, and generally a tensor of order N is denoted by $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$.
By fixing only a subset of a tensor's indices, subarrays of the original data tensor are selected [31]. A *mode-n fiber* is obtained by fixing all indices except for the nth. The mode-1, mode-2 and mode-3 fibers are referred to as column, row and tube fibers, respectively. Similarly, a tensor *slice* is a two-way subarray of the original tensor obtained by fixing all but 2 indices. For third order tensors, the horizontal ($\mathbf{X}_{i::}$), lateral ($\mathbf{X}_{:j:}$) and frontal ($\mathbf{X}_{::k}$) slices are obtained by fixing only the first, second and third index, respectively.
The *Frobenius norm* of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ is computed as [101]:

$$\|\mathcal{X}\|_F = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \ldots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \ldots i_N}^2}. \tag{3.12}$$

The *outer product* or *tensor product* of the N-way tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ and the M-way tensor $\mathcal{B} \in \mathbb{R}^{J_1 \times J_2 \times \ldots \times J_M}$ yields the (N+M)-way tensor $\mathcal{C} = \mathcal{A} \circ \mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N \times J_1 \times J_2 \times \ldots \times J_M}$ with entries

$$c_{i_1 i_2 \ldots i_N j_1 j_2 \ldots j_M} = a_{i_1 i_2 \ldots i_N} b_{j_1 j_2 \ldots j_M}. \tag{3.13}$$

Note that the entries of $\mathcal{C}$ consists of all possible pairwise products of the elements of $\mathcal{A}$ and $\mathcal{B}$ [32]. So, the outer product of the vectors $\mathbf{a} \in \mathbb{R}^{I_1}$ and $\mathbf{b} \in \mathbb{R}^{I_2}$ will result in a rank-1 matrix $\mathbf{C} = \mathbf{a}\mathbf{b}^T = \mathbf{a} \circ \mathbf{b} \in \mathbb{R}^{I_1 \times I_2}$. Generally, the tensor product of N vectors $\mathbf{a}^{(n)} \in \mathbb{R}^{I_n}$ (n = 1, 2,..., N) yields a rank-1 tensor:

$$\mathcal{X} = \mathbf{a}^{(1)} \circ \mathbf{a}^{(2)} \circ \ldots \circ \mathbf{a}^{(N)} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}. \tag{3.14}$$

Each entry of the tensor $\mathcal{X}$ is computed as the product of the corresponding vector elements [101]:

$$x_{i_1 i_2 \ldots i_N} = a_{i_1}^{(1)} a_{i_2}^{(2)} \ldots a_{i_N}^{(N)}. \tag{3.15}$$

3.2.3 Canonical polyadic decomposition

Various tensor decompositions and factorizations have been proposed during the past decades. Although the foundation of many of these tensor decompositions was laid a long time ago, these techniques have recently received increased attention from researchers in a variety of fields such as signal processing, mathematics, bioinformatics, computer vision and data mining [32]. Nowadays,

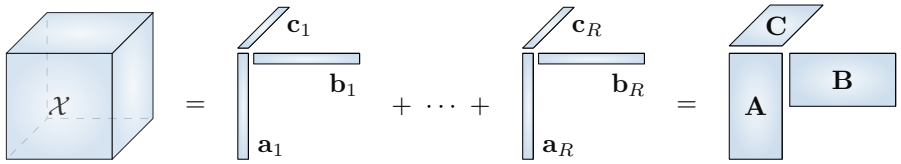


Figure 3.6: The canonical polyadic decomposition of a third-order tensor.

the use of tensor data mining tools is rapidly increasing, due to the ever-increasing amount of available data.
Two of the most commonly used tensor decompositions are the canonical polyadic decomposition (CPD) and the Tucker decomposition. Next to the CPD and the Tucker decomposition, other more advanced decompositions, for example the block term decomposition [46] and the tensor train decomposition [147], exist. Dependent on the specific application and the properties of the data, a different tensor decomposition might be appropriate. In the remainder of this section, we will focus on the CPD, because that is the decomposition used in Chapter 5. For more details about tensor decompositions and their application in signal processing, the reader is referred to [31, 174].

A polyadic decomposition (PD) approximates a tensor by a linear combination of R rank-1 tensors. If the number of rank-1 terms R is minimal, the decomposition is called canonical polyadic decomposition (CPD) and R is the tensor rank. For a third-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ this can be written as:

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!], \tag{3.16}$$

where R represents the number of components and $\mathbf{a}_r \in \mathbb{R}^{I_1}$, $\mathbf{b}_r \in \mathbb{R}^{I_2}$ and $\mathbf{c}_r \in \mathbb{R}^{I_3}$ are the factor vectors. These factor vectors $\mathbf{a}_r, \mathbf{b}_r$ and $\mathbf{c}_r$ are the rth $(1 \leqslant \mathrm{r} \leqslant \mathrm{R})$ column of the factor matrices $\mathbf{A} \in \mathbb{R}^{I_1 \times R}$, $\mathbf{B} \in \mathbb{R}^{I_2 \times R}$ and $\mathbf{C} \in \mathbb{R}^{I_3 \times R}$, respectively. A visual representation of this decomposition for a third-order tensor is given in Figure 3.6.

Stability

In order to find the components of the CPD for a given rank R, the least squares optimization problem:

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \quad f = \frac{1}{2} \left\| \mathcal{X} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \right\|_F^2 . \tag{3.17}$$

has to be solved [174]. Different approaches, such as the well-known alternating least squares (ALS) [101], algebraic methods [45] and nonlinear least squares (NLS) methods [177], have been proposed to minimize this objective function. Because the optimization problem in (3.17) is non-convex, these algorithms might converge to a local minimum. As a result, when decomposing the tensor multiple times using different initializations of the factor matrices, other solutions might be obtained even though the residual between the original data tensor and its CPD is similar. Therefore, it is recommended to run the decomposition multiple times using different initializations and assess the stability of the solution [197]. The initialization of the factor matrices can be done randomly or based on generalized eigenvalue decompositions, however when specific patterns are expected in one or multiple modes, this prior knowledge can be taken into account during the initialization. Note that stability issues might not only occur due to convergence to a local minimum, but it is also possible that the cost function does not have a minimum and as a consequence the optimal CPD solution is non-existent (diverging terms) [107].

The number of components

It is generally known that the column rank of a matrix corresponds to its row rank, and these are equal to the rank of the matrix. The rank of matrix can be computed by means of its singular value decomposition (SVD). Moreover, according to the Eckart-Young theorem, the best low-rank approximation of a matrix can be obtained by truncation of its SVD (only retaining the largest singular values) [60].

However, determining the rank of higher-order tensors is less straightforward. As mentioned before, the rank of a higher-order tensor $\mathcal{X}$ is the smallest number of rank-1 terms that sum to $\mathcal{X}$. The *mode-n rank* of a tensor corresponds to the dimension of the vector space spanned by the mode-n fibers [31]. Hence, the mode-n rank is computed by estimating the rank of the mode-n unfolding of the original tensor, which can be done using established matrix techniques. The N-tuple of mode-n ranks $(R_1, R_2, ..., R_N)$ is then referred to as the *multilinear rank* of the tensor [47]. In contrast to the matrix case, these mode-n ranks do not necessarily have to be the same. Moreover, the rank of the tensor does not have to be the same as one of the mode-n ranks, not even if they are equal [47]. In fact, the rank of a tensor can even be greater than its largest dimension and is per definition greater than or equal to the largest mode-n rank.

In practice, the rank of a tensor is not known a priori and defining the rank is an NP-hard problem. Therefore, different tools to assess the proper number of components have been proposed. First of all, the multilinear singular value

spectrum can give an indication of the appropriate number of components [31]. Moreover, Bro et al. have introduced the core consistency diagnostic (CORCONDIA) to determine the rank [23]. CORCONDIA relies on the fact that the CPD model can also be written as multilinear product with a diagonal core [31] and assesses the degree of superdiagonality of the core tensor for increasing values of the rank R (starting at R = 1). For a tensor $\mathcal{X}$ with core tensor $\mathcal{G}$, the core consistency is computed as:

$$\text{core consistency} = 100\left(1 - \frac{\|\mathcal{G} - \mathcal{I}\|_F^2}{R}\right), \tag{3.18}$$

where $\mathcal{I}$ is a superdiagonal tensor of size $R \times R \times R$ with R the number of components. The last model, so the one with the highest number of components, for which the CORCONDIA is still close to 100% will be selected. This metric was introduced in the field of chemometrics, but has also been used to estimate the number of components for EEG applications [86, 135]. Besides, diagnostics based on the fitting error have been developed. For example, difference of fit (DIFFIT) uses the change in fitting error with increasing number of components to estimate the rank [190]. The fitting/reconstruction error is computed as the squared Frobenius norm of the residual between a tensor $\mathcal{X}$ and its rank-r CPD approximation $[\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$:

$$e(r) = \|\mathcal{X} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]\|_F^2 \tag{3.19}$$

DIFFIT is then defining the optimal rank R by maximizing the ratio of the difference in fitting error between consecutive models [121]:

$$\text{DIFFIT}(r) = \frac{e(r-1) - e(r)}{e(r) - e(r+1)}. \tag{3.20}$$

Although several metrics to estimate the rank (e.g. core consistency and DIFFIT) have been proven successful in several applications, the determination of the optimal number of components is still a frequently occurring challenge. Therefore, it is good practice to compute multiple diagnostic tools and investigate the obtained factor matrices for different ranks [2]. Prior knowledge can often aid in assessing the model. Moreover the occurrence of cancelling diverging terms is an indication of a poor selection of the rank. This problem is referred to as degeneracy and can be overcome by imposing constraints such as nonnegativity, smoothness, sparsity, etc. These additional constraints on the factor matrices can also facilitate interpretation of the obtained signatures [31].

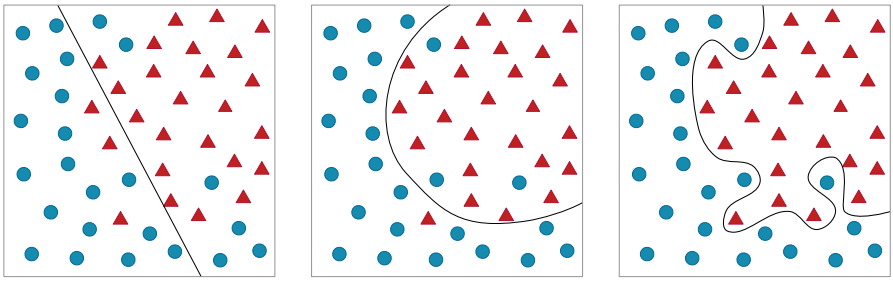
Uniqueness

In matrix decompositions constraints are required to ensure uniqueness, e.g. orthogonality for principal component analysis (PCA) and independence for independent component analysis (ICA). However, for many applications these constraints are too stringent and hamper interpretability of the components. Unlike low-rank matrix decompositions, low-rank tensor decompositions are unique, up to scaling and permutation ambiguity, under relatively mild conditions [59]. This property is one of the main strengths of low-rank tensor decompositions and plays an important role in various applications [31, 174].

3.3 Supervised learning

In supervised learning a training dataset, consisting of N training examples or inputs with their corresponding outputs: $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_N, y_N)\}$, is used to infer a mapping function $y = f(\mathbf{x})$ [67]. The inputs are usually D-dimensional feature vectors: $\mathbf{x}_i = [x_{i,1}\ x_{i,2}\ ...\ x_{i,D}] \in \mathbb{R}^D (1 \leqslant i \leqslant N)$. In *classification* problems the output is a discrete value (e.g. binary classification: $y \in \{-1, 1\}$), while the output is continuous in a *regression* problem.
The goal in supervised learning is to train the model such that the inferred mapping function f can predict the output of new, unseen data points. Therefore, the trained model should have good *generalization capabilities*, which is mostly determined by the model's complexity. On the one hand, if the learned function is too simple, it will not be flexible enough to capture the relationship between the input and output of the training data. As a consequence, unreliable predictions will be obtained for a new unseen test set. This phenomenon is called *underfitting*. On the other hand, a too complex model will tend to fit irrelevant noise in the particular training dataset, known as *overfitting*, thereby risking low generalization and poor predictions. The concepts underfitting and overfitting in the context of classification models are illustrated in Figure 3.7.

We will start by introducing the supervised classification models used in this manuscript. First, the most important concepts of support vector machines (SVMs) will be explained. This will be followed by a brief discussion on least squares SVMs (LS-SVMs) and fixed-size LS-SVMs. Afterwards, the concept of deep learning, more specifically the building blocks and properties of convolutional neural networks, will be presented. At last, linear regression models will be discussed.



(a) Underfitting (b) Good fit (c) Overfitting

Figure 3.7: (a) An underfitted model is too simple to capture the input-output interaction of the training data. Hence, it performs poorly on both training and test set. (b) Appropriate fitting will lead to a generalizable classification model. (c) Overfitting occurs when the model learns the noise in the training data, resulting in poor performance on unseen test samples.

3.3.1 Support vector machines

Support vector machines or SVMs have been introduced by Vapnik et al. in 1995 [202] and are nowadays among the most commonly used supervised learning methods for solving both classification and regression problems.

Let us consider a binary classification problem with training dataset: $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_N, y_N)\}$ composed of N feature vectors $\mathbf{x}_n \in \mathbb{R}^D$ and their class label $y_n \in \{-1, 1\}$ for $n = 1, 2, ..., N$. The SVM will use this labelled training data to find the optimal hyperplane which separates the classes while maximizing the margin between the decision boundary and the closest samples. Actually, only a small subset of the training samples, called support vectors, define the decision boundary. This decision function can subsequently be used to define the class of new, unknown test samples [19]. Figure 3.8 illustrates how the SVM finds the optimal hyperplane.

If the classes are not linearly separable, the feature space can be mapped to a higher-dimensional space where a linear classification of the training dataset is possible. This feature-space transformation of input vectors $\mathbf{x}$ is denoted as $\phi(\mathbf{x})$.

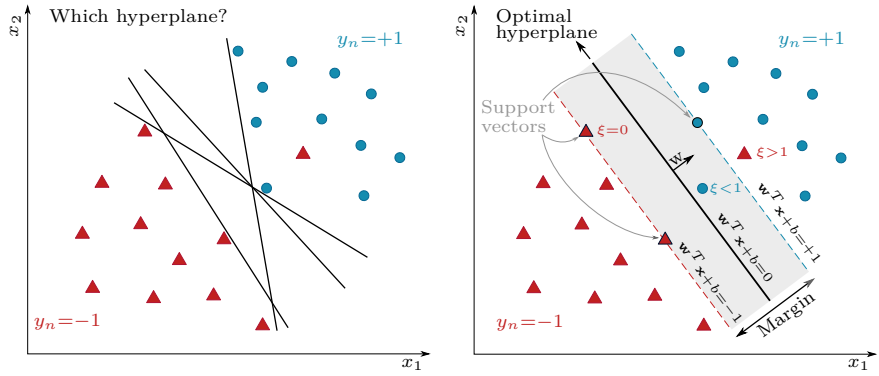


Figure 3.8: An SVM finds the optimal hyperplane by maximizing the margin. The closest data points to the decision boundary, called support vectors, define the position and orientation of the separating hyperplane.

Hard margin

A possible decision function for the binary classification problem in this higher-dimensional feature space can be written as:

$$\mathbf{w}^T\phi(\mathbf{x}) + b = 0 \tag{3.21}$$

where $\mathbf{w}$ is the weight vector normal to the hyperplane and b is the bias [15]. We want to find the two parallel hyperplanes (defining the boundary of the margin) with maximal margin such that for all training data points [185]:

$$\begin{cases} \mathbf{w}^T\phi(\mathbf{x_n}) + b \geqslant +1, & \text{if } y_n = +1 \\ \mathbf{w}^T\phi(\mathbf{x_n}) + b \leqslant -1, & \text{if } y_n = -1 \end{cases} \tag{3.22}$$

which is equivalent to:

$$y_n(\mathbf{w}^T\phi(\mathbf{x}_n) + b) \geqslant 1 \quad n = 1, ..., N. \tag{3.23}$$

As a result, the distance from the support vectors to the hyperplane is equal to:

$$d = \frac{|\mathbf{w}^T\phi(\mathbf{x}) + b|}{\|\mathbf{w}\|} = \frac{1}{\|\mathbf{w}\|}. \tag{3.24}$$

In order to find the optimal separating hyperplane, the margin equal to $2 \times d = \frac{2}{\|\mathbf{w}\|}$ has to be maximized, which is equivalent to minimizing $\frac{\|\mathbf{w}\|^2}{2}$. Therefore,

the optimal weight vector and bias parameter for a linearly separable training dataset can be found by solving the following optimization problem [15]:

$$\begin{aligned} \min_{\mathbf{w},b} J_1(\mathbf{w}) = \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} \\ \text{s.t.} \quad & y_n(\mathbf{w}^T\phi(\mathbf{x}_n) + b) \geqslant 1 \quad n = 1, ..., N. \end{aligned} \tag{3.25}$$

Soft margin

Because in practice the data is usually not linearly separable in the feature space $\phi(\mathbf{x})$, misclassification of some data points is tolerated. Slack variables $\xi_n \geqslant 0, \;\; n = 1, ..., N$ are introduced to penalize mislabelled training data points [19]. Training data points on the correct side of the decision boundary or on the margin are correctly classified and will get a slack variable ξ_n equal to zero. A slack variable penalty $0 < \xi_n \leqslant 1$ is given to data points at the correct side of the hyperplane which lie within the margin. At last, data points at the wrong side of the decision boundary are misclassified and receive $\xi_n > 1$, with a penalty proportional to the distance from the hyperplane [19]. The sum of all slack variables $\sum_{n=1}^{N} \xi_n$ can be seen as a measure of how many training samples are misclassified [15].
The optimization problem of this soft margin method is then equal to:

$$\begin{aligned} \min_{\mathbf{w},b,\xi_n} J_2(\mathbf{w}, \xi_n) = \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} + c\sum_{n=1}^{N} \xi_n \\ \text{s.t.} \quad & y_n(\mathbf{w}^T\phi(\mathbf{x}_n) + b) \geqslant 1 - \xi_n \\ & \xi_n \geqslant 0, \quad n = 1, ..., N. \end{aligned} \tag{3.26}$$

As can be seen from Equation (3.26) the goal of the objective function is dual. On the one hand the margin is maximized, on the other hand the number of misclassifications should be kept as low as possible. The parameter $c > 0$ controls the trade-off between the penalty for misclassification and the width of the margin [15, 19].

The Lagrange multipliers $\alpha_n \geqslant 0$ and $\mu_n \geqslant 0$ (n = 1, ..., N) are introduced to solve the optimization problem in (3.26):

$$\begin{aligned} L_2(\mathbf{w}, b, \xi_n, \alpha_n, \mu_n) = \quad & J_2(\mathbf{w}, \xi_n) - \sum_{n=1}^{N} \alpha_n \left[y_n(\mathbf{w}^T \phi(\mathbf{x}_n) + b) \right. \\ & \left. -1 + \xi_n \right] - \sum_{n=1}^{N} \mu_n \xi_n. \end{aligned} \tag{3.27}$$

If the partial derivatives of $L_2(\mathbf{w}, \text{b}, \xi_n, \alpha_n, \mu_n)$ with respect to $\mathbf{w}$, b and ξ_n are set equal to 0 and the results are substituted in equation (3.27), the Lagrangian in (3.27) can be reduced to the dual quadratic programming problem [19, 185]:

$$\begin{aligned} \max_{\alpha_n} \tilde{L}_2(\alpha_n) = \quad & \sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{n,m=1}^{N} \alpha_n \alpha_m y_n y_m K(\mathbf{x}_n, \mathbf{x}_m) \\ \text{s.t.} \quad & \sum_{n=1}^{N} \alpha_n y_n = 0 \\ & 0 \leqslant \alpha_n \leqslant C, \quad n = 1, ..., N \end{aligned} \tag{3.28}$$

with $K(\mathbf{x}_n, \mathbf{x}_m) = \phi(\mathbf{x}_n)^T \phi(\mathbf{x}_m)$ a kernel function. This *kernel trick* avoids the explicit computation of the mapping to the high-dimensional feature space $\phi(.)$, since this is done implicitly using the positive definite kernel function K [65]. Various kernel functions exist, two commonly used functions are the linear kernel

$$K(\mathbf{x}_n, \mathbf{x}_m) = \mathbf{x}_n^T \mathbf{x}_m \tag{3.29}$$

and the radial basis function (RBF) kernel

$$K(\mathbf{x}_n, \mathbf{x}_m) = \exp\left(-\frac{\|\mathbf{x}_n - \mathbf{x}_m\|^2}{\sigma^2} \right), \tag{3.30}$$

where σ is a tuning parameter.

The dual optimization problem in (3.28) represents a quadratic programming problem in which a quadratic function is optimized with a set of inequality constraints [19]. Only data points with α_n different from zero, the support vectors, are contributing to the predictive model. In this way sparsity is added and only the distance between this subset of data points and the hyperplane are maximized. By solving the Lagrangian and finding the optimal values for

α_n, the weight vectors can be found as:

$$\mathbf{w} = \sum_{n=1}^{N} \alpha_n y_n \phi(\mathbf{x}_n). \tag{3.31}$$

and the bias term can be determined using the Karush-Kuhn-Tucker (KKT) conditions [19].
The SVM decision function is then equal to:

$$\begin{aligned} y(\mathbf{x}) &= \text{sign } \left[\mathbf{w}^T \phi(\mathbf{x}) + b\right] \\ &= \text{sign } \left[\sum_{n=1}^{N} \alpha_n y_n K(\mathbf{x}, \mathbf{x}_n) + b\right] \end{aligned} \tag{3.32}$$

3.3.2 Least squares support vector machines

Least squares SVMs have been proposed by Suykens et al. [185]. In the LS-SVM the objective function of the optimization problem in (3.26) is formulated as a least squares problem. Moreover, equality constraints instead of inequality constraints are imposed:

$$\begin{aligned} \min_{\mathbf{w},b,e_n} J_3(\mathbf{w}, e_n) = \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} + \gamma\frac{1}{2}\sum_{n=1}^{N} e_n^2 \\ \text{s.t.} \quad & y_n(\mathbf{w}^T\phi(\mathbf{x}_n) + b) = 1 - e_n \quad n = 1, ..., N, \end{aligned} \tag{3.33}$$

where γ is a regularization parameter. The Lagrangian is then defined as:

$$L_3(\mathbf{w}, b, e_n, \alpha_n) = J_3(\mathbf{w}, e_n) - \sum_{n=1}^{N} \alpha_n \left[y_n(\mathbf{w}^T\phi(\mathbf{x}_n) + b) - 1 + e_n\right] \tag{3.34}$$

with Lagrange multipliers α_n. After solving for the conditions of optimality, this can be written as a system of linear equations [185]:

$$\left[\begin{array}{c|c} 0 & \mathbf{y}^T \\ \hline \mathbf{y} & \mathbf{\Omega} + \gamma^{-1}\mathbf{I} \end{array}\right] \left[\begin{array}{c} b \\ \hline \alpha \end{array}\right] = \left[\begin{array}{c} 0 \\ \hline 1_N \end{array}\right],$$

where $\alpha = [\alpha_1 \ \alpha_2 \ ... \ \alpha_N]^T$ are the Lagrange multipliers, $\mathbf{y} = [y_1 \ y_2 \ ... \ y_N]^T$, $1_N = [1 \ 1 \ ... \ 1]^T$ and $\Omega_{n,m} = y_n y_m \phi(\mathbf{x}_n)^T \phi(\mathbf{x}_m) = y_n y_m K(\mathbf{x}_n, \mathbf{x}_m)$ for n,m = 1,..., N.
As a consequence , the formulation of the LS-SVM classifier can be found by

solving a system of linear equations instead of a quadratic programming problem [185]. Moreover, in contrast with SVMs, there is no sparsity since all support values α_n are different from zero. As a result, all training data points contribute to the model.

Fixed-size LS-SVM

For very large training datasets, evaluation of the kernel function for each training example can be computationally expensive. Therefore, a number of methods to deal with these large scale datasets have been suggested (e.g. Nyström method) [186]. Here we will focus on the fixed-size LS-SVM, where a representative subset of the training set ($M \ll N$) is selected ensuring that the underlying distribution of the original data is retained. So, a predefined number of training examples are iteratively selected by maximization of the quadratic Renyi entropy [65, 186]:

$$H_R = -\log \int p(x)^2 dx \tag{3.35}$$

which can be approximated by

$$\int \hat{p}(x)^2 dx = \frac{1}{N^2} 1_N^T \mathbf{\Omega} 1_N^T, \tag{3.36}$$

where $\mathbf{\Omega}$ and N denotes the kernel matrix and the total number of training samples, respectively.

3.3.3 Deep learning

Finding an appropriate set of useful discriminative features to represent the input data is an important step in the traditional machine learning pipeline. This feature engineering is a time-consuming, problem-specific and challenging task, which heavily relies on prior domain knowledge [117]. Deep learning, by contrast, can learn the relevant features directly from the raw input data, making manual feature engineering redundant [117].
Various types of deep neural networks exist (e.g. autoencoders, deep belief networks, recurrent neural networks, convolutional neural networks). Convolutional neural networks or CNNs are among the most commonly used ones and are biologically inspired by the structure and function of the mammalian visual cortex [170]. As a result, they have been abundantly applied for image recognition [108, 116], but also in other domains such as speech processing tasks [1], recommender systems [42] and biomedical signal processing [3, 184]. The

building blocks and properties of CNNs will be explained in more detail in the next section.

Convolutional neural networks

The architecture of a convolutional neural network consists of a sequence of different types of layers: convolutional layers, nonlinear activation layers and pooling layers. These stacked layers automatically extract descriptive features from the input. The input is usually a three-dimensional volume with modes: width, height and depth. The final layers of the network are fully connected and use the extracted features to perform the classification. The output of the network is a class score, representing the probability of belonging to one of the output classes.

Convolutional layer Convolutional (Conv) layers are the core building blocks of a CNN and are characterized by some specific properties. A first important property of the convolutional layer is that, unlike traditional artificial neural networks, the neurons are not fully connected. This *local or sparse connectivity* means that each neuron is only connected to a small region of the input volume, called the *receptive field*. The spatial extent of this subregion is defined by the filter size. It is important to note that this local connectivity only holds in the spatial mode (width and height) and the neuron is fully connected along the depth of the input volume [72].
A second property of CNN is the use of *parameter sharing*. Due to the shared parameters, the CNN has shift invariance properties.Moreover, these properties lead to a drastic reduction in the number of parameters, resulting in a substantial increase of the computational efficiency.

In the convolutional layer a dot product between the filter weights and the entries in the subregion of the input volume is computed. This operation is performed in a sliding window fashion, where the *stride* defines how much the kernel or filter moves in each direction. The output of this layer is called an activation or feature map. This operation is represented schematically in Figure 3.9. Mathematically, this computation corresponds to a convolution, hence the name [117].

For a three-dimensional input I(i,j,p), this convolution operation can be written as:

$$\mathcal{O}(i,j,k) = \sum_{p=1}^{P} \sum_{n=1}^{N} \sum_{m=1}^{M} \mathcal{F}_k(m,n,p)\mathcal{I}(i-m+1, j-n+1, p) + b_k$$

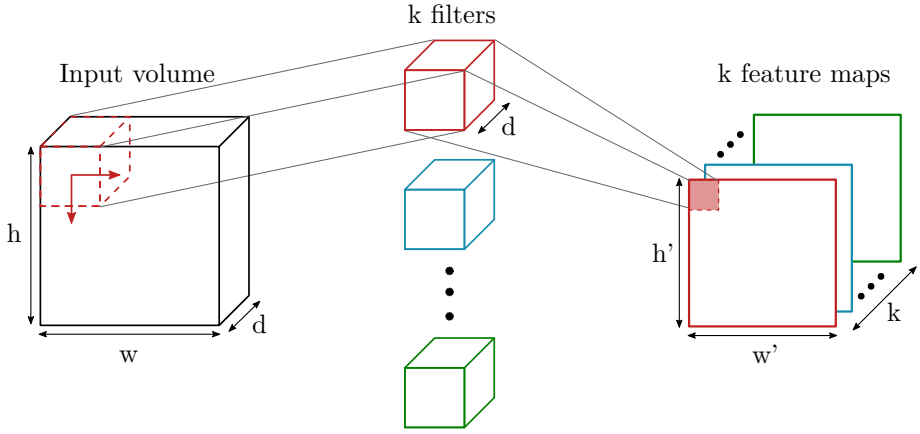


Figure 3.9: During the convolution operation dot product of the input volume and the filter is computed in a sliding window fashion resulting in a single feature map. Hence, the application of k filters will result in k feature maps.

where k is the index of the filter, $\mathcal{O}$(i,j,k) is the output of the convolutional layer, b_k is the bias of the kth filter and $\mathcal{F}_k$ is the filter that is convolved with the input in the first and second mode. This filter consists of M $\times$ N $\times$ P coefficients, where M and N represent the size of the filter, while P represents the number of feature maps in the previous layer. $\mathcal{O}$(i,j,k) is the output of convolutional layer and subsequently the input of the next layer.

The hyperparameters of the convolutional layer that have to be selected when designing the network are the number of filters (defining the depth of the output volume), the filter size, the stride and the amount of zero-padding. The dimension of the third mode of the output is defined by the number of filters, while the dimensionality in the first spatial modes depend on these hyperparameters according to the following formula:

$$O = \frac{I - F + 2Z}{S} + 1 \tag{3.37}$$

where O and I are the size (width/height) of the output and input, F is the filter size and Z and S correspond to the zero padding and stride, respectively.

Nonlinear activation layer Nonlinear activation layers are implemented in order to introduce nonlinearity into the network. The most popular activation function in CNNs is a rectified linear unit or ReLU. This activation function replaces each negative element by zero, which can be written as: $g(z) =$

$max(z, 0)$. This layer works element-wise and hence does not change the dimension of the input.

Pooling layer The pooling layer performs a nonlinear downsampling. Different functions can be used to progressively reduce the size of the input, the most common ones are averaging (avg pooling) and taking the maximum (max pooling) in a small subregion of the input volume. Pooling is performed on every depth slice independently, so it only affects the size of the input volume in the spatial modes [72]. The pooling layers enhance the translational invariance of the network [73]. This operation does not need any parameters, however the spatial extent and stride have to be set during the design of the network.

Fully connected layer The last layers of the CNN are usually fully connected, so each neuron is connected to each neuron of the previous layer. These fully connected layers use the extracted high level features to perform the classification. The activation function typically used in the final layer is a softmax function. Finally, the output of the network is a score with dimensions [1 × 1 × N] reflecting the probability of belonging to one of the N classes [72].

Training procedure Before the network is ready for use, the internal parameters of the machine have to be optimized using backpropagation during the training phase. This means that training examples are presented to the network (during the forward pass) and the loss function, a measure of the error between the class score of the network and the real class labels, is computed. Afterwards, the gradients are backpropagated via the chain rule during the backward pass. Eventually, the filter weights are adjusted via gradient descent in order to minimize loss across the training data set [72, 117]. Different variations of gradient descent exist. Optimization algorithms where the complete training set is used to update the weights is called batch gradient descent. In stochastic gradient descent (SGD) only a single training example is used to adjust the weights. Mini batch stochastic gradient descent is an in-between solution using a batch of training examples [73]. The learning rate defines how much the weights are adjusted during each optimization step. This is one of the hyperparameters that has to be chosen by the designer of the network [72].

3.3.4 Linear regression

Regression models are used to examine the relationship between a dependent variable and one or more independent variables. The most simple linear

regression model is a univariable regression, where the linear relationship between a single independent variable or predictor and the continuous dependent variable or response variable is fitted. Multiple regression refers to models with multiple independent variables [164, 171]. Note that these independent variables can be either continuous or categorical.
Generally, the multiple regression model describing the relation between the dependent variable **y** and the p independent variables **x** can be written as:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots \beta_p x_{ip} + \epsilon_i, \tag{3.38}$$

where i indicates the observation, i = 1, 2, ... n. There are p independent variables with a corresponding regression coefficient β. These regression coefficients or slopes describe the rate of change of the dependent variable per unit change of its corresponding independent variable, while holding all other predictors constant. β_0 is the intercept and ϵ represents the error term [164]. If $\hat{\mathbf{y}}$ are the regression model's estimated values, then the residual for the ith observation e_i is computed as:

$$e_i = y_i - \hat{y}_i. \tag{3.39}$$

These residuals measure the deviation of the observed data from the fitted regression line, as illustrated in Figure 3.10 [164]. The parameters of the regression model, the intercept and regression coefficients, are estimated in a least squares sense by minimizing the sum of squares of the residuals $SS_{res} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = e_i^2$.

The coefficient of determination or R^2 is a measure to assess how well the model can explain and predict the response variable. It describes the proportion of variation in the dependent variable explained by the best-fit regression line and is computed as:

$$R^2 = \frac{\text{explained variation}}{\text{total variation}} = \frac{SS_{reg}}{SS_{tot}} = 1 - \frac{SS_{res}}{SS_{tot}} \tag{3.40}$$

with the explained sum of squares SS_{reg} and the total sum of squares SS_{tot} equal to

$$SS_{reg} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \tag{3.41}$$

$$SS_{tot} = \sum_{i=1}^{n}(y_i - \bar{y})^2, \tag{3.42}$$

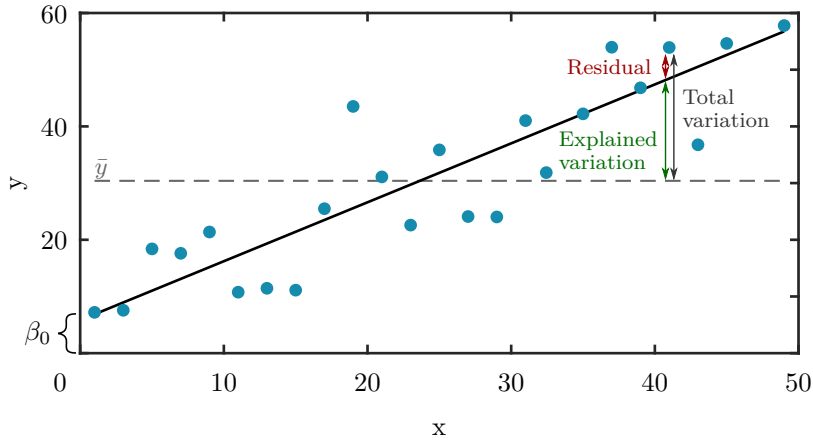


Figure 3.10: Illustration of a simple linear regression. The blue circles represent the data points and the solid black line the least squares regression line. The dashed line indicates the mean of the dependent variable Y. For one particular data point, the residual, variation explained by the model and total variation are indicated.

where $\bar{y}$ denotes the mean of the dependent variable among all observations.

The multiple linear regression described above assumes that all observations are independent. However, in case of repeated measurements, for example when multiple recordings are measured from the same subject, this assumption is violated. Therefore, a linear mixed-effects (LME) model containing both fixed and random effects should be used to account for this patient-to-patient variability [164].

3.4 Performance metrics

3.4.1 Classification

After feature extraction, selection and training a classification model, we need performance measures to evaluate how well the model can predict the class of new, unseen samples. Various metrics to assess the performance of a classifier have been developed. The performance measures used in this thesis will be

briefly discussed for binary classification problems, however they can be easily extended to multiclass problems.

Confusion matrix

Typically, the first step in evaluating the classification performance is to construct a confusion matrix. This matrix consists of four values: the number of true positives/negatives and false positives/negatives:

- True Positives (TP)/True Negatives (TN): these are samples that are correctly predicted as positive/negative by the model.
- False Positives (FP)/False Negatives (FN): these are samples that are incorrectly predicted as positive/negative by the model.

Sensitivity, specificity and accuracy

Based on these metrics the accuracy, sensitivity and specificity are usually computed:

- Accuracy is the ratio of the number of correctly predicted samples to all predicted samples:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}. \tag{3.43}$$

  However, it is not a reliable performance metric in case of imbalanced datasets (e.g. 99 percent of the dataset belongs to class 1 and 1 percent of the dataset belongs to class 2 results in an accuracy of 99 percent if all samples are classified as class 1).

- Sensitivity or recall is the fraction of positive samples that are correctly classified by the model:

$$\text{Sensitivity} = \frac{TP}{TP + FN}. \tag{3.44}$$

- Specificity is the fraction of negative samples that are correctly classified by the model

$$\text{Specificity} = \frac{TN}{FP + TN}. \tag{3.45}$$

Cohen's Kappa

As mentioned above, in case of imbalanced datasets the accuracy will be biased and is not a reliable performance metric. In contrast to accuracy, Cohen's Kappa coefficient is a performance measure that can handle imbalanced datasets. Cohen's Kappa score (κ) was originally proposed as a metric to assess the inter-rater agreement, but can also be employed to assess the agreement between the predicted labels of the model and the ground truth. This statistical measure takes the agreement occurring by chance into account, hence it is a more robust and more conservative measure compared to accuracy [35, 131]. It is computed according to the following formula:

$$\kappa = \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e} \tag{3.46}$$

where p_o and p_e represent the observed agreement and expected agreement, respectively. The observed agreement is computed as the accuracy in (3.43), while the expected agreement due to chance p_e is obtained as [175]

$$p_e = \frac{\frac{f_1 \times g_1}{n} + \frac{f_2 \times g_2}{n}}{n} = \frac{f_1 \times g_1 + f_2 \times g_2}{n^2}, \tag{3.47}$$

where n denotes the total number of samples. $f_{1/2}$ and $g_{1/2}$ represent the number of times rater 1 (f) and rater 2 (g) predicted class 1/2, respectively. So, Cohen's Kappa is indicating the classification model's performance corrected for chance agreement. Contrary to the previously described performance metrics, Cohen's Kappa is not a value between 0 and 1, but can range from -1 to 1. A Kappa statistic equal to 1 indicates a perfect agreement and 0 represents the agreement expected by chance. Therefore, an agreement worse than random chance is reflected in a negative Kappa value [131].

Receiver operating characteristic

However, the above mentioned performance criteria depend on the chosen decision threshold. To account for this, the receiver operating characteristic (ROC) curve can be computed. The ROC curve displays the true positive rate versus the false positive rate of the classification model for all classification thresholds. The higher the true positive rate and the smaller the false positive rate, the better the classifier. Therefore, the performance of the classification model is often measured as the area under the curve (AUC) of the ROC. It is a measure of the discrimination power of the classifier across all decision thresholds and its value is bounded between zero (worthless classifier) and one (perfect classifier).

3.4.2 Regression

Coefficient of determination

Various metrics exist to assess the fit of regression model. As proposed in the previous section, the goodness-of-fit of the regression line can be quantified using R-squared statistic computed as in (3.40). The coefficient of determination is in the range from 0 to 1. A model that fits the data well will have an R-squared close to 1, while an invalid model is characterized by an R-squared close to 0.

Root mean square error

Another commonly used performance metric is the root mean square error (RMSE). This metric evaluates how close the data points are to the least squares regression line by computing the standard deviation of the residuals:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{3.48}$$

where n represents the number of observations, y_i is the ith observation of the response variable and $\hat{y}_i$ its estimate. In contrast to the coefficient of determination this is an absolute measure of fit and is expressed in the same unit as the response variable. The lower the RMSE, the better the model can predict the response variable.

3.5 Conclusion

This chapter provided the necessary technical background of the methods used throughout the thesis. Time domain features assessing the continuity and the complexity of the EEG signal were described. Moreover, frequency domain features commonly used to classify neonatal sleep stages and to assess brain maturation are presented. Generalizations of vectors and matrices to higher dimensions, tensors, are introduced. Moreover, the canonical polyadic decomposition and its properties are discussed. Afterwards, different approaches for supervised classification and the basics of linear regression models are presented. At last, the metrics used to assess the performance of classification and regression models are introduced.

Part II

Automated neonatal EEG sleep staging

Chapter 4

Neonatal sleep stage classification based on EEG complexity features

As described in section 2.2.2, the sleep architecture of preterm infants changes gradually during early brain development. Therefore, automated classification of the neonatal sleep stages can provide information about the brain maturation of the infant. This chapter presents the first automated method for neonatal sleep stage classification proposed in this thesis. This algorithm follows the classical machine learning pipeline for classification problems. So, first a set of discriminative features is extracted from the preprocessed EEG time series. These features are then used to train a supervised learning model. Finally, the trained model can predict the labels of new, unseen test samples.
In this study, a novel set of features quantifying the complexity of the EEG signal is explored to discriminate quiet sleep from non-quiet sleep. This analysis is performed on a large database consisting of 97 EEG recordings from 26 prematurely born infants, measured between 27 and 42 weeks postmenstrual age. The results suggest that the complexity of the brain dynamics is different dependent on the neonate's vigilance state, that's why it is a highly useful feature to perform automated sleep stage classification. Chapter 5 builds on these findings and exploits the complexity of the EEG time series in a tensor framework to identify quiet sleep segments. A deep learning approach for neonatal sleep stage classification is presented in Chapter 6. The current sleep stage classifier based on EEG complexity features is thoroughly compared to the other proposed algorithms in Chapter 7. This work has been published in

modified form as part of [50].

4.1 Introduction

The sleep state organization of preterm infants undergoes fast development during their stay in the neonatal intensive care unit (NICU) and provides information about their functional brain maturation. For that reason, several algorithms have been developed for automated sleep stage classification in preterm neonates [56, 103, 162]. These algorithms typically rely on the difference in continuity or spectral content of the EEG time series to distinguish quiet sleep from non-quiet sleep. More recently, attention has focused on nonlinear dynamical analysis of neonatal EEG as well. Few studies have carried out a dimensional analysis of the neonatal EEG during sleep [88, 169]. According to Janjarasjitt et al. [89], the dimensional complexity tends to be higher during active sleep than during quiet sleep. In addition, the sample entropy of the neonatal sleep EEG has been explored by Zhang et al. [218]. They reported a higher sample entropy during active sleep compared to quiet sleep, which supports the findings of Janjarasjitt et al. [89].

The studies presented thus far provide evidence that the EEG complexity in neonates is dependent on the vigilance state. More specifically, that the complexity of the EEG is higher during active sleep compared to quiet sleep. In this study, another metric, multiscale entropy, will be employed to quantify the complexity of the EEG time series. This method calculates the sample entropy over multiple time scales, exploiting the fact that complex systems are characterized by dynamics at multiple scales.
The purpose of this study is to investigate whether the EEG complexity quantified using multiscale entropy is indeed different depending on the sleep state of the neonate and can thus be used for automated sleep stage discrimination. Therefore, a sleep stage classifier based on complexity features derived from the multiscale entropy curve is developed.

The remainder of this chapter has been organised as follows. First, an overview of the database and the preprocessing of the EEG is given. Secondly, the extraction of EEG complexity features is described. In the following section, the classification model will be thoroughly explained. Finally, the findings of this research will be presented and discussed.

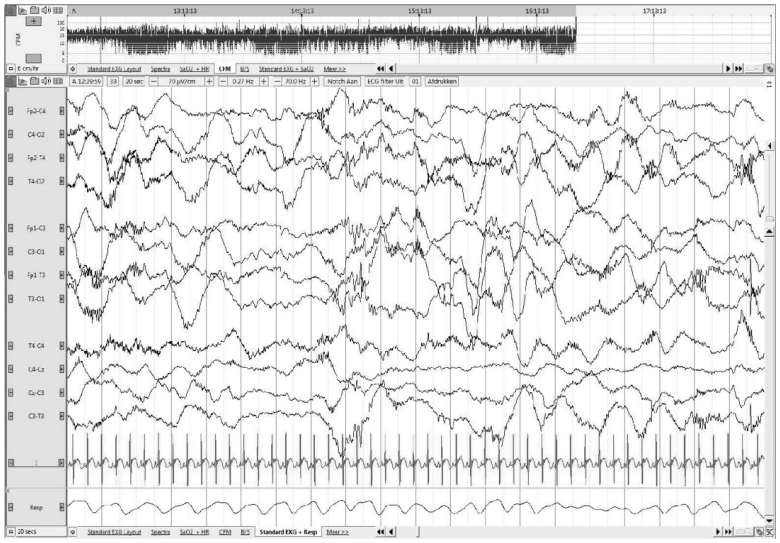
4.2 Materials and methods

4.2.1 Database

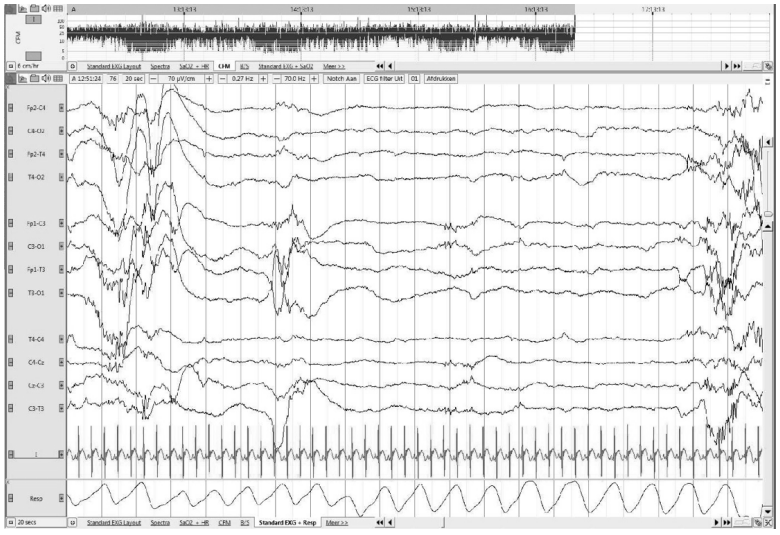
The dataset used in this study consists of 97 multichannel EEG recordings from 26 preterm neonates with a postmenstrual age (PMA) ranging from 27 weeks to 42 weeks (33.96 $\pm$ 3.33 weeks). The neonates included in this study were born prematurely before 32 weeks of gestation (28.74 $\pm$ 2.33 weeks). On average, the first EEG recording of each patient was measured 2 weeks after birth (2 days – 5 weeks plus 3 days), and serial recordings were performed with at least two recordings per subject. All babies recruited for this study had a normal neurodevelopmental outcome score at 9 and 24 months corrected age (Bayley Scales of Infant Development-II, mental and motor score > 85). Moreover, none of the subjects were under sedative or anti-epileptic medication during the EEG registration or had severe cerebral lesions (normal cerebral ultrasonography or intraventricular haemorrhage grade $\leqslant$ II, no periventricular leukomalacia or ventricular dilatation $>$ p97).

The EEG signals were recorded between 2012 and 2014 at the Neonatal Intensive Care Unit of the University Hospitals Leuven, Belgium. Informed parental consent was obtained for each recruited patient and the study was approved by the Ethics Committee of the University Hospitals of Leuven. The EEG was recorded using nine electrodes: F1, F2, C3, C4, T3, T4, O1, O2 and reference electrode Cz, according to the modified international 10-20 system [30]. The monopolar EEG set up was used and the reference electrode Cz was disregarded during the analysis phase. All EEG time series were acquired using BrainRT equipment (OSG bvba, Rumst, Belgium). The data were initially filtered between 0.3 and 70 Hz and sampled at 250 Hz. Neither preselection of data segments, nor artifact rejection has been applied.

The duration of the EEG recordings ranged from 2 h to 25 h 57 min, with an average recording length of 6 h 57 min. Two independent expert clinicians carried out annotation of quiet sleep segments upon consensus based on visual analysis of video EEG. Other vigilance states, such as active sleep and awake, are considered non-quiet sleep. An example of a labelled non-quiet sleep and quiet sleep EEG segment at a postmenstrual age of 32 weeks are presented in Figure 4.1. In some recordings (mainly in the long EEG measurements from 36 weeks PMA onwards), the quiet sleep (QS) labelling was stopped before the end of the recording. As a result, the average length of the labelled EEG recordings is 5 h 4 min (1 h 33 min – 10 h 9 min).



(a) Non-quiet sleep



(b) Quiet sleep

Figure 4.1: Example of a non-quiet sleep and quiet sleep EEG segment at 32 weeks and 2 days PMA. (a) Continuous tracing during active sleep. Delta brushes in temporal and occipital regions, irregular breathing pattern. (b) Discontinuous tracing during quiet sleep. IBI shorter than 15 s, temporal theta activity and occipital delta brushes, more regular breathing pattern.

4.2.2 Preprocessing

The first preprocessing step consisted of resampling all time series to 125 Hz to reduce the computational time. Next, the EEG signals were bandpass filtered between 1 and 20 Hz, in order to avoid distortion by artifacts while retaining most of the neural activity. The filter was applied twice, once forwards and once reversed, resulting in zero phase filtering.

4.2.3 Multiscale entropy computation

A well-known method to assess the complexity of physiological signals is multiscale entropy. Multiscale entropy, developed by Costa et al. [39], quantifies the degree of irregularity of the time series across multiple scales. In contrast with the traditional single-scale entropy measures, multiscale entropy can account for structures with long-range correlations on multiple time scales present in complex systems. As a consequence, it will not assign a high complexity to completely random signals which are not truly complex [38, 39].

The procedure to compute the multiscale entropy (MSE) of a signal consists of two steps. The signal is first coarse-grained according to the scale factor τ. This coarse-graining consists of averaging all data points within nonoverlapping windows of length τ. Next, the sample entropy of each coarse-grained time series is computed. In this study, the multiscale entropy is computed for scales from 1 to 20. The parameters used to compute the sample entropy were set as follows: the embedding dimension m was chosen equal to 2 and tolerance r was set as 0.2 $\times$ standard deviation of the EEG segment [125, 218]. Eventually, the output of the multiscale entropy analysis is a curve showing sample entropy in function of the scale factor τ, referred to as the multiscale entropy curve. For a more elaborate description of the MSE computation, the reader is referred to section 3.1.2.

Prior to computing the multiscale entropy of the neonatal EEG time series, the signals were segmented into epochs of length N. There are multiple factors to take into account when optimizing the window length. To acquire a reliable entropy estimate, the number of data points should be large enough [39]. If the window length is too short, it is possible that no template matches will be found, hence the entropy will be undefined. An important consideration in the choice of the window length N is that the length of the coarse-grained sequence reduces progressively with increasing scale τ. In this study, the largest analysed scale is equal to 20, thus the shortest sequence will have length $\frac{N}{20}$. Figure 4.2a shows the multiscale entropy of a neonatal EEG recording (channel C3) during quiet sleep evaluated at three different window lengths: 10 s (1250 samples),

100 s (12,500 samples), and 200 s (25,000 samples). The graph shows that from a certain window length on, the mean is no longer affected, while the standard deviation slightly reduces for longer epochs. However, the computational time is inversely proportional to the window length. As a result, the choice of window length is a trade-off between accuracy, robustness, and computational time. Since the reduction in standard deviation for 200 s versus 100 s is negligible and the computational time is reasonable, the window length is chosen equal to 100 s. As a consequence, the shortest coarse-grained time series (at $\tau = 20$) will consist of 625 samples, which complies with the rule that at least $10^m - 20^m$ data points are required to obtain a robust estimate of the sample entropy [165].

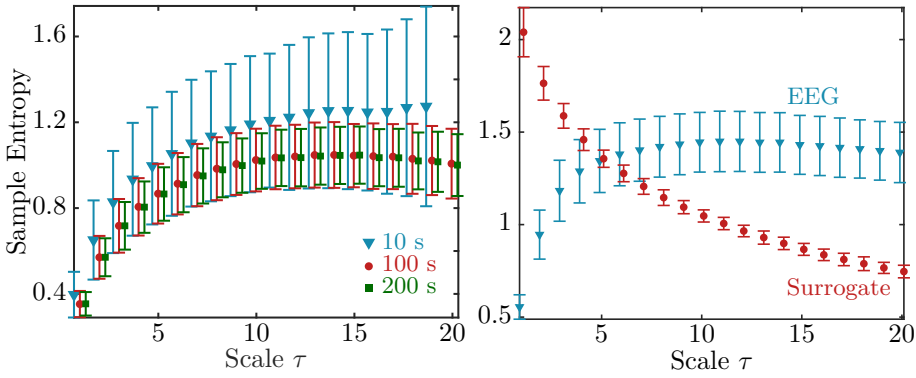
Moreover, a surrogate time series was constructed by randomly shuffling the time order of the data points within each 100 s segment of an EEG recording. In this way, the temporal correlations of the signal are destroyed, while the statistical properties of the distribution are preserved [40]. The symbols and error bars in Figure 4.2b show the mean and standard deviation of the multiscale entropy for both the original EEG recording and its surrogate. The multiscale entropy curve of the surrogate time series resembles that of white Gaussian noise, having a high sample entropy at small scales, but quickly decreasing with increasing scale factor. Although the sample entropy of the surrogate is greater than that of the original time series at small scales, the physiologic time series is more complex than the surrogate as can be seen from the greater entropy values at larger scales [40].

4.2.4 Feature extraction

After quantifying the multiscale entropy of each multichannel 100 s EEG segment, four features were extracted from the multiscale entropy curve: (1) the area under the multiscale curve (this will be referred to as the complexity index), (2) the average slope of the multiscale entropy curve in the small scales (scale 1 – 5), (3) the average slope of the curve in the large scales (scale 6 – 20), and (4) the maximum value of the multiscale entropy curve. Thus, in total, a set of 32 (8 channels $\times$ 4) features are extracted. These EEG complexity features are then used to train a classification model to discriminate quiet sleep from non-quiet sleep. The model and training procedure will be described in detail in the next section.

4.2.5 Classification model and training procedure

To investigate how powerful the complexity features are in identifying neonatal sleep stages, a sleep stage classifier was developed using LS-SVMlab [43, 186].



(a) Window length optimization (b) Surrogate testing

Figure 4.2: (a) The multiscale entropy computed using three different window lengths: 10 s (blue triangles), 100 s (red stars), and 200 s (green squares). The symbols represent the mean and the error bars the standard deviation. The consistency of the entropy estimation is lost if the window length is too short. The sample entropy is even undefined at scale 20 for a window length of 10 s. (b) The blue triangles and error bars represent the mean and standard deviation of the multiscale entropy of all 100 s segments of an EEG recording, respectively. The red circles and error bars show the mean and standard deviation of the randomly shuffled surrogate of this EEG time series. The entropy of the original time series is greater than that of the surrogate at large scales, indicating more complex dynamics.

The complete feature matrix $\mathbf{X} \in \mathbb{R}^{N \times d}$ consists of $N = 17{,}600$ data points ($N_{QS} = 4356$, $N_{NQS} = 13{,}244$) and $d = 32$ features. To cope with the large number of data points, a fixed-size least squares support vector machine (LS-SVM) was adopted. This method selects a (small) fixed number of training data points M ($M \ll N$) representing the underlying distribution of the dataset through maximization of the quadratic Renyi entropy [65]. A Radial Basis Function (RBF) kernel was used during this active selection of the support vectors, and its bandwidth parameter was computed according to the rule of thumb: $\sigma^2 = \text{factor} \times d \times \text{mean}[\text{var}(\mathbf{X})]$, where d is equal to the number of features, and $\text{var}(\mathbf{X}) = [\text{var}(\mathbf{x}_1)\ \text{var}(\mathbf{x}_2)\ \ldots\ \text{var}(\mathbf{x}_d)]$ with $\mathbf{x}_i \in \mathbb{R}^N$ [203]. The factor was tuned experimentally and set equal to 0.1. The number of support vectors M that are selected in an iterative way was set equal to 1500.

As the data are recorded in a wide range of age and we expect the maturational effect to play a role during sleep stage classification, the observations were

divided into three groups according to their PMA: brain activity measured before 31 weeks ($N_{<31} = 2395$), in the range from 31 to 37 weeks ($N_{31-37} = 10{,}901$), and EEGs recorded beyond 37 weeks PMA ($N_{>37} = 4304$) [111]. To assure that not all training data points were drawn from 1 sleep state, the number of training datapoints selected from a specific sleep state was proportional to its representation in the complete dataset. Similarly, the ratio of the number of observations for a specific age group was preserved.

The hyperparameters of the LS-SVM model with Gaussian RBF kernel were then tuned using 10-fold cross-validation, resulting in a regularization parameter $\gamma = 52.02$ and a squared bandwidth $\sigma^2 = 283.88$.

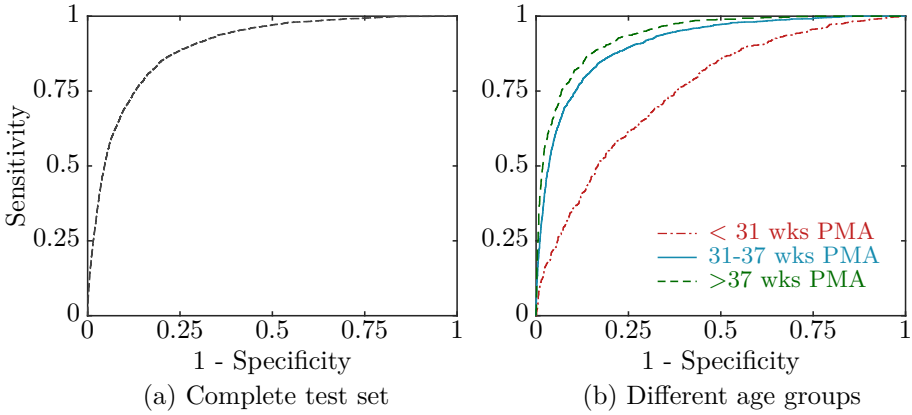
After training the LS-SVM classifier, the classification performance was tested on all remaining samples. As the complexity of the appearance of the EEG is expected to evolve throughout age as well, we are interested in both the overall classification performance and in the classification power for each of the age groups separately. The receiver operating characteristic (ROC) curves will be constructed for the four test sets (PMA $<$ 31 weeks, PMA: 31 – 37 weeks, PMA $>$ 37 weeks, complete test set) and the performance will be measured as the area under the ROC curve (AUC).

4.3 Results

The performance of the LS-SVM classifier is shown by means of ROC curves in Figure 4.3. The left panel shows the ROC curve when the classifier is evaluated on the complete test set consisting of all EEG segments not used for training the model. The area under the ROC curve is 90%, indicating a good classification performance. Figure 4.3b illustrates how the classifier performs on the different age groups. The best performance is reached for the oldest group (PMA $>$ 37 weeks) with an AUC of 94%, the middle group (PMA: 31 - 37 weeks) has an AUC of 91%, and the classifier performed worst at the youngest age (PMA $<$ 31 weeks) with an AUC of only 76%.

4.4 Discussion

In this study the complexity of the neonatal EEG is quantified by means of multiscale entropy. Features extracted from the multiscale entropy curves were then used to build a neonatal sleep stage classifier. The performance of the trained fixed-size LS-SVM classifier was assessed on a large test set of EEG recordings measured in a wide range of PMA.



(a) Complete test set (b) Different age groups

Figure 4.3: (a) The receiver operating characteristic (ROC) curve, with an area under the curve (AUC) of 90%, showing the performance of the sleep stage classifier on the complete test set. (b) The ROC curves when the least squares support vector machine (LS-SVM) classifier was applied on recordings from the different age groups. The red curve is the ROC curve of the recordings below 31 weeks PMA and has an AUC of 76%. The blue curve represents the ROC of the recordings between 31 and 37 weeks PMA. For this age group the AUC is 91%. The green curve shows the ROC for all recordings of the oldest group. The classifier performed best on this age group with an AUC of 94%.

The high performance of the proposed sleep stage classifier indicates that the complexity of brain dynamics exhibit fundamental differences between vigilance states in preterm infants. These results corroborate the findings of previous work [89, 218] and contribute additional evidence suggesting that brain dynamics are different in quiet sleep compared to non-quiet sleep.

It is interesting to note that the performance of the proposed sleep stage classifier depends on the PMA. The classifier reaches the highest performance (AUC = 0.94) on neonates near term age and has the lowest performance on EEG recordings of very young neonates (AUC = 0.76). Note that the performance on the complete test set is biased as the majority of the EEG recordings are measured between 31 and 37 weeks PMA. Therefore, it is no surprise that the classifier on the complete test set (AUC = 0.90) has approximately the same performance as the classifier tested on the middle group (AUC = 0.91).
Even though sleep state organization starts at around 27 weeks, sleep–wake cycles can only be reliably differentiated from the EEG at around 31 weeks PMA [30]. Therefore, sleep stage discrimination below 31 weeks is challenging both for clinicians and for the algorithm. This might explain why the worst performance

is obtained for EEG signals recorded before 31 weeks PMA. Another factor that may explain the low performance in the young neonates is the imbalance of the age groups. With only 14% of the training data coming from recordings of the youngest age group, this group is underrepresented in the training set. As a consequence, the classifier might be biased towards other, more represented, age groups. In future studies this problem can be tackled by balancing the age groups in the training data. Moreover, the use of weighted LS-SVM classifiers could be explored in order to deal with this data imbalance problem. In contrast with the approach taken here, it is also possible to train a separate classifier for the different age groups. As the models can learn age-specific patterns, they will probably outperform the presented classifier. However, since these models will then be tailored to a specific PMA range, they will be less useful in clinical practice. Besides, the current study was limited by the fact that the performance was evaluated on all data points not used for training. This means that some EEG segments of a particular recording are used for training and others for testing. As a result, the classifier might be able to learn patient specific patterns during the training procedure. To get an idea of the performance on completely unseen data, the algorithm should be evaluated on a patient independent division of the dataset into training and test set. This analysis will be carried out and discussed in Chapter 7.
Future research should also be undertaken to investigate the influence of the number of selected support vectors in the fixed-size LS-SVM on the performance. At last, only EEG complexity features were considered to develop the proposed algorithm. Yet, previous studies have demonstrated that other EEG features, such as the (relative) power in specific frequency bands or measures of the continuity of the EEG, are suitable to perform sleep stage classification. Therefore, future studies should focus on combining all established discriminative features to design a robust and reliable neonatal sleep stage classifier.

4.5 Conclusion

In this chapter, we investigate how the complexity of the EEG can be used to classify sleep stages in preterm infants. An LS-SVM classifier relying on solely EEG complexity features could identify neonatal sleep stages with an AUC of 90%. The performance has also been evaluated on three different age groups. The highest performance was obtained for the oldest neonates (AUC = 94%), followed by the measurements between 31 and 37 weeks PMA (AUC = 91%) and the lowest performance was observed for the very young infants (AUC = 76%).

Chapter 5

Decomposition of a multiscale entropy tensor for sleep stage identification in preterm infants

In the previous chapter we presented an algorithm for automated classification of neonatal sleep stages based on EEG complexity features. The high performance of the classifier indicated that the EEG complexity of preterm neonates depends on the infant's behavioural state. This chapter relies on these results and exploits tensor algebra to perform neonatal sleep stage identification. The multiscale entropy of each EEG recording is computed to represent the EEG signal as a tensor. The decomposition of this multiscale entropy tensor will then be used to extract the sleep-cycling of the infant. In contrast to the previous chapter, this tensor-based algorithm is unsupervised, which allows to apply it in other clinical centers using different EEG acquisition setups. The performance of this novel tensor-based algorithm will be assessed for different ranks on a database of 97 neonatal EEG recordings. This chapter is based on [51].

5.1 Introduction

In human infants, the emergence of sleep cycles occurs at approximately 26 to 28 weeks postmenstrual age (PMA) [75]. During maturation of the sleep architecture, the distribution and duration of specific sleep states change gradually. Very young preterm infants have an abundant amount of sleep, and active sleep is the dominant sleep stage. From then on, the proportion of time spent asleep decreases, while the relative amount of quiet sleep increases. Near term age, both active sleep and quiet sleep constitute approximately half of the total sleep time [27, 74]. Existing research recognises the importance of sleep in early brain development [27, 56, 75]. Sleep and established sleep cycling play a vital role in normal neurosensory development, learning processes, memory consolidation and in the protection of the infant's brain plasticity [75]. Moreover, studies such as that conducted by Shellhaas et al [172] have shown that the presence of sleep cycling and the quantity and quality of each sleep state are associated with neurodevelopmental outcomes [12, 97, 148].

Most prematurely born infants stay in the neonatal intensive care unit (NICU) during the first critical weeks of rapid growth and development of the brain. In the NICU, neonates are exposed to a myriad of noxious environmental stimuli, such as high noise and light levels and painful procedures, which might disrupt their sleep state organization. In recent years, there has been an increasing interest in strategies to promote sleep in the NICU environment (e.g., kangaroo care, massage therapy, cycle lighting, etc.) [12, 196].
Real-time automated identification of behavioural states can be used to optimize the planning of NICU caregiving in order to reduce disturbance of sleep-wake cyclicity [210]. Moreover, an automated sleep staging algorithm can be used to assess the sleep architecture and by that the functional brain maturation. In view of all that has been mentioned so far, one may suppose that there is a need to assess the sleep staging of neonates in order to provide developmentally appropriate care.

A number of algorithms for sleep stage classification in preterm neonates have been developed. The majority of these approaches are supervised and combine a set of electroencephalography (EEG) features (e.g., temporal features, spectral features, spatial features, complexity features) with a classification algorithm [50, 103].
This chapter proposes a novel unsupervised method to discriminate quiet sleep from non-quiet sleep in preterm infants. In this study, a tensor-based method exploiting the differences in EEG complexity between different vigilance states will be used. Due to the increasing amount of data being collected, and the specific properties of tensor decompositions, multiway analysis has received increasing attention during recent decades. Tensor algebra has been used in

a broad range of applications, such as image and video processing, machine learning and biomedical applications [31, 220]. To the best of our knowledge, this is the first study where tensor decompositions are used to discriminate sleep stages in preterm neonates. Therefore, this research can serve as a proof of concept and illustrate how tensor decompositions can be used in biomedical applications, and more specifically in a classification problem based on neonatal EEG.

The remaining part of this chapter proceeds as follows. The first section will describe the dataset. It will then go on to the explanation of the different steps of the proposed method. Afterwards the results of the algorithm will be reported and discussed.

5.2 Database

The proposed method is evaluated on the same dataset as used in Chapter 4. A brief overview of the database will be given here, more details can be found in section 4.2.1. All neonates included in the study were born at a gestational age below 32 weeks and had multiple EEG measurements at a postmenstrual age (PMA) between 27 and 42 weeks. In total 97 EEG measurements with a monopolar setup of eight EEG channels (Fp1, Fp2, C3, C4, T3, T4, O1, O2) were used in the analysis. Quiet sleep periods were annotated by two independent expert clinicians upon agreement. All other sleep states are merged and will be referred to as non-quiet sleep. The goal of the proposed algorithm is to automatically label EEG segments as either quiet sleep (QS) or non-quiet sleep (NQS). The analysis of the data was carried out in Matlab 2017b (The MathWorks, Inc., Natick, MA, USA), and the tensor decompositions were performed using Tensorlab [205].

5.3 The proposed tensor-based sleep stage identification method

The pipeline of the proposed algorithm consists of the following six steps: (1) Preprocessing of the EEG; (2) Assessment of the EEG complexity via computation of the multiscale entropy; (3) Tensorization of each EEG recording, (4) Decomposition of the multiscale entropy tensor; (5) Selecting the component of interest; and (6) Postprocessing and clustering of the temporal signature. Each step of the algorithm will be extensively described in the next paragraphs.

Finally, the metrics to assess the classification performance and the statistical testing will be explained.

5.3.1 EEG preprocessing

In order to avoid distortion of the EEG time series by high or low frequency artifacts, a finite impulse response bandpass filter between 1 and 40 Hz was applied on each EEG channel. Moreover, an additional notch filter at 50 Hz is used to remove any remaining powerline interference. The EEG signal is then downsampled by a factor of two to reduce the computational complexity.

5.3.2 Multiscale entropy computation

After filtering the data, the complexity of the EEG signal is assessed by means of multiscale entropy. Therefore, the multichannel EEG is segmented into nonoverlapping windows of 100 s as decided in section 4.2.3. Per channel, the multiscale entropy of each 100 s EEG segment is then computed for scales τ from 1 to 20. For the computation of the sample entropy, the template length m is set equal to 2 and the tolerance is defined as 0.2 $\times$ the standard deviation of the original time series. Once the sample entropy is computed for each of the coarse-grained time series, a multiscale entropy (MSE) curve can be constructed. This curve shows the sample entropy in function of the scale factor τ. Hence, it reflects the regularity of the signal across multiple scales.
Figure 5.1 presents an example of the average multiscale entropy curve of all quiet sleep and non-quiet sleep segments within a single EEG recording in green and pink, respectively. At all scales the sample entropy is lower during quiet sleep compared to non-quiet sleep, which indicates a reduced EEG complexity during quiet sleep. The algorithm will rely on this property to discriminate quiet sleep from non-quiet sleep segments.

5.3.3 Tensorization

The entropy values of each EEG recording are then organized in a third order tensor with modes: channels, scales and time segments. So, the multiscale entropy curves of consecutive time segments are stacked in the third mode of the tensor $\mathcal{X} \in \mathbb{R}^{N \times S \times T}$. Using this tensorization, the structural information among the leads is preserved. In this study, the number of EEG channels N is equal to 8, the number of scales S for which the sample entropy is computed is equal to 20 and the number of time segments T is variable depending on

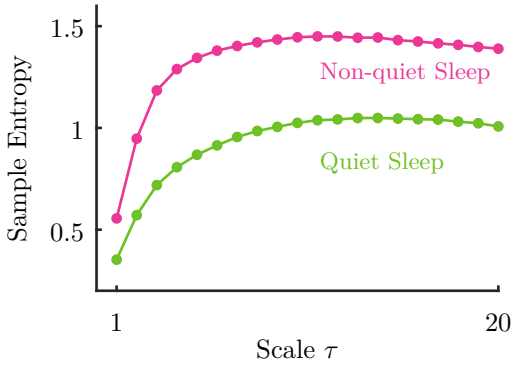


Figure 5.1: The pink curve represents the average multiscale entropy curve of all non-quiet sleep segments of an EEG recording measured at 32 weeks and 5 days PMA. The green curve shows the average multiscale entropy curve of all quiet sleep segments within the same recording. The multiscale entropy curve is lower during quiet sleep compared to non-quiet sleep, indicating reduced EEG complexity during quiet sleep.

the length of the EEG recording. Thus, the data of each EEG recording is transformed into a third order tensor $\mathcal{X} \in \mathbb{R}^{8\times 20\times T}$, where each row fiber (mode-2 fiber) of the tensor represents a multiscale entropy curve from a specific EEG channel and time segment. In the remainder of this paper, this tensor will be referred to as the multiscale entropy tensor.

5.3.4 Tensor decomposition

The canonical polyadic decomposition (CPD) or parallel factor analysis (PARAFAC) of a rank-R tensor $\mathcal{T}$ factorizes the tensor in a sum of R rank-1 tensors [101]. This can be written as:

$$\mathcal{T} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \tag{5.1}$$

where the factor vectors $\mathbf{a}_r, \mathbf{b}_r$ and $\mathbf{c}_r$ are the rth $(1 \leqslant r \leqslant R)$ column of the factor matrices $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$, respectively. The factor matrices $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ are obtained by solving the least-squares optimization problem with objective function:

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \ \frac{1}{2} \|\mathcal{T} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]\|_F^2 . \tag{5.2}$$

The advantage of the CPD compared to matrix factorizations is that the decomposition is unique (up to scaling and permutation ambiguity) under mild conditions [59]. In addition, prior knowledge of the data properties can easily be taken into account by imposing constraints on the factor matrices (e.g., sparsity, smoothness, nonnegativity, etc.) [31]. Because entropy values are positive, nonnegativity is enforced during the decomposition of the multiscale entropy tensor. Moreover, the nonnegativity constraint circumvents the occurrence of degenerate components.

The CPD of the third-order multiscale entropy tensor with rank R will result in three nonnegative factor matrices: the spatial signatures will form the columns of $\mathbf{A} \in \mathbb{R}^{N \times R}$, the scale signatures the columns of $\mathbf{B} \in \mathbb{R}^{S \times R}$ and the temporal signatures the columns of $\mathbf{C} \in \mathbb{R}^{T \times R}$. An example of this decomposition is shown in Figure 5.2. The factor vectors in the first mode $\mathbf{a}_r$, will show the variation over the different EEG channels, the factor vectors in the second mode $\mathbf{b}_r$, contain information about the distribution over scales, while the factor vectors in the third mode $\mathbf{c}_r$, will capture the variation of the EEG complexity over the different time segments.

Starting from the assumption that the EEG complexity is different depending on the neonate's sleep stage [50], we expect that one of the temporal signatures will reflect the sleep cycling. Since the goal of the algorithm is to perform automated sleep staging, only the factor vectors in the third mode, the temporal signatures, are of interest.

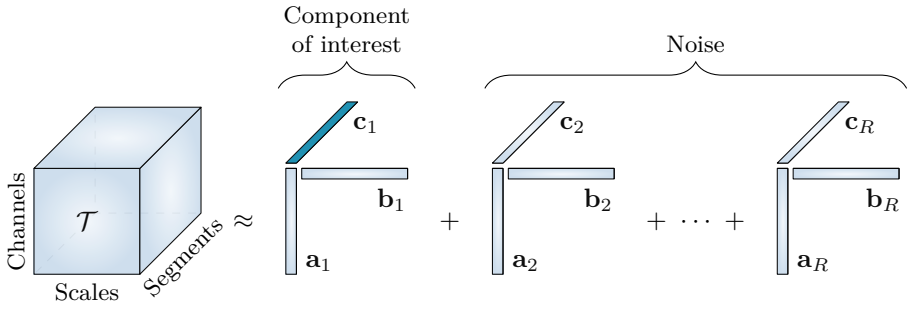


Figure 5.2: The rank-R canonical polyadic decomposition of a multiscale entropy tensor $\mathcal{T}$. When the rank R is greater than 1, the component of interest related to the neonatal sleep staging has to be selected. The temporal signature of interest $\mathbf{c}_1$, which is highlighted in dark blue, will then be used to discriminate the neonatal sleep stages. All other components will be discarded during further analysis.

Detection of stable solution

Due to the fact that (5.2) is a non-convex optimization problem, we may end up in a local minimum rather than a global minimum. Thus, different factor matrices can be obtained depending on their initialization [197]. To ensure that a reliable solution is found, the decomposition is repeated 50 times with different random initializations. Next, we want to detect and select the most occurring, stable solution among these 50 repetitions. To this end, the similarity between all possible combinations of components is assessed by means of cosine similarity. In this study we assume that each component should match in all its modes, so the triple cosine product or congruence is computed. To assess the similarity of two rank 1 tensors $\mathcal{X} = \mathbf{k} \circ \mathbf{l} \circ \mathbf{m}$ and $\mathcal{Y} = \mathbf{p} \circ \mathbf{q} \circ \mathbf{r}$ this can be written as [22, 124, 193]:

$$\text{cong}(\mathcal{X}, \mathcal{Y}) = \cos(\mathbf{k}, \mathbf{p}) \cos(\mathbf{l}, \mathbf{q}) \cos(\mathbf{m}, \mathbf{r}) = \frac{\mathbf{k}^T \mathbf{p}}{\|\mathbf{k}\| \, \|\mathbf{p}\|} \frac{\mathbf{l}^T \mathbf{q}}{\|\mathbf{l}\| \, \|\mathbf{q}\|} \frac{\mathbf{m}^T \mathbf{r}}{\|\mathbf{m}\| \, \|\mathbf{r}\|}. \quad (5.3)$$

The R components of each repetition are sorted based on (5.3) to account for the permutation indeterminacy of the decomposition. Subsequently, the similarity between the corresponding components of different repetitions is investigated. On the assumption that all R components should match, the congruence of the R components is multiplied. This can be organized in a symmetric similarity matrix $\mathbf{S} \in \mathbb{R}^{it \times it}$, where it represents the number of iterations (in this case 50). So, s_{ij} represents how similar the factor matrices of the ith and jth iteration are. Finally, the solution of the iteration with the highest similarity to the other repetitions is selected as the stable and reproducible solution.

Number of components

One of the principal challenges when applying tensor decompositions is selecting an appropriate number of components for the problem at hand [174]. A multitude of strategies to define the rank have been proposed. First of all, we investigated the multilinear singular value spectrum in order to get an initial estimate of the rank. Then, we performed the decomposition for rank R going from 1 to 5. To compare the different ranks, we investigated the core consistency diagnostic (CORCONDIA), which assesses how close the core tensor is to being superdiagonal. In addition, a diagnostic based on the relative reduction of the fitting error for an increase of the rank, called the difference of fit (DIFFIT) [190], was computed to assist in determining the appropriate number of components.

5.3.5 Selection of the component of interest

When decomposing the multiscale entropy tensor, we expect that one component will reflect the sleep staging of the neonate. However, if the number of components R is greater than one, an automatic selection of the component of interest is required. So, the goal is to find the temporal signature related to sleep staging. In Figure 5.2 this is the temporal signature of the first component, $\mathbf{c}_1$, which is marked in blue.

We expect that the component related to sleep staging will have a more regular, cyclic pattern compared to the other (noise) components. Therefore, the autocorrelation function of the temporal signatures are computed. Since the temporal signature reflecting the sleep staging has a stronger correlation in time, we expect that its area under the absolute value of the autocorrelation function (ACF) will be larger [52, 100]. Therefore, the component whose area is the largest will be selected as the component of interest and used for further processing.

Figure 5.3 shows an example of this procedure for a rank-2 CPD of a multiscale entropy tensor. The autocorrelation of the two temporal signatures is plotted on top. The area under the absolute value of the autocorrelation is equal to 31.85 and 13.82 for the temporal signature of the first (blue) and second (red) component, respectively. Therefore, the first component will be selected as the component of interest. The temporal signatures of the two components are plotted at the bottom with highlighted periods of quiet sleep in light grey. From these graphs we can see that the first component is indeed reduced during quiet sleep segments compared to non-quiet sleep segments, whereas the second component is not related to the sleep stages.

5.3.6 Postprocessing and clustering

Once the nonnegative polyadic decomposition of the multiscale entropy tensor is performed, the temporal signature of the selected component is used to define the neonate's sleep stage. This temporal signature shows a (slow) cyclic pattern reflecting the sleep staging of the infant. However, there are high frequency oscillations superimposed on this pattern, which could lead to incorrect sleep stage identification. Therefore, a postprocessing step consisting of smoothing the temporal signature using a moving average filter with length L equal to 5 is applied. This moving average filter is applied in both directions (to avoid phase distortion), so actually a weighted moving average filter is used with triangular shape and length $2 \times$ L. This smoothing operation accounts for the fact that a sleep stage does not change instantly. More specifically, sleep periods were only

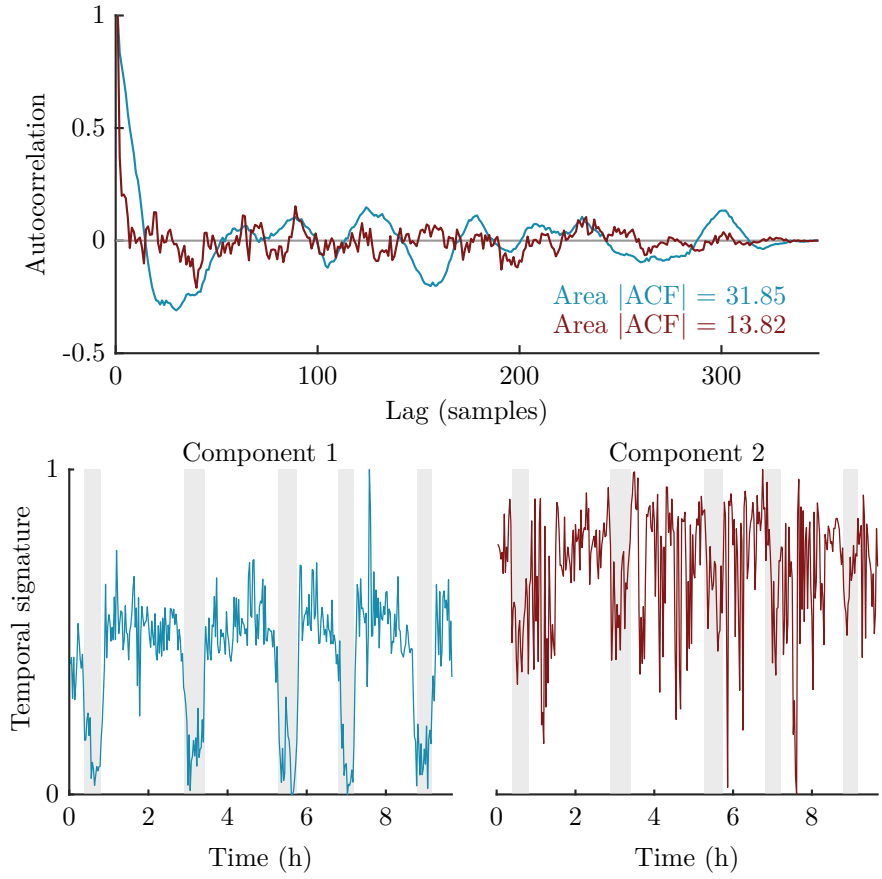


Figure 5.3: Illustration of the component selection procedure. On top, the autocorrelations of the temporal signatures of a rank-2 canonical polyadic decomposition (CPD) of a multiscale entropy tensor are shown. The area under the absolute value of the autocorrelation (ACF) is equal to 31.85 and 13.82 for the first component (in blue) and the second component (in red), respectively. Hence, the first component will be selected as the component of interest. At the bottom, the corresponding temporal signatures are plotted with the quiet sleep periods annotated by the clinicians highlighted in light grey. The temporal signature of the first component shows a clear reduction during quiet sleep, while the second component is not correlated with the sleep cycling.

visually labelled as quiet sleep if 3 consecutive minutes or 3 out 4 min were clinically detected as quiet sleep.

In order to divide the data then into two distinct clusters, k-means clustering (k = 2) is performed using the smoothed temporal loading. So, each data point of the smoothed temporal signature of interest $\mathbf{c} \in \mathbb{R}^{T\times 1}$ will be assigned to one of the two clusters. Thus, the result of the clustering is a vector of length T containing a cluster label for each of the 100 s EEG segments. Since k-means clustering is heavily dependent on its initialization, k-means clustering is repeated 100 times with different initial cluster centroids and the clustering with the lowest sum of within-cluster point-to-centroid distances is selected. The cluster with the lowest EEG complexity will be assigned the label of quiet sleep, while the other cluster is labelled as non-quiet sleep [50, 89].

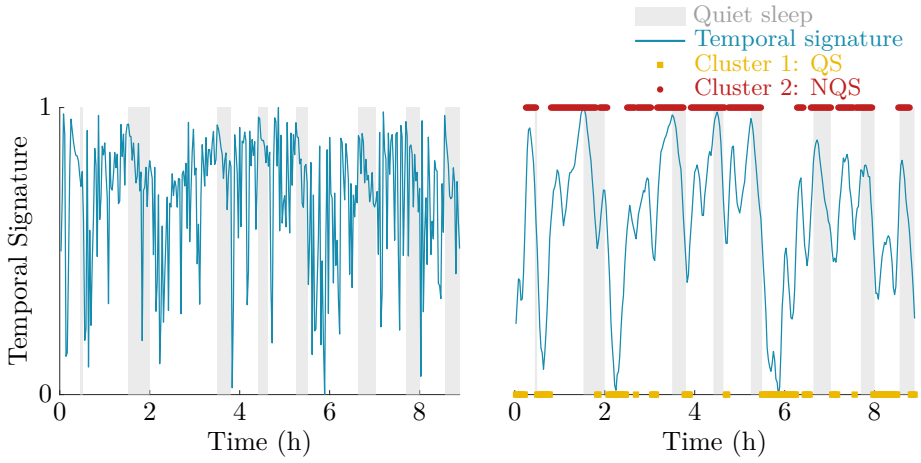
The effect of the smoothing and the result of the clustering is illustrated in Figure 5.4. In this example, the multiscale entropy tensor of an EEG recording is decomposed for rank R equal to 1 and 2. The raw temporal signatures of the rank-1 and rank-2 decomposition (only the temporal signature of interest) are shown on the left, while their smoothed versions are plotted on the right. The smoothing clearly removes the unwanted variations. The sleep labels obtained by the algorithm based on k-means clustering are marked by yellow squares (cluster 1 corresponding to quiet sleep) and red dots (cluster 2 corresponding to non-quiet sleep). The rank-1 CPD does not correlate well with the clinical annotations of quiet sleep highlighted in light grey, while the smoothed temporal signature of the rank-2 CPD is clearly reduced during quiet sleep.

5.3.7 Classification Performance

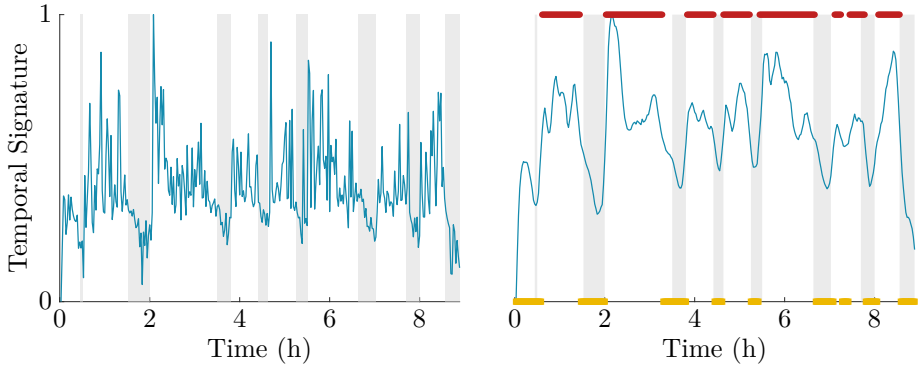
The performance of the algorithm is evaluated using the annotations by expert clinicians. The sensitivity, specificity, accuracy and Cohen's Kappa will be computed for each EEG recording and the mean and standard deviation will be reported. Moreover, to investigate the performance without the k-means clustering step, receiver operating characteristic (ROC) curves are constructed based on the smoothed temporal signatures and the clinical labels. The mean and standard deviation of the area under the ROC curve (AUC) will be presented.

5.3.8 Statistical Analysis

Statistical analysis is performed to gain insight in the choice of the parameters. A statistical test is used to investigate whether the area under the absolute value of



(a) Rank-1 CPD.



(b) Rank-2 CPD.

Figure 5.4: Illustration of the postprocessing and clustering of the temporal signature for a rank-1 (a) and rank-2 (b) decomposition of the multiscale entropy tensor (PMA = 40 weeks 5 days). On the left half of the figure, the temporal signature of interest is plotted in blue. The right half of the figure shows the smoothed temporal signature after applying the weighted moving average filter. Moreover, the quiet sleep and non-quiet sleep segments estimated by the algorithm are marked by the yellow squares and red dots, respectively. The clinically labelled quiet sleep periods are highlighted in light grey. The rank-1 CPD does not give a good indication of the sleep stages, while the (smoothed) temporal signature of the rank-2 CPD is clearly reduced during the quiet sleep periods.

the autocorrelation function is a good feature to detect the component of interest. In order to do this, the agreement between each of the R temporal signatures obtained from a rank-R CPD and the clinical sleep labels are evaluated using the Kappa score. The temporal signatures are then sorted in descending order based on their Kappa statistic. As a result, the first component is the most related to sleep staging and should have the largest area under the absolute value of its autocorrelation function (ACF). A statistical test is then performed to assess whether the area under the ACF of the component of interest is significantly different from the other components.
In addition, a statistical test is carried out to investigate the influence of the rank R on the performance. More specifically, we compared Cohen's Kappa between models with different values of the rank R. In all analyses, the Shapiro-Wilk test is used to test for normality. If the data is normally distributed, one-way ANOVA is used, otherwise a Kruskal-Wallis test is performed. The significance level is always set equal to 0.05.

5.4 Results

Comparison of the factor matrices obtained using the rank indicated by CORCONDIA or DIFFIT with the ones for other ranks revealed that the diagnostics were often not suitable to define the appropriate number of components in this application. Therefore, the rank has not been fixed for each recording, instead the results are reported for different values of the rank R. Table 5.1 presents the mean and standard deviation of the performance measures among the 97 EEG recordings. The first five rows show the performance for a fixed rank R for all recordings with R going from 1 to 5. From these data, we can see that the average performance is slightly higher for rank 2 compared to rank 1 (higher AUC and kappa). Moreover, Table 5.1 also shows that the performance decreases gradually for an increasing rank beyond 2. However, a Kruskal-Wallis test with multiple comparisons revealed that this performance difference between the rank-1 and rank-2 model is not significant, while Cohen's kappa of the rank-1 and rank-2 model are both significantly better compared to the rank-5 model. In order to get more insight in the performance difference between a rank-1 and rank-2 CPD, the performance of each recording is plotted as a function of the postmenstrual age in Figure 5.5. In this figure each green circle and pink square represents the area under the ROC curve of an EEG recording for a rank-1 and rank-2 decomposition, respectively. The dashed line indicates 37 weeks postmenstrual age. The most interesting aspect of this graph is that when a rank-1 decomposition is used, a high performance is only obtained up to around 36 to 37 weeks PMA. From that age onwards a clear drop in the AUC can be observed in Figure 5.5. The rank-2 CPD on the other hand,

Table 5.1: The classification performance of the proposed tensor-based method for different values of the rank R. The mean (standard deviation) of the sensitivity, specificity, accuracy, area under the ROC curve (AUC) and Cohen's Kappa are presented.

	Sensitivity	**Specificity**	**Accuracy**	**AUC**	**Kappa**
Fixed rank					
R = 1	0.73(0.29)	0.79(0.16)	0.78(0.15)	0.84(0.19)	0.47(0.34)
R = 2	0.82(0.27)	0.74(0.23)	0.76(0.18)	0.87(0.19)	0.49(0.33)
R = 3	0.74 (0.34)	0.71(0.2)	0.72(0.17)	0.80(0.24)	0.38(0.36)
R = 4	0.75(0.33)	0.67(0.22)	0.70(0.18)	0.80(0.24)	0.35(0.36)
R = 5	0.73(0.35)	0.64(0.24)	0.67(0.21)	0.75(0.27)	0.31(0.38)
Age-dependent rank					
Preterm: R = 1, term: R = 2	0.80(0.23)	0.79(0.17)	0.79(0.14)	0.87(0.16)	0.53(0.28)
Optimal rank					
Automatic component selection	0.76(0.32)	0.78(0.18)	0.79(0.17)	0.85(0.21)	0.50(0.38)
Optimal component	0.86(0.32)	0.81(0.18)	0.82(0.17)	0.91(0.21)	0.60(0.38)

has a high performance for most recordings at term equivalent age, but has a lower performance for some measurements recorded at a younger age.

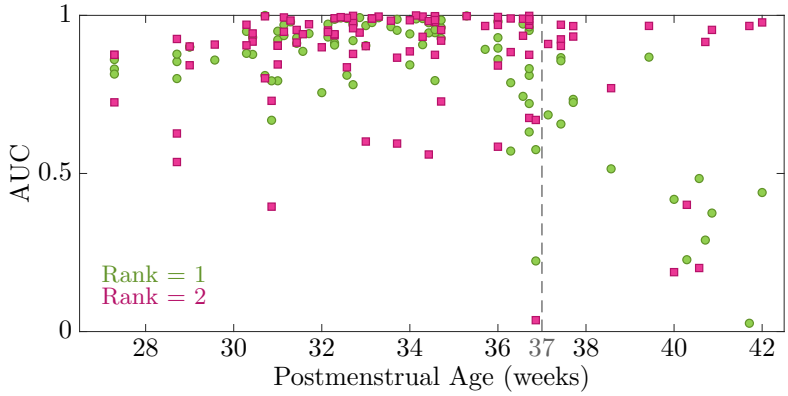


Figure 5.5: The area under the ROC curve as a function of the postmenstrual age at the moment of the recording. Each green circle shows the performance for one of the EEG recordings for a rank-1 CPD, whereas the pink squares represent the performance for a rank-2 CPD.

Based on these findings we decided to also report the results for an age-dependent rank, where the rank is chosen equal to 1 for preterm recordings and equal to 2 for EEGs recorded from 37 weeks PMA onwards. The performance of this approach is set out in the sixth row of Table 5.1. It is clear that the average performance of using an age-dependent rank is better compared to a fixed rank for all recordings, however this performance difference is not significant.

Finally, the last two rows of the table show the classification performance when the optimal rank (between 1 and 5) for each recording was selected based on Cohen's Kappa computed with the ground truth annotations. Specifically, the last row corresponds to the highest attainable performance (optimal rank and best component), while in the penultimate row the optimal rank was used but the component of interest was automatically selected using the strategy based on the autocorrelation explained above. As a consequence, the difference between these two rows is caused by failures of the automatic component selection procedure.

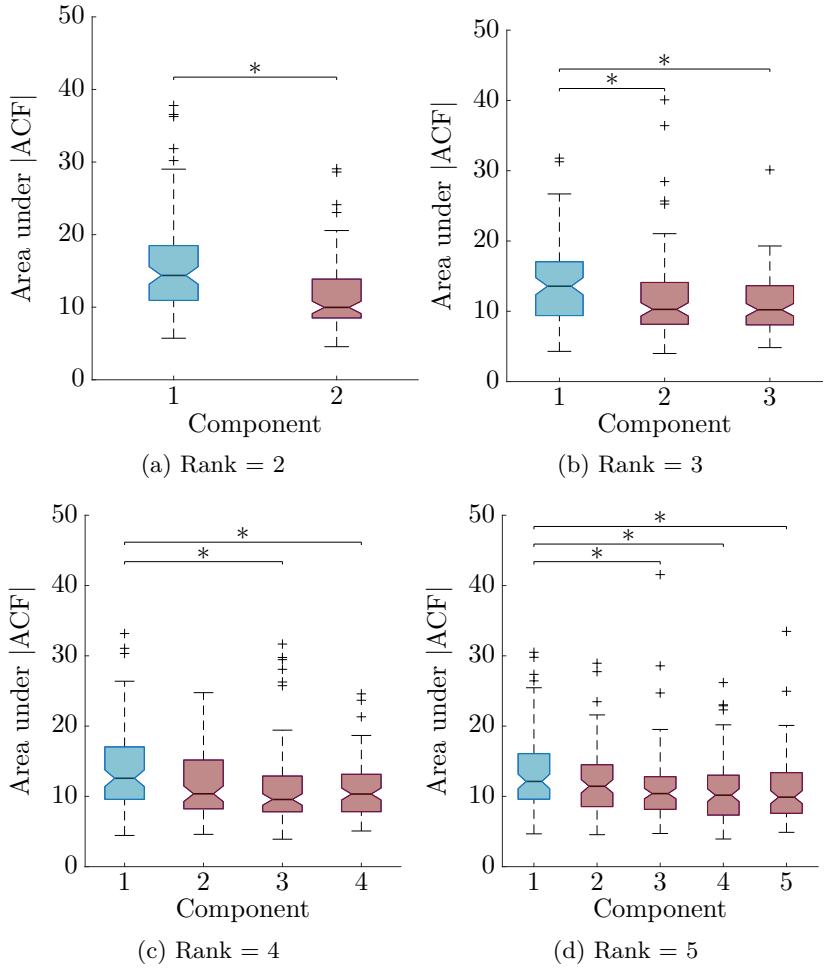
As explained in section 5.3.8, a statistical analysis is performed to examine whether the area under the ACF is a suitable feature to rely on for the automatic component selection. The boxplots in Figure 5.6 show the area under the absolute value of the ACF after sorting the temporal signatures based on their agreement with the clinical annotations. The component of interest, with the

largest Kappa score, is the first component and is marked in blue. A Mann-Whitney U test and Kruskal-Wallis test with multiple comparisons determined that the first component is significantly different from all other components for the rank-2 and rank-3 CPD, respectively (Figure 5.6 a,b). For the rank-4 and rank-5 CPD the area under the ACF of the sleep staging component is significantly different compared to all other components except the second one. This is also indicated on the boxplots in Figures 5.6 (c,d).

5.5 Discussion

This study aimed to discriminate quiet sleep (QS) from non-quiet sleep (NQS) using the nonlinear dynamics of the EEG signal in a data-driven way. The proposed algorithm relies on the fact that the EEG complexity is different depending on the sleep state. In order to quantify the complexity of the EEG signal, multiscale entropy is computed for consecutive segments of the multichannel EEG recording. These multiscale entropy values are then used to construct a third-order tensor with modes channels, scales and segments. Subsequently, the rank-R nonnegative CPD of the multiscale entropy tensor is computed. The temporal signature of interest is detected based on the area under the absolute value of its autocorrelation function. After smoothing, this temporal signature shows a cyclic pattern reflecting the neonatal sleep staging. Clustering is then performed to discriminate quiet sleep from non-quiet sleep.

The performance of the algorithm is reported for different ranks. The average Kappa is equal to 0.47 and 0.49, for a rank-1 and rank-2 CPD, respectively. However, if a rank-1 decomposition is used for EEGs recorded at a postmenstrual age below 37 weeks and a rank-2 decomposition is used for EEGs recorded at an older age, a Kappa of 0.53 is obtained. This indicates that a higher rank is preferred for neonates at term equivalent age, while rank 1 is suitable for most of the preterm recordings. The fact that for some recordings rank 1 is not sufficient, is demonstrated by the example in Figure 5.4. An even higher performance could be obtained if the optimal rank is used and the component of interest is detected correctly. However, as can be seen from Table 5.1, the performance when the optimal rank is used and the component of interest is selected automatically is slightly lower compared to an age-dependent rank. This reduction in performance is due to the fact that the procedure to select the component of interest is more likely to fail for a higher number of components. From the data presented in Table 5.1 we can see that the performance gradually decreases if a rank higher than two is used. This finding suggests that for the majority of the recordings in the current database a lower



(a) Rank = 2 (b) Rank = 3

(c) Rank = 4 (d) Rank = 5

Figure 5.6: Boxplots of the area under absolute value of autocorrelation of the temporal signatures after sorting them in descending order according to their Kappa score. The component with the largest Kappa score and therefore most related to sleep staging is the first one and is marked in blue, while the other (noise) components are marked in red. '*' indicates a statistically significant difference with $p < 0.05$. For the rank-2 CPD (a) and the rank-3 CPD (b) there is a significant difference between the area under the ACF of the sleep staging component and the noise component(s). For a rank-4 (c) and rank-5 CPD (d) the component most related to sleep staging is significantly different from all other components except one.

rank is more appropriate. Nevertheless, this performance reduction can also be partly attributed to the more difficult component selection for higher ranks.

There are various possible explanations for the need for a higher rank for recordings at term-equivalent age. First of all, it is likely that this is due to the increased amount of artifacts in near term EEG recordings. It is expected that an older neonate will move more, which could lead to more severe distortions and lower quality of the EEG signal. The higher performance for higher ranks in noisy preterm EEG recordings confirms this reasoning. Secondly, the emergence of four distinct sleep states around 36 weeks PMA might also play a role in this phenomenon [56].

The performance of the described algorithm is lower compared to state-of-the-art algorithms [9, 57, 162]. However, there are multiple advantages of the proposed method. The major advantage of the algorithm is that it is data-driven and unsupervised. As a consequence, the method can be easily applied on a completely new dataset and used in new centers, where there is little expertise about EEG sleep labelling. Moreover, in this study the algorithm is assessed on eight channel EEG recordings measured using the restricted 10–20 system for neonates. However, as long as the data is of appropriate quality, the same approach can also be applied on datasets with fewer or more electrodes without adjustments. Finally, the tensor decomposition can be updated in an efficient way whenever a new batch of EEG data is available [200]. This allows real time tracking of neonatal sleep states.

The main weakness of this study is the lack of a method to define the rank for each new EEG recording automatically in a data-driven way. This is a general issue in tensor applications, therefore it is suggested to combine different tools to asses the rank [2, 37]. We tested several diagnostics, but none of them resulted in a reliable estimate of the number of components. More research is needed to determine the optimal number of components for this specific application. An additional drawback of the proposed algorithm is the procedure for selecting the component of interest. The temporal signature of interest does not always have the largest area under the absolute value of the autocorrelation function. For that reason more advanced strategies for component selection should be investigated in future studies. Another limitation of the study is that artifacts commonly lead to misclassification of the segment. More specifically, EEG segments distorted by artifacts are often classified as quiet sleep because they have a reduced EEG complexity. The cause of this is two-fold. First of all, artifacts are more predictable and less complex compared to noise-free EEG [128]. Secondly, a high amplitude artifact will drastically increase the standard deviation of the segment, resulting in a much higher tolerance r and thus more template matches and consequently lower entropy. Detecting artifacts as quiet sleep periods will lead to a reduced performance, since the majority of

the (motion) artifacts occur during non-quiet sleep. The smoothing operation is implemented to reduce the influence of artifacts and can deal with short duration artifacts (sudden drop in EEG complexity). Yet, the postprocessing cannot avoid false detections caused by longer artifacts (on multiple channels). Further research might explore incorporating information about the noise in the algorithm to make it more robust to artifacts. Besides, an estimation of the amount of noise could be used to get an initial estimate of the proper number of components. Notwithstanding these limitations, the study suggests that a decomposition of a multiscale entropy tensor can be used to discriminate neonatal sleep stages.

This study is an exploratory analysis on the use of tensor decompositions for sleep stage identification in preterm infants. Only the differences in complexity between different sleep stages are used. However, in future work the temporal signature could be combined with other discriminating features (e.g., spectral edge frequency, power in specific EEG frequency bands, etc.) to boost the performance. Moreover, the heart rate variability, respiration and body movements of the infant are also dependent on the sleep state. Hence, in future research the combination of these complementary modalities can be studied. Finally, other tensorization techniques can be examined in future investigations.

5.6 Conclusion

This study confirmed that the EEG complexity is significantly lower during quiet sleep compared to non-quiet sleep. This property is exploited to develop an unsupervised algorithm that can detect quiet sleep in preterm infants based on the data-driven factorization of the multiscale entropy tensor.

Chapter 6

Quiet sleep detection in preterm infants using deep convolutional neural networks

So far, two algorithms for automated sleep staging in preterm infants have been proposed in this thesis. The first algorithm, explained in Chapter 4, followed the typical machine learning approach with the extraction of complexity features from the EEG which are then used to train an LS-SVM classifier. The second method, presented in Chapter 5, proposes a new pipeline based on tensor decompositions to identify the neonatal sleep stages in an unsupervised manner. While the tensor-based approach is not fully exploiting the data and available clinical annotations to achieve the optimal performance, the first algorithm is limited by the set of handcrafted features which only assess the complexity of the EEG. In this chapter, we will address these problems by introducing a new algorithm for automated EEG sleep staging based on deep learning.
The proposed 17-layer convolutional neural network is fed with raw multichannel EEG data and gives a sleep stage label, quiet sleep or non-quiet sleep, as output. The network is trained on 54 recordings from 13 prematurely born neonates and the performance is assessed on 43 recordings from 13 independent patients. This research was performed in close collaboration with Dr. Amir Hossein Ansari and led to a joint publication [9].

6.1 Introduction

Preterm infants are born during a critical period of rapid growth and development of the brain and central nervous system. In these newborns, the fast neural growth that normally occurs in the womb during the third trimester of gestation is interrupted and has to take place in the neonatal intensive care unit (NICU). Therefore, a primary concern of NICUs is to support optimal brain development of these vulnerable infants, who are at increased risk of neurodevelopmental disorders [144, 158]. Continuous electroencephalography (EEG) monitoring is considered a valuable noninvasive tool to assess and track brain maturation of newborns in the NICU. Sleep wake cycling (SWC) undergoes fast changes and is one of the main hallmarks of neurodevelopment in preterm infants. Since sleep state organization reflects the level of functional brain development, a first step towards automated analysis of cerebral function is to identify the different sleep stages [56, 168].
Moreover, evidence suggests that sleep is of utmost importance in the development and maturation of neural pathways. As a consequence, the neonate's sleep should be protected and promoted during their stay in the NICU [6, 196]. Automated algorithms for real-time neonatal sleep state identification could optimise the scheduling of neonatal interventions and care procedures in the NICU. Up to now, the gold standard for the assessment of sleep EEG is visual analysis by an expert clinician. However, visual interpretation of neonatal EEG is a tedious and time consuming task for the clinician and requires expertise that might not always be available in the NICU [56]. This can be overcome by developing automated algorithms to discriminate neonatal vigilance states.

A growing body of literature has investigated algorithms for neonatal EEG sleep scoring [28, 103, 159], which typically rely on the difference in continuity [56, 103] and frequency content [162] between neonatal sleep states. Palmu et al. [152] compared the proportional duration of bursts, called Spontaneous Activity Transients (SAT%), in the EEG between neonatal sleep states. They confirmed the existence of genuine vigilance states in early preterm infants and found that the fluctuation of the SAT% signal corresponds to sleep stage cycling. Stevenson et al. [181] continued this work and derived quantitative measures of brain activity cycling from the frequency domain representation of the SAT% time series. Besides analysis of the SAT% signal, EEG time profiles have been used for automated sleep stage detection as well [57, 105, 106]. The EEG time profiles are constructed by adaptive segmentation of the EEG and subsequent clustering of features extracted from each segment. Thresholding of the processed time profile, which shows the class membership of the EEG segments over time, will then lead to labels of the sleep states. This concept of EEG time profiles for sleep stage modelling was explored by Barlow et al. [13],

extended by Krajca et al. [105, 106] and further improved by Dereymaeker et al. [57]. In addition to the approaches described above, a number of studies have developed a neonatal sleep stage classifier based on a set of discriminative EEG features in combination with a machine learning classifier [50, 71, 103, 111].

While a wide range of features have been used to develop neonatal sleep stage classifiers, there is still no consensus on the optimal combination of features. Moreover, those handcrafted features are based on prior human knowledge about EEG sleep patterns, hence only a limited set of features have been examined. In general, the process of selecting the optimal feature set and classifier is a challenging task.

This study will address these problems by adopting a deep convolutional neural network (CNN) to detect quiet sleep in preterm infants. Deep neural networks learn the features directly from the data, which eliminates the need for manual feature extraction. CNNs have been extensively used in the field of image and speech processing [1, 33, 108, 115, 116, 176]. Only recently, deep neural networks have been attracting a lot of interest for biomedical applications, such as medical image analysis [173], detection of myocardial infarction [3] and (neonatal) seizure detection [4, 8]. CNNs have been applied for sleep stage scoring in adults [29, 134, 156, 178, 184, 194], however to the best of our knowledge this is the first study adopting CNNs for neonatal sleep stage classification.

The aim of this paper is to design and train a deep convolutional neural network to identify sleep stages in preterm infants. The performance of this novel classifier will be compared to two reference algorithms described in the literature: (1) a support vector machine (SVM) classifier using a set of spectral features [162] and (2) CLuster-based Adaptive Sleep Staging (CLASS) developed by Dereymaeker et al. [57].

The remaining part of the paper proceeds as follows. We start by describing the database that has been used to develop, train and test the classifier. We will then go on to a thorough description of the architecture and training procedure of the network developed for sleep scoring. The pipeline of the two reference algorithms described in the literature will be briefly explained as well. The following sections are concerned with the comparison of the proposed algorithm and the two reference algorithms from the literature in the sense of performance, error correlation and computational time. Next, the performance of the proposed CNN and two reference algorithms will be reported and compared. At last, the advantages and disadvantages of the proposed method and future directions will be discussed.

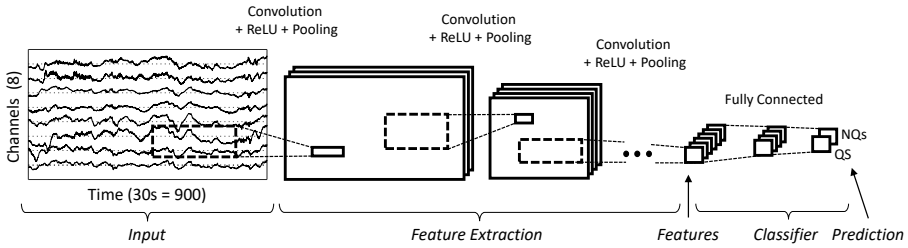


Figure 6.1: Architecture of the convolutional neural network.

6.2 Materials and methods

6.2.1 Database

The same dataset as described in section 4.2.1 is used to train and test the convolutional neural network. The database consisting of 8-channel EEG recordings from 26 prematurely born infants was split into a training set and a test set. The training set containing 54 recordings from 13 patients was used to develop and train the convolutional neural network. The total duration of training data is equal to 269 h, with 69 h of quiet sleep, and 200 h of non-quiet sleep. The test set consisting of 43 recordings from 13 independent patients was used to assess the classification performance. The test set consists of 223 h of EEG with 53 h during quiet sleep and 170 h during non-quiet sleep. This interpatient data splitting avoids that the model is tested on data from a patient that is also used to train the model. In this way, we expect that patient specific characteristics cannot bias the classification performance.

6.2.2 The proposed CNN for sleep stage classification

Preprocessing

Prior to using the EEG data as input for the CNN, some preprocessing steps are performed. First of all, the EEG is bandpass filtered between 1 and 15 Hz. The filtered EEG is then downsampled to 30 Hz in order to reduce the complexity of the network. The training data is normalized so that the mean and standard deviation of each EEG channel across the whole dataset is equal to zero and one, respectively. The parameters from the training data are then used to normalize the test data. At last, the EEG is segmented into windows of length 30 s, resulting in EEG segments of size 8 (channels) $\times$ 900 (samples).

Table 6.1: Layers of the designed network.

	Layer	Type	Size	No. filters	No. parameters	Stride	Padding	Output dimension
Feature Extraction	0	Input						(8,900,1)
	1	Conv	(1,10)	3	33	(1,1)	(0,9)	(8,900,3)
	2	ReLU						(8,900,3)
	3	Conv	(3,1)	3	30	(1,1)		(6,900,3)
	4	ReLU						(6,900,3)
	5	Mpool	(2,3)			(2,2)	(0,2)	(3,450,3)
	6	Conv	(1,5)	5	80	(1,1)	(0,4)	(3,450,5)
	7	ReLU						(3,450,5)
	8	Conv	(3,1)	5	80	(1,1)		(1,450,5)
	9	ReLU						(1,450,5)
	10	Mpool	(1,5)			(1,3)	(0,2)	(1,150,5)
	11	Conv	(1,5)	7	182	(1,1)	(0,7)	(1,150,7)
	12	ReLU						(1,150,7)
	13	Mpool	(1,6)			(1,4)		(1,37,7)
	14	Conv	(1,37)	10	2600	(1,1)		(1,1,10)
FC	15	Sigmoid						(1,1,10)
	16	Conv	(1,1)	2	22	(1,1)		(1,1,2)
	17	Softmax						(1,1,2)
Total Number of Parameters:					3027			

Conv: Convolutional layer
ReLU: Rectified linear unit
MPool: Pooling by maximum operator
FC: Fully connected classifier

CNN architecture

These preprocessed multichannel EEG segments are then fed into a 17-layer convolutional neural network, whose architecture is shown in Figure 6.1 and Table 6.1. The architecture of the network was designed based on our previous experience for neonatal seizure detection [8], and trial and error using the training and validation data. The CNN is designed and implemented in Matlab using the MatConvNet toolbox [204]. The first 14 layers perform feature extraction and the last 3 layers including the fully connected layers perform classification. The mask (kernel) size, stride, padding, and the number of feature maps used in the conv and/or pooling layers are also mentioned in the table. In order to train the CNN using the backpropagation algorithm, 20% of the

training data was selected as validation set and the remaining 80% of the training data was used to train the network's parameters. During this splitting of training data into training and validation set, the ratio of quiet sleep to non-quiet sleep segments is retained. The learning rate changes from 10^{-2} to 10^{-4} dependent on the training epoch and the layer. The weight decay was equal to 10^{-6}. For initializing the network, all bias terms were initiated with zero and the filter coefficients were set with Gaussian random noise with mean and standard deviation equal to 0 and 0.4 respectively. Batch learning was adopted with the batch size equal to 20 EEG segments and the maximum number of training epochs equal to 1000. Early stopping was used in order to avoid overfitting and ensure network generalization. So, the validation set error was monitored during the training phase, and training was stopped at epoch 493 which has the lowest validation set error.

Postprocessing

According to the literature a sleep stage should last at least three consecutive minutes or three out of four consecutive minutes [104]. Relying on this fact, a moving average filter with length 6 (6 segments of 30 s = 3 min) is used as a postprocessing step.

6.2.3 Spectral feature based neonatal sleep stage classifier

This neonatal sleep stage classifier is based on a set of spectral features described by Piryatinska et al. [162]. These features are fed into a support vector machine (SVM) classifier.

After bandpass filtering between 1 and 20 Hz and segmenting the multichannel EEG in epochs of 30 s, 9 spectral features are extracted from each EEG segment: (1-4) relative power in 4 EEG frequency bands (δ: 0.5 - 4 Hz, θ: 4 - 8 Hz, α: 8 - 12 Hz, β: 12 - 15 Hz), (5-6) spectral edge frequency (75% and 90%), (7) spectral moment, (8) spectral entropy and (9) amplitude entropy. Each multichannel EEG segment is then characterized by a total of 72 (9×8 channels) features and is marked as quiet sleep or non-quiet sleep according to the clinical label. The hyperparameters of the SVM with radial basis function (RBF) kernel are set using 5-fold cross validation on the training data. This SVM classifier is then trained using the training set and the performance is evaluated on the test set. As in the proposed CNN, a moving average filter with length 6 is added to remove the transient discontinuities.

6.2.4 Cluster-based Adaptive Sleep Staging (CLASS)

This algorithm is based on the relatively higher discontinuity in quiet sleep compared to non-quiet sleep and has been developed by Dereymaeker et al. [56]. Briefly, the pipeline of the algorithm is as follows. First, a preprocessing step consisting of a bandpass filter from 1 to 40 Hz and a notch filter at 50 Hz is performed. Artefact subspace reconstruction is then adopted to reject remaining artefacts, which could otherwise be wrongly detected as EEG discontinuities. After artefact removal, the cleaned EEG is downsampled with a factor three (sampling frequency = 83 Hz) to reduce the computational time during the subsequent steps. To deal with the nonstationarity of the EEG signal, the recordings are adaptively segmented resulting in quasi-stationary segments of variable length. Features extracted from these segments are then used to cluster segments with similar characteristics. A cluster time profile is obtained by representing each sample by its cluster label. Once the cluster time profile is extracted for each EEG channel, some processing steps are performed to obtain a single smooth envelope which can then be thresholded to detect quiet sleep segments. This threshold is calculated as the mean of the envelope plus a factor multiplied with the standard deviation of the signal envelope. In the final stage of the algorithm, quiet sleep detections shorter than three minutes are removed [104].

6.2.5 Classification performance

In order to assess the performance of the sleep stage classifier, the sensitivity, specificity and Cohen's Kappa coefficient κ have been computed. All of the above mentioned evaluation metrics are measured for each recording of the test set, and the median and interquartile range are reported. The computation of these measures require a fixed decision threshold. In this study, the optimal threshold was defined so that the sensitivity equals the specificity on the training set. Since the threshold used in the CLASS algorithm is computed as threshold = mean($\mathbf{y}$) + factor $\times$ std($\mathbf{y}$), where $\mathbf{y}$ denotes the smooth envelope, the factor instead of the complete threshold is optimized on the training set.

In addition, the receiver operating characteristic (ROC) curve is computed for each recording of the test set. This allows to analyse the effectiveness of the classification without defining a fixed threshold. The mean and median ROC curve over all recordings are then obtained by taking the mean and median of the true positive rate and the false positive rate at a specific threshold. These mean and median ROC curves are plotted and the area under the curve (AUC) is computed.

6.2.6 Error correlation

In order to investigate the error correlation between the three sleep stage classifiers, the percentage of test segments correctly classified by all three algorithms, by only two or one of the algorithms, or by none of the algorithms are computed. In addition, Cohen's Kappa is computed among each of the sleep classification algorithms and the clinical sleep labels in order to investigate the agreement between the different algorithms.

6.2.7 Computational time

It is generally known that deep neural networks are highly complex and computationally expensive to train. However, when using a sleep stage classifier in clinical practice, it is not the training time, but the time required to classify a new incoming EEG segment that is of interest.

In order to compare the computational evaluation time of the three algorithms, sleep stage classification of the complete test set in blocks of 2 h of EEG is performed with each of the algorithms. The three algorithms were run chronologically and the mean and standard deviation of the computational time are reported to avoid variable CPU loading. Since the complete test set is classified, more than 100 iterations are performed. This experiment was conducted on a workstation with an Intel(R) Core(TM) i7 3.6 GHz processor and 16 GB RAM, implemented in MATLAB R2016a software (The MathWorks, Natick, MA, USA).

6.3 Results

6.3.1 Feature evolution during sleep-wake cycling

To illustrate how each of the 10 features extracted by the convolutional neural network behave during each of the sleep states, boxplots of the features during sleep-wake cycling are shown in Figure 6.2. The top of Figure 6.2 illustrates the amplitude integrated electroencephalography (aEEG) derived from the bicentral channels C3/C4 from a test recording at 31 weeks and 3 days postmenstrual age. aEEG is commonly used in the NICU to monitor the functional brain integrity and is a suitable tool to assess sleep-wake cycling of the infant [49]. More discontinuous activity during quiet sleep is recognized by a widening of the aEEG trace, while the more continuous activity during active sleep or wakefulness is characterized by a narrow trace [12, 49]. The aEEG of the quiet

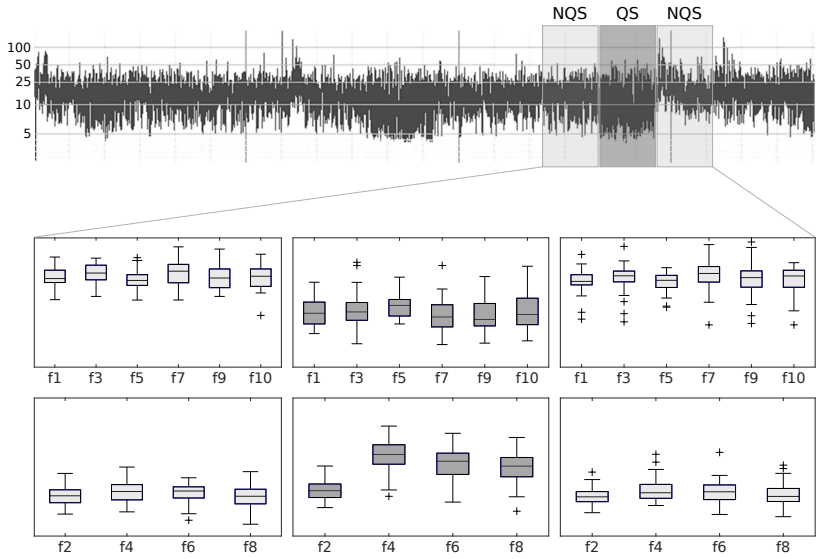


Figure 6.2: The aEEG trace derived from the bicentral channels C3 and C4 from a recording of the test set at 31 weeks and 3 days PMA is shown on top. The selected quiet sleep (QS) segment and preceding and subsequent non-quiet sleep (NQS) segments are indicated on the aEEG. The boxplots show the features during each of these selected segments. The top half of the boxplots show the six features that reduce during quiet sleep (f1, f3, f5, f7, f9, f10), the bottom half shows the four features that increase during quiet sleep (f2, f4, f6, f8). Most of the features are significantly different for the two sleep stages.

sleep segment selected in Figure 6.2 has a clearly wider bandwidth compared to the non-quiet sleep segments before and after. The top row of boxplots shows features 1, 3, 5, 7, 9 and 10, which are reduced during quiet sleep (the middle box) compared to non-quiet sleep (left and right box). The bottom row provides the boxplots of features 2, 4, 6 and 8, which increase during quiet sleep. From the boxplots in Figure 6.2, it can be seen that most of the features are well discriminating quiet sleep from non-quiet sleep.

6.3.2 Classification performance

The classification performance of the proposed CNN is illustrated in Figure 6.3, which presents the ROC curves for each of the test recordings separately, and the mean and median ROC curve. The area under the mean and median ROC

curve are equal to 92% and 98%, respectively. The histogram of the training data segments and the AUC for each recording of the test set as function of postmenstrual age are plotted in Figure 6.4. It can be seen from the histogram that training data is available at all ages. From the AUC in function of PMA, it is apparent that most of the test recordings are measured between 30 and 38 weeks of PMA. Moreover, the figure shows that the algorithm performs well over a wide range of PMA. It is notable that in Figure 6.3 as well as in Figure 6.4 two outliers with a low AUC ($\leqslant 0.65$) can be observed.
Table 6.2 compares the AUC, sensitivity, specificity and Cohen's Kappa across the 3 methods. It can be seen from the table that the proposed CNN has similar performance compared to the CLASS and feature-based algorithm. In addition, it is clear that the postprocessing step significantly improves the performance of the feature-based sleep stage classifier.

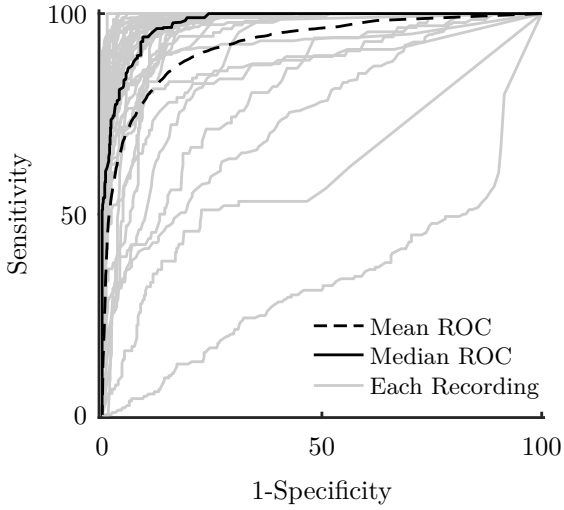


Figure 6.3: The ROC curves for the proposed CNN sleep stage classifier. The light grey ROC curves show the performance of the classifier for each recording of the test set. The black dashed and full line represent the mean ROC and median ROC curve, respectively.

6.3.3 Error correlation

Figure 6.5a illustrates the percentage of segments correctly labelled by all the algorithms, by only two or one of the algorithms, or misclassified by all methods. The complete bars represent the complete test set, the dark grey part

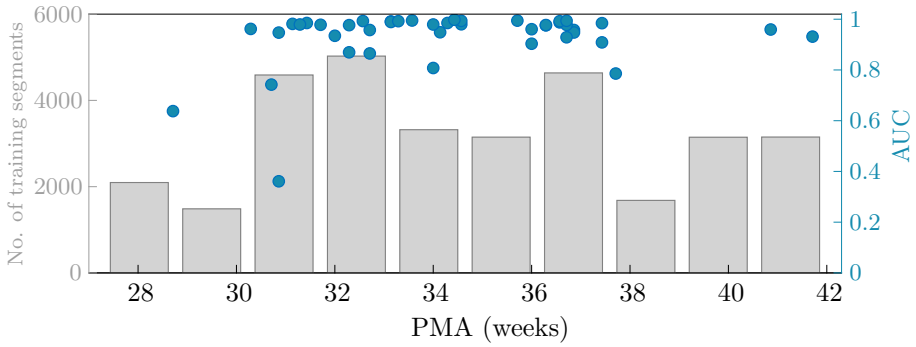
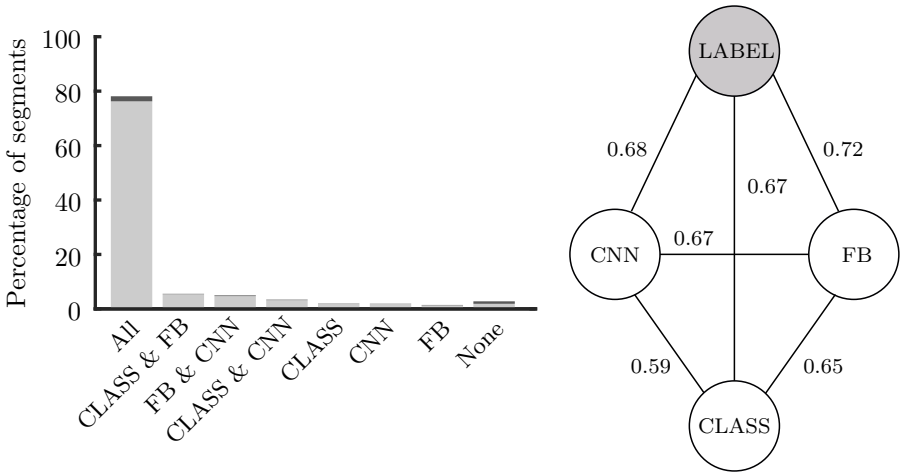


Figure 6.4: The histogram of the training data segments is displayed in light grey (left y-axis). The blue circles show the AUC for each recording from the test set (right y-axis).

Table 6.2: The classification performance for the proposed CNN, the CLASS algorithm and the feature-based approach without (NP) and with (PP) postprocessing respectively. The area under the mean ROC curve (AUC) and median(IQR) of the sensitivity, specificity and Cohen's Kappa are set out.

	AUC	Sensitivity	Specificity	Kappa
CNN	0.92	0.88(0.26)	0.93(0.13)	0.74(0.17)
CLASS	0.92	0.72(0.19)	0.97(0.06)	0.75(0.19)
Feature-based NP	0.83	0.74(0.21)	0.87(0.13)	0.59(0.25)
Feature-based PP	0.93	0.83(0.28)	0.97(0.07)	0.77(0.23)

corresponds to the segments of the two outliers in Figure 6.4, while the light grey part of the bars correspond to the remaining segments. From this bar graph, it is clear that most of the segments (nearly 80%) are correctly classified by all algorithms. Approximately 14% of the segments are correctly identified by two of the algorithms, and only 8% of the segments were only correctly classified by one of the described methods. Less than 3% of the segments were wrongly classified by all the algorithms. However, a considerable amount of the segments that were misclassified by all algorithms are part of the two recordings with a low AUC. More specifically, from all segments of the outliers, almost 25% is not identified by any of the algorithms. This illustrates the high agreement among the three algorithms, which is also confirmed by the Kappa values shown in Figure 6.5b.



(a) (b)

Figure 6.5: The error correlation between the three sleep stage classifiers. (a) The bar graph shows the percentage of EEG segments correctly classified by all, two, only one or none of the algorithms. The dark grey part of the bars correspond to the segments of the two outliers, while the light grey part corresponds to the remaining test segments. (b) This graph represents the agreement among the three algorithms (CNN, FB, CLASS) and the clinical labels (LABEL) computed using Cohen's Kappa coefficient.

6.3.4 Computational Time

The bar graph in Figure 6.6 illustrates the mean evaluation time for 2 h EEG segments. The dark blue box represents the average total time for loading, preprocessing and classifying 2 h EEG data. The light blue box shows the time required for classification. In case of CLASS, the classification consists of performing CLASS, starting from the artefact subspace reconstruction up to defining the quiet sleep periods. For the feature-based approach, the computational time for feature extraction, classification using the SVM and the postprocessing are merged in the classification time. Finally, the time for classification and postprocessing are combined and marked as classification time for the CNN.

From this chart, it can be seen that overall the CNN (10 $\pm$ 1.0 s) is 9 times faster than CLASS (92 $\pm$ 3.8 s) and slightly faster than the feature-based (12 $\pm$ 1.3 s) approach. The most remarkable difference is the classification time,

which is significantly lower for the CNN (0.16 ± 0.04 s) compared to CLASS (85 ± 3.74 s) and the feature-based classification (2.5 ± 0.28 s).

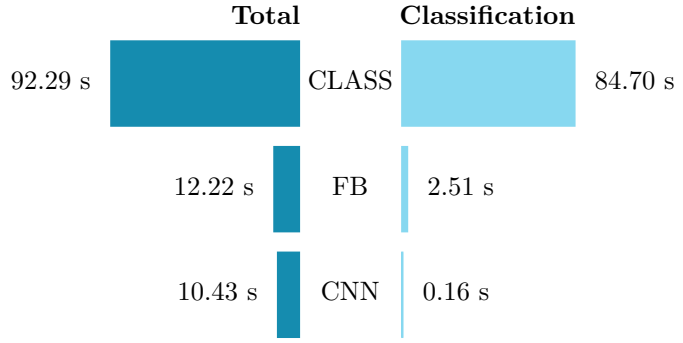


Figure 6.6: The average computational time for 2 h multichannel EEG segments for each of the three algorithms. The total computational time (left) and the time for only classification (right) are shown.

6.4 Discussion

The aim of the current study was to implement a convolutional neural network for neonatal sleep stage classification. By adopting convolutional neural networks which are able to learn relevant features automatically from the raw EEG data, the difficult and time-consuming process of selecting the proper feature set, which requires domain knowledge from experts, can be avoided. Moreover, the CNN optimizes the feature extraction and classification simultaneously. This novel data-driven approach successfully identified quiet sleep in preterm infants with an average and median AUC of 92% and 98% respectively.

While this performance is comparable to the CLASS algorithm developed by Dereymaeker et al. [56], the CNN has some advantages compared to the cluster time profiles approach. First of all, except for bandpass filtering the proposed method does not use any advanced artefact removal, whereas the CLASS algorithm uses artefact subspace reconstruction as additional artefact rejection. Secondly, the proposed CNN can be retrained if additional data collection is performed, while this is not straightforward for the CLASS algorithm. As sleep patterns will be better represented in a larger training database, the classification performance will further improve. At last, it must be pointed out that the threshold used in the CLASS algorithm is based on the mean and the standard deviation of the signal envelope. Therefore, the complete

or at least a representative part of the recording should be available before the classification can be performed. In contrast to CLASS, the proposed CNN uses a fixed threshold which can be decided based on the training set and does not need to be adjusted to the new recording. This facilitates real time sleep scoring. Although the proposed CNN uses only 10 features as input for the fully connected classification stage, it reaches the same performance as the feature-based approach which is based on 72 features [162]. To conclude, it can be seen in Figure 6.6 that the presented CNN is considerably faster compared to the current state-of-the-art.

One of the key challenges of neonatal sleep stage classification is to deal with the fast alterations in the EEG patterns during sleep ontogenesis. It is apparent from Figure 6.4 that the network is able to train the age specific EEG characteristics, resulting in a good performance over a wide PMA range (30 - 38 weeks PMA). Due to the small number of recordings below 30 weeks and beyond 38 weeks PMA, we cannot make any statements about the performance in those age ranges. Both in Figure 6.3 and 6.4, two outliers can be observed. These recordings also had a low AUC in the CLASS and feature-based approach. According to an expert clinician, the outlier at 28 weeks and 5 days (AUC equal to 65%) is due to poor EEG quality probably as a result of electrode impedance problems, while the low performance of the recording at 30 weeks and 6 days (AUC equal to 36%) is caused by an intravenous infusion motor artifact.

In this study, the overall AUC is computed as the area under the mean or median ROC curve. As explained above, this mean ROC curve is obtained by averaging the sensitivity and specificity at an incremental threshold between zero and one, which is called threshold averaging [66]. It is important to note that this procedure of first computing the average ROC and then the AUC is not the same as computing the AUC for each recording separately and then taking the average. Computing the mean of individual AUCs is called vertical averaging, which corresponds to taking the average of the true positive rates at a fixed false positive rate [66]. In this application we opted for threshold averaging rather than vertical averaging, because implementation of the algorithm in a brain monitor for use in clinical practice requires the choice of one fixed threshold.

The present study has relevance for clinical practice. First of all, the proposed CNN has a low computational time, hence it is suitable for real time sleep stage monitoring. Real time sleep scoring is important in the NICU as it can aid in making the neonatal care more patient-driven and avoid disruption of the infant's sleep. In addition to implementation of the algorithm in a brain function monitor to promote neonatal sleep in the NICU, this algorithm is of great interest in maturation studies as we have to account for the cyclic nature of EEG during brain maturation analysis. Furthermore, the sleep EEG undergoes fast maturational changes during early brain development.

The main limitation lies in the fact that the training phase is computationally expensive and requires a large database. However, training time is secondary since the network only has to be trained once. After training, the recall computational time is the important parameter, which is much smaller for CNN compared to the existing algorithms. A weakness of this study was the paucity of EEG recordings below 30 weeks and beyond 38 weeks postmenstrual age. Another drawback of CNNs is the lack of interpretability of the features. Further work needs to be done to establish which features are extracted by the network.

The results of the current study suggest that convolutional neural networks are a promising approach for sleep stage classification in preterm infants. However, several steps can be explored to further improve the performance of the sleep stage classification. A number of recommendations for future research are given. First of all, as mentioned before, the current architecture of the network has been defined based on trial and error. As a result, there is abundant room for further progress in designing the network, especially with the aid of the high-speed graphics processing unit (GPU). Some of the model design parameters (e.g. number of layers) could for instance be considered as hyperparameters and could be systematically tuned in order to obtain a more optimal and reproducible network architecture. In addition, the proposed method uses a postprocessing step to take into account that the sleep state cannot change instantly. In future investigations, it might be possible to incorporate this in the network by including information of the preceding and subsequent segments by using another deep learning algorithm such as Long Short Term Memory networks (LSTM). Moreover, clinicians do not rely only on EEG characteristics, but assess other non-cerebral criteria as well during visual sleep scoring [56, 157]. Behavioural correlates that are essential to distinguish quiet sleep and non-quiet sleep are body movements, eye movements and cardiorespiratory regularity. Therefore, we expect that the performance can be improved by including additional modalities such as EMG, respiration and EOG. Our results indicate that the proposed network can reliably classify sleep in preterm infants with a PMA from 30 to 38 weeks. Future studies should aim to replicate the results for a larger age range, by retraining the network using data from an older, term neonates population. At last, interpretation of the features extracted by the network is an important issue for future research.

6.5 Conclusion

In this study, we designed and implemented a deep convolutional neural network which automatically extracts optimal features to discriminate quiet sleep from

non-quiet sleep in preterm infants. The proposed network achieved the state-of-the-art performance over a wide range of PMA without using domain-specific knowledge. Furthermore, the presented sleep stage classifier has a low computational time, which makes the CNN based sleep stage classification attractive for real time sleep scoring in the NICU.

Chapter 7

Comparison of neonatal sleep stage classification algorithms

In this chapter, the sleep stage classification models developed and explained in the previous chapters (Chapter 4-6) will be compared. For the sake of completeness, two algorithms proposed in the literature, the CLASS algorithm and an approach based on spectral features in combination with an SVM, have also been included in this comparative analysis. Both algorithms are briefly discussed in Chapter 6, for a more elaborate explanation the reader is referred to [56, 162].

This chapter begins by explaining which division into training and test set was considered and how each of the algorithms was adjusted to evaluate the performance on this particular test set. Section 7.1 will then go on to present and discuss the performance of each classifier on the same test set. In section 7.2 the effect of brain maturation and the concomitant changing EEG patterns on the performance of the sleep stage classifier will be explored. At last, the generalizability and computational time of the proposed methods will be compared in section 7.3 and 7.4

7.1 Performance comparison

7.1.1 Database and performance evaluation

To allow a proper comparison of the different algorithms, they all have to be trained and evaluated on the same training and test set, respectively. Since retraining the deep learning algorithm is not that straightforward and a patient independent splitting of the dataset was used to develop the network, this division into training and test set has also been used for this comparative analysis. As a consequence, the performance of the CNN and the two reference algorithms is the same as presented in Chapter 6, but now the mean and standard deviation instead of the median and interquartile range will be presented.
Given that the tensor-based sleep stage identification is unsupervised, no extra computations were required. However, the rank for the decomposition of each of the multiscale entropy tensors has to be defined. We opted to present the results with an age-dependent rank, meaning that a rank equal to 1 is used in EEG signals recorded before 37 weeks PMA and a rank equal to 2 for older neonates.
At last, the sleep stage classifier based on complexity features derived from the multiscale entropy curve was trained on a different training set, so a new LS-SVM classifier has to be trained using the current training set for proper comparison. As in the proposed algorithm (Chapter 5) 1500 data points representing the underlying distribution of the dataset are selected from the original training dataset to train a fixed-size LS-SVM. The set of support vectors is found by maximizing the Renyi-entropy and k-medoids is used for the initialization in order to reduce the number of iterations [203]. After training the fixed-size LS-SVM classifier, the algorithm is evaluated on the same test set. Similarly to the other algorithms, the decision threshold is fixed and defined based on the training dataset in order to report realistic performance measures. The decision threshold is selected so that the average sensitivity equals the average specificity on the training set. Based on this decision threshold, the sensitivity, specificity, accuracy and Cohen's Kappa are then computed for each EEG recording and the average and standard deviation among the different test recordings will be presented. Moreover, the area under the average ROC curve is extracted.

Table 7.1: Performance comparison of various algorithms for automated sleep stage classification in preterm infants. The top three algorithms have been proposed in the previous chapters, while the bottom three are reference algorithms from the literature. The mean (standard deviation) of the performance metric among the different test recordings is presented for each of the considered algorithms.

	Sensitivity	**Specificity**	**Accuracy**	**AUC**	**Kappa**
Proposed algorithms					
CNN	0.80(0.22)	0.90(0.12)	0.88(0.10)	0.92	0.68(0.24)
CPD-MSE	0.75(0.25)	0.81(0.17)	0.79(0.14)	0.87(0.15)	0.50(0.29)
Complexity features	0.66(0.24)	0.87(0.13)	0.83(0.10)	0.85	0.52(0.23)
Reference algorithms					
CLASS	0.69(0.20)	0.95(0.06)	0.88(0.09)	0.92	0.66(0.24)
Feature-based NP	0.73(0.16)	0.84(0.11)	0.82(0.09)	0.83	0.53(0.17)
Feature-based PP	0.77(0.20)	0.92(0.11)	0.89(0.09)	0.93	0.70(0.21)

7.1.2 Results

Table 7.1 provides an overview of the performance of each of the proposed classification models on the same test set consisting of 43 EEG recordings from 13 prematurely born neonates. Note that for the algorithm based on the decomposition of the multiscale entropy tensor both average and standard deviation of the AUC are reported. This has to do with the fact that the final classification decision is based on clustering of the temporal signature of interest rather than applying a fixed decision threshold. As a result, the ROC curve and the corresponding AUC can be computed for each recording separately without loss of generalizability. A detailed overview list the performance for each of the test recordings for the different algorithms is presented in Tables A.1 – A.6 in Appendix A.

7.1.3 Statistical analysis

Statistical testing is used to compare the performance of the different methods for neonatal sleep stage classification. We selected Cohen's Kappa as the performance measure of interest, since it uses a fixed decision threshold, as would be the case in clinical practice. Moreover, it is a comprehensive performance metric which is also reliable in unbalanced datasets. The normality of the distribution of the Kappa scores is tested using a Shapiro-Wilk test. Since the null-hypothesis of normal distribution was rejected, statistical significance

was analysed using the Kruskal-Wallis test with multiple comparisons. The significance level was set at 0.05.
This statistical testing revealed that Cohen's Kappa of the CNN, the spectral feature-based approach with postprocessing and the CLASS algorithm are significantly higher compared to the tensor-based method or the algorithms relying on either complexity or spectral features without postprocessing step. No significant difference within each of these three algorithms could be observed.

7.1.4 Discussion

From this analysis, it is apparent that the convolutional neural network outperforms the algorithms proposed in the previous chapters. This comes as no surprise, as this approach fully exploits the training data to learn the EEG patterns corresponding to the different sleep stages, while the two other algorithms only consider information about the signal's complexity. However, the performance of the deep learning algorithm is comparable to the CLASS algorithm and the feature-based algorithm with an additional postprocessing step. From the data in Table 7.1 and the statistical test, it can be seen that this postprocessing step is of key importance and leads to a significant increase of the classification performance. The sleep stage classifier based on solely EEG complexity features attains approximately the same performance as the feature-based algorithm without postprocessing step. Hence, both complexity and spectral content of the EEG contain a considerable amount of information about the sleep-wake cycling of neonates. Note that a similar postprocessing step, consisting of a smoothing operation using on moving average filter, could also be applied on the output of the LS-SVM classifier using complexity features. This might boost the performance of the approach based on complexity features. However, the complexity features are computed for 100 s segments instead of the conventional 30 s segments. As a result, less segments can be considered during smoothing, hence a smaller rise in classification performance is expected. These longer epochs, which are required to obtain a reliable estimate of the multiscale entropy, also hamper the accurate localization of sleep stage transitions and detection of short duration sleep cycles.

7.2 Maturational effect

The challenge of developing a sleep stage classifier for preterm infants is to ensure that the algorithm performs well on a wide range of PMA. In order to illustrate that all of the proposed methods are able to deal with the maturational

trend of the neonatal sleep-wake cycling, the AUC of each test recording is plotted as function of the PMA at the moment of recording in Figure 7.1.
It is clear that the majority of the recordings of the test set are measured between 30 and 38 weeks PMA. Only few younger and older neonates are included in the test set. Overall, it can be seen that the classifiers work well throughout the complete age range. The data in Figure 7.1 show that both CNN and the complexity feature based algorithm perform poorly on two recordings, one measured at 28 weeks and 5 days and one at 30 weeks and 6 days. As mentioned in Chapter 6, this is probably due to inadequate recording quality and infusion motor artifacts, respectively. Remarkably, the CPD-based algorithm is able to deal with these distortions of the EEG and attains a higher performance for these recordings. However, the tensor-based approach has difficulties to classify other recordings correctly, especially the ones near 37 weeks PMA. This is probably caused by a poor choice of the rank, which is the main drawback of the CPD-based algorithm. For example, the recording at 36 weeks and 6 days has AUC of 0.22 for a rank-1 decomposition, while an AUC of 0.84 would be obtained when a rank-3 decomposition is performed. Hence, automated decision of the optimal rank is an important issue for improvement of this algorithm.

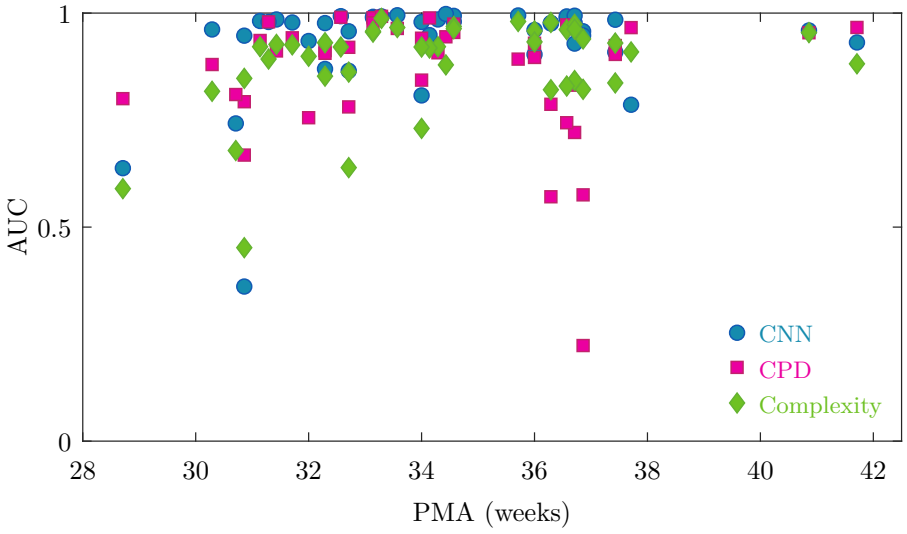


Figure 7.1: The AUC for each test recording using the three proposed algorithms for neonatal sleep stage classification.

7.3 Generalizability

The ultimate goal of algorithm development in the setting of biomedical signal processing is to develop an algorithm that can be used in clinical practice at the benefit of the patient. Hence, we want to build a sleep stage classification model that does not only perform well on the dataset used to develop and evaluate the model, but that can also be applied in other clinical centers where little expertise is available (hence no clinical sleep annotations are at disposal).
In the sense of generalizability, the aspects that have to be taken into account for neonatal EEG sleep staging is that many centers measure aEEG rather than full montage conventional EEG. As a result, only two up to four EEG channels will be available for analysis. Moreover, different EEG monitors, each with their own characteristics (e.g. amplifier, sample rate, filter settings, etc.), are being used. This will affect the appearance of the EEG time series.

All supervised algorithms suffer from the fact that they have been explicitly trained on the characteristics of this specific training dataset. As a result, it is likely that the performance will drop when the same model is tested on a new database acquired in a different center. Moreover, when fewer electrodes are used to record the EEG or an electrode detaches during the measurement, the classifier's performance will be affected.
In order to illustrate this, we have performed a channel analysis on the sleep staging algorithm based on deep learning. At first, we simulated the detachment of a single electrode by replacing the time series of the corresponding channel by zeros. The performance for the case of one missing electrode obtained on the complete test set is visualized in Figure 7.2. It can be seen that for most electrodes the performance is similar to the performance reported in Table 7.1 using all EEG channels. Only removal of the T3 electrode induces a larger drop of the classifier's performance.

Secondly, we have also explored different electrode setups. As the aEEG is commonly measured using two or four electrodes in the C3, P3, C4 and P4 positions, we have evaluated the algorithm for the electrode pair C3 – C4 and for all 4 electrodes. However, the parietal electrodes are not available in the current dataset, so they are replaced by occipital channels O1 – O2. The results for these two montages are shown in Figure 7.3. The reduction of the number of electrodes to four electrodes, covering the central and occipital regions, resulted in a clear drop in performance. However, the AUC is still above chance level. The reduction from four to two EEG channels only leads to a small performance decline. A similar decrease in performance is expected in the supervised algorithms based on either complexity or spectral features. In these methods a reduced number of electrodes will lead to missing features, which on their turn will not be considered by the classifier. In future studies,

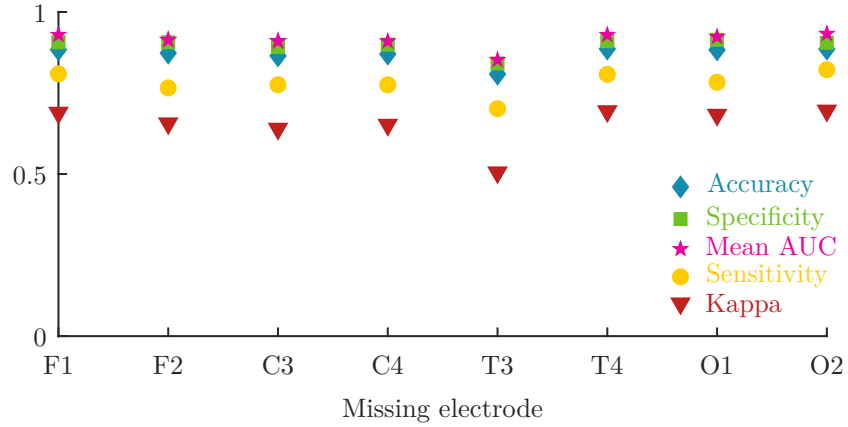


Figure 7.2: Performance of the CNN sleep stage classifier on the complete test set for one missing electrode.

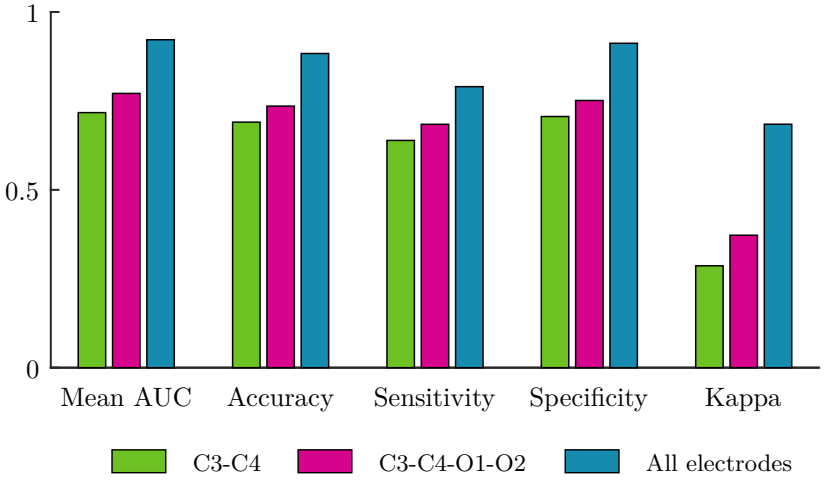


Figure 7.3: Performance of the CNN sleep stage classifier on the complete test set for a reduced montage. The green bars represent the performance for 2-channel EEG measured using the central electrodes C3-C4. The pink bars illustrate the performance for EEG measured using C3-C4-O1-O2. The blue bars show the performance using the full-montage including all 8 electrodes.

this problem could be tackled by training a separate classifier for each EEG channel and merge the outputs of these classifiers. This late integration ensures that the algorithm can also be used with fewer electrodes, but requires training of multiple models and is hence more complex compared to early integration approaches.

In contrast to the algorithms using a supervised classifier (i.e. the CNN and the algorithms based on complexity or spectral features), the tensor-based approach and the CLASS algorithm are expected to be more easily transferable to different centers. First of all, the decision threshold is either determined based on k-means clustering or set using the statistics of the envelope of the cluster time profile (with an additional scaling factor). Therefore, we expect that even when the EEG time series is measured with a different machine, the algorithm will work reasonably well. Some recalibration of the scale factor to define the threshold for the CLASS algorithm might be required though.
In addition, these algorithms are expected to be less sensitive to the number of electrodes. The tensor-based algorithm does not necessarily need an 8-channel recording. The same pipeline can also be used using less electrodes and proper sleep stage identification is expected as long as some of the retained channels are of decent quality. Likewise, the cluster time profile used to define the sleep stage in the CLASS algorithm can also be computed based on fewer EEG channels. To conclude, the major advantage of the unsupervised learning methods is their generalizability properties.

7.4 Computational time

The computational time of the CNN, the CLASS algorithm and the spectral feature-based algorithm have been compared in Chapter 6. This revealed that the CNN approach is computationally faster that the CLASS and feature-based algorithm, especially for the classification time. While the classification step of the tensor- and complexity feature-based algorithms are reasonably fast, they cannot compete with the algorithms mentioned above due to the computationally expensive feature extraction. In order to compute the multiscale entropy, the sample entropy has to be computed at multiple scales, which is the bottleneck in the computational time of these algorithms.

Table 7.2: Overview of the pros and cons of the proposed preterm sleep stage classification algorithms.

Algorithm	Pros	Cons
CNN	• Optimal use of training dataset • High performance • Low computational time	• Poor interpretability • Tailored to 8-channel EEG • Requires large training dataset
CPD-MSE	• Unsupervised • Generalizability • Interpretability	• Decision of the rank • Suboptimal performance
Complexity features	• Interpretability • Can be easily extended	• Tailored to 8-channel EEG

7.5 Conclusion

To conclude, three methods for automated EEG sleep staging in preterm infants have been proposed. Each has its advantages and drawbacks, an overview is given in Table 7.2. Dependent on the properties of the dataset and the performance requirements, another algorithm might be optimal. Therefore, we are convinced that the best approach to perform automated sleep staging on a completely new, unlabelled dataset is to run multiple algorithms and compare the obtained sleep labels. A high agreement among the different algorithms indicates that it is likely that the sleep stages are correctly identified, while caution is advised when poor agreement is observed.

Part III

Automated brain maturation quantification

Chapter 8

Assessing brain maturation in preterm infants using EEG complexity features

In premature babies, especially those born very early, many organs are underdeveloped at birth. These immature organs are particularly susceptible to abnormal structural and functional development, which can lead to long-term health issues. Especially the early development of the neonatal brain has a large impact on the neurodevelopmental outcome of these vulnerable neonates. Therefore, careful monitoring of the neurological function is a primary concern in the current NICU setting.

Automated analysis of the electroencephalographic (EEG) data has been proven useful in tracking the evolution of the brain maturation during the first postnatal weeks. In this chapter, we will investigate whether the complexity of the EEG signal can provide information about the brain function of preterm neonates. First, we examine how the complexity of the neonatal EEG signal evolves with increasing postmenstrual age. In addition to this correlation analysis, a brain-age regression model is developed to estimate the postmenstrual age (PMA), and thus the neonate's electrocortical maturation, based on a combination of complexity features. The study is also described as part of [50].

8.1 Introduction

According to the World Health Organization (WHO), approximately 15 million babies are born prematurely every year, and the number of births before 37 completed weeks of gestation is still rising [20]. Due to the advancements in the neonatal intensive care units (NICUs), the survival rates of these infants have increased over the course of several decades. However, these babies are more susceptible to brain damage and are prone to neurological impairment. As a result, focus has shifted towards close monitoring of the brain development during the first critical weeks of life [166]. Assessment of brain maturation by automated analysis of the neonatal EEG can assist clinicians in starting timely treatment, monitoring the evolution of the baby, and predicting the neurodevelopmental outcome. As a result, a great deal of literature has focused on the extraction of features from the EEG that reflect electrocortical maturation of the neonate. The vast majority of these studies explored features related to the developmental increase of continuity of the EEG [54, 141], the shift towards higher frequencies [142, 150] and the changes in functional brain connectivity [114, 133]. Moreover, previous studies have investigated how the correlation dimension of a single EEG channel evolves at increasing age [88, 169]. They reported a positive correlation between the dimensional complexity and the neonate's age. Furthermore, they observed a significant difference in EEG complexity between the preterm and full term population assessed at the same postmenstrual age [88].

In this chapter, multiscale entropy will be employed to quantify the complexity of the EEG time series. This method has been successfully applied to a wide range of biological signals, such as heart rate variability [209], EEG [136], and electromyography (EMG) [219]. During ageing or disease, a reduction of the complexity of the physiological system is generally observed due the reduced capacity of the system to adapt to stress [120]. The underlying idea for this study is that—in contrast with the loss of the complexity during ageing—the rapid changes in the developing brain will result in more complex brain dynamics. We expect that this increase of complexity of the brain dynamics will be reflected in the EEG.
The specific objective of this study was to examine whether EEG complexity measured by multiscale entropy is related to the postmenstrual age. If these complexity features can be used to accurately estimate the infant's postmenstrual age, the discrepancy between the estimated age and the real postmenstrual age of the infant can serve as an indication of the neonate's brain maturity [150]. A large deviation from the real age (typically more than two weeks) might indicate a maturational delay, while an accurate estimation indicates normal brain development [150, 154]. In this study, a correlation analysis is

performed to examine the relationship between the EEG complexity and the postmenstrual age. Afterwards, a linear regression model is trained to estimate the neonate's postmenstrual age based on a set of EEG complexity features. The performance has been tested on a large database of preterm patients with good outcome to prove that the extracted complexity features can serve as a brain maturation index.

The chapter first gives a brief overview of the database and the procedure to compute multiscale entropy. The correlation and regression analysis used to examine the relationship between the EEG complexity features and the postmenstrual age are then thoroughly explained. The last section of the methodology is concerned with the topological changes of the complexity features during maturation. Afterwards, the results of the correlation, regression and topological analysis are presented. Finally, the findings of the study will be discussed and suggestions for further work will be given.

8.2 Materials and methods

The same database and feature set is used as in Chapter 4, so we will only provide a brief overview here. More detailed information can be found in section 4.2.1 and 4.2.3, respectively.

8.2.1 Database

The dataset used to investigate the EEG complexity in relation to the postmenstrual age consists of serial EEG measurements from 26 infants born before 32 weeks of gestation. Their electrical brain activity was measured at least twice during their stay in the NICU at a postmenstrual age between 27 and 42 weeks. The 8-channel EEG (Fp1, Fp2, C3, C4, T3, T4, O1, O2) was registered at a sampling rate of 250 Hz. Quiet sleep periods were annotated by two expert clinicians upon agreement.

8.2.2 EEG preprocessing

First, the EEG was downsampled, with an anti-aliasing low pass filtering, to 125 Hz. The resampled EEG signals were then bandpass filtered between 1 and 20 Hz to reduce the contamination by artefacts.

8.2.3 Multiscale entropy

The multiscale entropy of each preprocessed EEG recording is then computed in order to assess its complexity. For this purpose, the EEG time series was segmented into nonoverlapping windows of 100 s. Per channel, the 100 s EEG segments are coarse-grained for scales τ from 1 to 20. Next, the sample entropy of each coarse-grained time series is computed to assess the regularity or predictability of the signal. The embedding dimension m was chosen equal to 2, while the tolerance r is set at 0.2 $\times$ the standard deviation of the EEG segment. Eventually, this results in a measure of the regularity of the signal at 20 scales, which can be represented by a multiscale entropy curve.

Figure 8.1 illustrates how the multiscale entropy curve of an EEG recording (channel C4) changes with increasing postmenstrual age. Five different multiscale entropy curves are plotted, corresponding to 5 serial EEG recordings measured between 29 and 39 weeks PMA from a neonate born at 24 weeks of gestation. Each multiscale entropy curve is computed as the average across all quiet sleep segments within that particular EEG recording. In Figure 8.1 there is a clear trend of increasing entropy values for increasing postmenstrual age. This shift indicates that the EEG recordings increase in complexity with development of the brain.

8.2.4 Feature extraction

Four features are then extracted from each multiscale entropy curve. First, the complexity index which is computed as the area under the multiscale entropy curve. Second, the average slope of the multiscale entropy curve between scales 1 and 5. Third, the average slope in the large scales (scale factor τ: 6 - 20). At last, the maximum value of the multiscale entropy curve. As all features are computed for each EEG channel separately, this will result in 32 features (8 channels $\times$ 4 features).
Based on the example in Figure 8.1, an increase of the complexity index, the slope in the small scales and the maximum value is expected. However, the trend for the average slope in the large scales is less clear. In order to investigate the relationship between the complexity features and the patient's age properly, a correlation and regression analysis will be performed. The findings of Chapter 4 indicate that the EEG complexity of the EEG is different depending on the sleep state, therefore it is important to account for them during the maturational analysis. Since quiet sleep is less prone to artefacts and shows more subtle alterations of brain function [57], only quiet sleep periods are considered during the correlation and regression analysis.

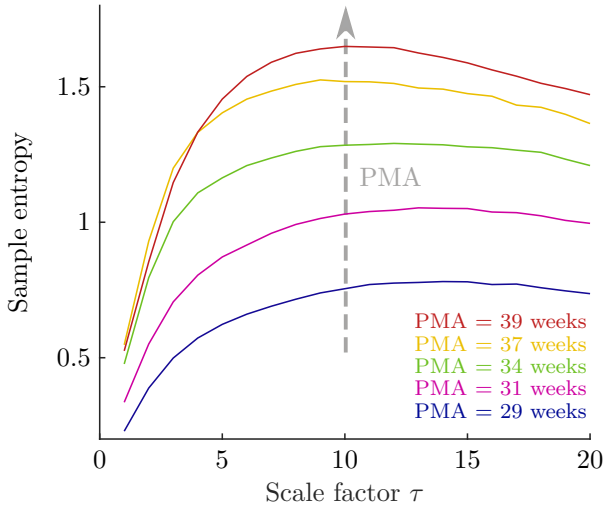


Figure 8.1: Multiscale entropy curves of EEG recordings measured between 29 and 39 weeks postmenstrual age. The multiscale entropy curve shifts upwards with increasing PMA, indicating an increase of EEG complexity with brain maturation.

8.2.5 Correlation and linear regression analysis

To prepare the feature matrix for the correlation and linear regression analysis, all data points during quiet sleep within a recording were averaged, resulting in one data point for each recording (97 in total). Thus, the feature matrix can be written as $\mathbf{X} \in \mathbb{R}^{N \times d}$, with $N = 97$ recordings and $d = 32$ features.

The Pearson correlation coefficient ρ between each of the EEG complexity features and the response variable, the PMA, was computed. Moreover, the p-value corresponding to the null hypothesis stating that there is no relationship between the complexity feature and the PMA is stored. Significance level was set at 0.05.

To increase the robustness and reliability of the results, the regression analysis was repeated 100 times, and in each iteration the dataset was randomly split into a training and test set with a $70 - 30$ ratio. The training set was used to build the model, while the prediction error of the model was assessed on the independent test set. The median and interquartile range of the performance measures are reported. For each complexity feature, a linear regression model explaining the relationship between the EEG feature and the PMA was estimated on the

training set. The coefficient of determination, R^2, was then assessed on the test set as:

$$R^2 = 1 - \frac{\sum (y_{test} - \hat{y}_{test})^2}{\sum (y_{test} - \bar{y}_{test})^2}, \tag{8.1}$$

where y_{test} denotes the observed response variable, $\bar{y}_{test}$ is its mean, and $\hat{y}_{test}$ is the corresponding predicted value [5]. Finally, the PMA estimation is evaluated on the test set using the root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum (y_{test} - \hat{y}_{test})^2}. \tag{8.2}$$

In the second step, several complexity features were combined in a multiple linear regression model. First, features that did not have a significant correlation with the response variable were removed. Afterwards, selection of the most relevant features was carried out using the least absolute shrinkage and selection operator (LASSO) algorithm. This method adds a regularization term to the loss function of the regression. This penalty term forces regression coefficients to zero and tends to avoid collinear predictors [189]. The LASSO regularization was performed 100 times, and the features that were selected more than 80 times were retained to build the multiple linear regression model. To assess the performance of this age-prediction model, the coefficient of determination and the RMSE are evaluated on the test set during 100 iterations. Their median and interquartile range will be reported.

8.2.6 Topological analysis

In addition to the evolution of EEG complexity with age, topological differences are of great interest as well. Since the complexity features were computed for each of the EEG channels, it is possible to investigate how the complexity changes are dependent on the brain region. For this analysis we focused on the complexity index (area under the multiscale entropy curve) and we divided the recordings into four age groups: postmenstrual age below 31 weeks, from 31 to 34 weeks, from 34 to 37 weeks, and above 37 weeks. The spatial map was then constructed by computing the grand average of the complexity index during a specific sleep stage (quiet sleep, non-quiet sleep) for a specific age group. In addition, we explored whether there is a difference in EEG complexity between the two brain hemispheres.

Table 8.1: The top of the table provides the results of the correlation and regression analysis for each feature that has been selected by the least absolute shrinkage and selection operator (LASSO) algorithm. At the bottom of the table the performance of the multiple regression model combining all 16 features is presented. The correlation coefficient ρ for the complete dataset, and the median (interquartile range) of the R^2 and root mean squared error (RMSE) for the test set are set out. The last column indicates how many times each feature was selected by LASSO.

Feature(s)		$\boldsymbol{\rho}$	$\mathbf{R^2_{test}}$	**RMSE (weeks)**	**# selected LASSO**
Complexity Index	C3	0.76	0.54 (0.19)	2.17 (0.52)	99
	T3	0.81	0.66 (0.14)	1.93 (0.42)	100
Average slope in small scales	Fp1	0.76	0.56 (0.19)	2.17 (0.49)	99
	Fp2	0.75	0.53 (0.19)	2.19 (0.39)	99
	C3	0.78	0.55 (0.19)	2.16 (0.45)	99
	C4	0.80	0.64 (0.14)	1.96 (0.40)	100
	T4	0.80	0.61 (0.15)	2.01 (0.56)	100
	O1	0.78	0.61 (0.17)	2.03 (0.38)	99
	O2	0.77	0.56 (0.16)	2.15 (0.36)	100
Maximum of MSE curve	Fp1	0.77	0.59 (0.15)	2.12 (0.30)	99
	Fp2	0.76	0.52 (0.19)	2.25 (0.52)	99
	C4	0.79	0.60 (0.18)	2.03 (0.39)	96
	T3	0.8	0.61 (0.15)	1.98 (0.49)	99
	T4	0.79	0.61 (0.17)	2.00 (0.44)	100
	O1	0.75	0.55 (0.18)	2.19 (0.38)	99
	O2	0.74	0.52 (0.11)	2.31 (0.29)	99
Combination of 16 selected features			0.69 (0.13)	1.88 (0.36)	

8.3 Results

8.3.1 Correlation and linear regression analysis

The significance of the correlation between each EEG complexity feature and the PMA at the recording time was examined. The eight features representing the average slope in the large scales were removed during further analysis, since for the majority of the channels this feature did not have a significant correlation with age. All the remaining EEG complexity features had a significant correlation coefficient in the range from 0.74 to 0.81, indicating a strong positive correlation with PMA.
Out of the remaining 24 complexity features, the LASSO algorithm selected a subset of 16 features. These features are represented in the first column of Table 8.1. As can be seen from the last column of Table 8.1, these features were selected at least 96 times. The results of the correlation and simple linear regression models for each of the chosen complexity features are presented in the top part of Table 8.1. Each of the single predictors could estimate the age of the neonate with an RMSE between 1.93 and 2.31 weeks. The performance of the multiple regression model, where all 16 features selected by LASSO are combined, is presented in the last row of Table 8.1. Using this model, the age of the neonate could be estimated with a root mean square error of 1.88 weeks and a coefficient of determination equal to 0.69.

Figure 8.2 visualizes how the EEG complexity changes during maturation. Figure 8.2a illustrates the fitted linear regression model for the complexity index of electrode T3. The boxplots in Figure 8.2b show the complexity index averaged among all EEG channels for both sleep stages grouped according to PMA. It is apparent from these boxplots that the average complexity index increases with age in both sleep states. Moreover, there is a clear difference between the two sleep stages, confirming the results of Chapter 4.

8.3.2 Topological analysis

The topoplots of the grand average of the complexity index during quiet sleep and non-quiet sleep are presented in Figure 8.3a and 8.3b, respectively. The color indicates the value of the complexity index at the different brain areas, and the black dots represent the position of the eight EEG electrodes.
To begin with, the colorbars show a rise of the complexity index with postmenstrual age within each sleep state. On top of that, comparison of the colorbars between the vigilance states for the same age group reveals that complexity index is generally higher during non-quiet sleep compared to quiet

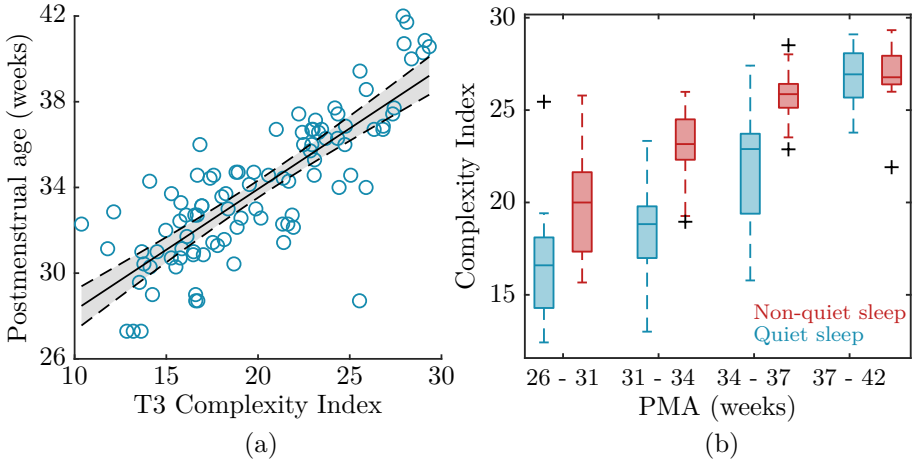


Figure 8.2: (a) The relationship between the complexity index of channel T3 and the postmenstrual age (PMA) fitted by simple linear regression. Each blue circle represents a data point (corresponding to 1 EEG recording), while the black solid line shows the regression line. The 95% confidence bounds are indicated by the dashed lines and the corresponding confidence interval is highlighted in light grey. (b) Boxplots of the complexity index averaged over all channels for both quiet sleep (blue) and non-quiet sleep (red). A clear increase of electroencephalogram (EEG) complexity can be observed in both sleep stages. Moreover, the complexity index is higher during non-quiet sleep compared to quiet sleep.

sleep. These findings confirm the results of the correlation analysis described above and is in line with the trends that can be seen in the boxplots in Figure 8.2b.

From Figure 8.3a, it can be seen that the central derivation is dominant below 31 weeks PMA during quiet sleep. In the second age group, from 31 to 34 weeks PMA, the EEG complexity exhibits the highest value in the central and occipital regions. From 34 weeks PMA onwards, the temporal electrodes gain importance as well. Beyond 37 weeks PMA, the spatial maps remain static, with highest EEG complexity in the central, temporal, and occipital regions.

During non-quiet sleep, the topological changes in brain complexity are less pronounced. As during quiet sleep, central complexity predominates at the youngest age. Between 31 and 34 weeks PMA, a spread towards the frontocentral region can be observed. In the oldest age groups, the frontotemporal areas exhibit the highest complexity index. However, conclusions drawn from these topoplots should be interpreted with caution, as only the group average was investigated and the variability among the recordings was disregarded. Finally,

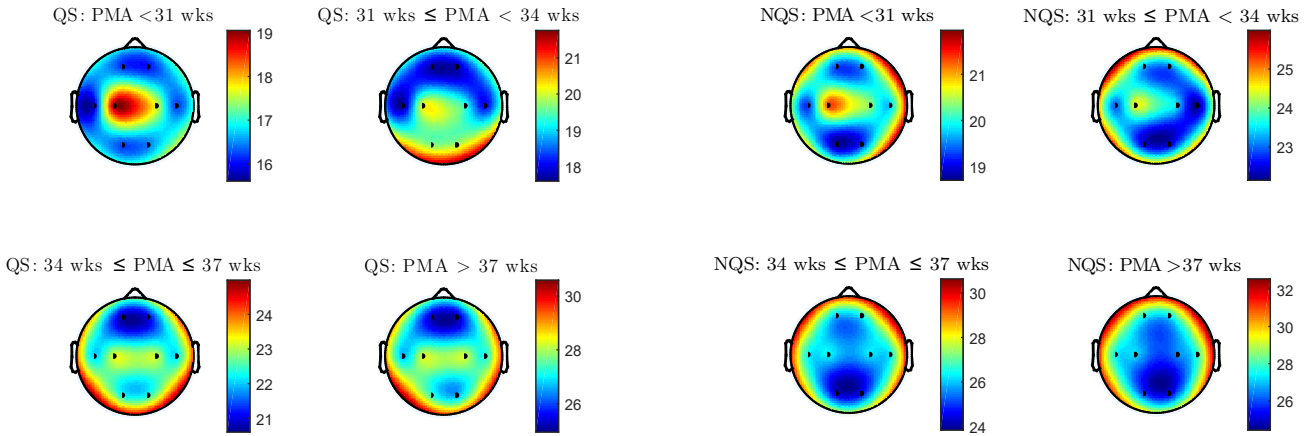
we also compared the complexity of left hemisphere to the complexity of the right hemisphere, but could not observe significant differences.

8.4 Discussion

This study was set out with the aim of assessing the brain maturity in preterm infants based on the neural complexity of multichannel EEG recordings. The first part of the analysis demonstrated that there was a significant positive correlation between the features derived from the multiscale entropy curve of the EEG signals and the PMA. Moreover, we proposed a multiple regression model which can accurately estimate the neonate's age using a set of EEG complexity features. To conclude, the spatiotemporal plots of the complexity suggest that the central area is most complex below 31 weeks in both sleep states. During quiet sleep, the predominance of the central area shifts towards the occipital and temporal regions with increasing age.

The maturational trend observed in the complexity features confirms the hypothesis and is in line with the results of prior studies [88, 169]. This increase in EEG complexity can be attributed to changes in the dynamics of the underlying neural networks during cortical maturation [89]. In this paper a regression model, relying solely on complexity features, is used to predict the patient's PMA. The best estimate of the PMA has a RMSE of 1.88 weeks, which is comparable to the results reported in [113, 150]. In addition, this is similar to an expert electroencephalographer who can estimate the neonate's PMA within approximately 2 weeks based on the EEG appearance [118, 154].

As the complexity was assessed for each EEG channel separately, spatiotemporal information could be extracted. This topological analysis can shed more light on the electrophysiological aspects of the study. One of the hallmarks of preterm EEG are delta brushes. At first these complexes tend to appear in the central brain regions; from around 31 weeks onwards they become more predominant in the temporal-occipital regions [7]. Eventually, they are mainly observed in the occipital regions (Figure 2.5). This pattern, with the dominant area evolving from the central towards temporal-occipital and occipital regions, corresponds to what is observed in the topoplots of the complexity during quiet sleep. The patterns in the spatial maps during non-quiet sleep are more difficult to interpret. The dominance of the central regions at the youngest age corresponds to what has been found during quiet sleep, but we could not find a clinical explanation for the spread towards the frontal and temporal areas with increasing age. This might be due to the fact that non-quiet sleep incorporates both active sleep and being awake. Moreover, artifacts are more likely to occur



(a) Quiet sleep

(b) Non-quiet sleep

Figure 8.3: The topoplot of the grand average of the complexity index during (a) quiet sleep (QS) and (b) non-quiet sleep (NQS). During quiet sleep the predominant region shifts from the central lobe towards the temporal-occipital regions. During non-quiet sleep, the central regions are initially predominant. Afterwards, the frontotemporal areas exhibit the largest complexity index.

during non-quiet sleep and might have affected the spatial maps.
The presented research has important clinical implications. To begin with, the results of the analysis show that the neonatal EEG complexity is a biomarker of brain maturation and can assist clinicians in evaluating the neurological function of the infant. Moreover, aberrations in the complexity features can play a major role in identifying neonates at increased risk for neurodevelopmental delay which need close cot-side EEG monitoring. In this way, neuroprotective interventions can be started early after diagnosis and the efficacy of the therapy can be monitored. In the present study, promising results were obtained using multiscale entropy to analyse the complexity of the neonatal EEG. However, extensions of multiscale entropy or other advanced methods to assess signal complexity can be adopted in future investigations [137, 217]. One of the limitations of the current study is that the EEG was only recorded in neonates with normal neurodevelopmental outcome up to 42 weeks PMA. According to the existing literature and the well-known "loss of complexity", a reduced complexity of the brain dynamics is expected in preterm infants with poor outcome [120, 125]. As a result, we expect that the brain-age regression model would underestimate the postmenstrual age of neonates with poor outcome. In order to confirm this hypothesis, further research in a cohort with unfavourable prognosis must be carried out. Moreover, to develop a full picture of the EEG complexity, future investigations must establish the behaviour of the EEG complexity beyond 42 weeks PMA. On top of that, an important issue for future research is to reveal how the observed complexity changes are related to structural changes in the underlying brain structure. For this purpose, a simultaneous neuroimaging study would be of great interest. Since the results presented in this paper strengthen the idea that the EEG complexity is highly useful in tracking brain maturation, and linear and nonlinear features can provide complementary information, a natural progression of this work is to combine the presented nonlinear complexity features with previously developed features to quantify neurodevelopment.

8.5 Conclusion

This study has identified a significant positive correlation between the complexity of the EEG signal and the postmenstrual age in preterm neonates. A multiple regression model based on a combination of EEG complexity features derived from the multiscale entropy curves could estimate the age of the neonates accurately with a RMSE of 1.88 weeks. These findings indicate that EEG complexity is able to quantify and track the early functional development of the brain. In summary, these results show that EEG complexity is a valuable feature in the neuromonitoring of preterm infants.

Chapter 9

Relationship between early functional and structural brain development in preterm infants

The previous chapter established that EEG complexity features are highly useful to track the brain maturation in preterm infants. This chapter starts with a multicenter validation of these results on a dataset consisting of 115 aEEG recordings from 52 neonates measured between 24 and 31 weeks postmenstrual age. In addition, this chapter aims to provide insight into the relationship between early functional and structural brain development in preterm infants. For this purpose, the association between features extracted from two complementary modalities, the EEG and MRI, is inspected. The relationship between a set of maturational EEG features quantifying early brain function and MRI metrics, characterizing the structural brain development and injury, is examined.

9.1 Introduction

The immature brain of extremely preterm infants is prone to brain injuries, therefore monitoring and promoting optimal structural and functional brain development is key in the current neonatal intensive care units (NICUs). Electroencephalography is a cheap and non-invasive tool widely used to perform long-term brain monitoring of preterm infants in the NICU. A great deal of previous research has focused on how the patterns and characteristics of the EEG evolve during maturation. Various studies have shown that these maturational features extracted from serial EEG recordings can be used to construct a brain-age model and thereby assess the neonate's brain function [50, 54, 150, 179]. Moreover, there is growing evidence that these maturational features are useful in the prognosis of long-term neurodevelopmental outcome [78, 99, 172, 212]. In addition to this, magnetic resonance imaging (MRI) is a valuable neuroimaging modality to detect and assess pathologies in the preterm brain, such as white matter injury and cerebellar haemorrhages, which are difficult to detect using cranial ultrasound [41, 84]. Moreover, sequential magnetic resonance imaging allows longitudinal assessment of the structural brain development, such as integrity and growth of brain structures [14].

Although extensive research has focused on evaluating the preterm brain function using (automated) analysis of the EEG and many studies have investigated MR images to examine the structural development of the brain, little is known about the relationship between the function and structure of the developing brain in preterm infants. Only a relatively small body of recent literature explores the relationship between early brain function assessed via EEG, and brain morphology evaluated using MRI in the preterm population. Benders et al. have investigated the correlation between early brain activity and structural brain development [16]. On the one hand, early brain activity was assessed by quantifying the EEG amplitude and the occurrence of spontaneous activity transients (SATs) during the first postnatal days. On the other hand, the structural metrics were measured based on segmentation of two sequential MRI recordings, one performed as early as possible after birth (around 30 weeks postmenstrual age) and one at term equivalent age (TEA). This study revealed that patients with increased level of early brain activity have a greater volumetric brain growth during the following weeks. Similarly, Tataranno et al. studied the association between early cortical network activity and the subsequent growth of cerebellar and cortical volumes and microstructural brain development [187]. A significant positive correlation between early brain activity, quantified using SAT rate and the inter-SAT interval (ISI) duration, and cerebellar and cortical grey matter growth between 30 and 40 weeks PMA was found. Moreover, a higher SAT rate was accompanied by increased fractional anisotropy in the

corpus callosum. These results are supported by a recent study by Hüning et al., which explores the relationship between aEEG and MRI features, and their power to predict the neurodevelopmental outcome at corrected age of 24 months [84]. The patterns of the aEEG recordings measured during the first 72 h after birth were assessed by means of the Burdjalov score [24]. The Burdjalov scoring system is based on visual evaluation of the continuity, sleep-wake cycling, bandwidth and amplitude depression of the aEEG in order to objectively assess the cerebral maturity [24]. Volumetric brain measurements were extracted from the MRI recordings and a Bayley Scales of Infant Development II was used to assess the neurodevelopmental outcome at 2 years corrected age. They confirmed the strong relationship between early brain activity and subsequent brain growth and observed that the combination of functional and structural brain parameters can aid in outcome prognosis [84].

So, few studies have investigated and confirmed the relationship between early brain function and structure in a systematic way. However, none of these studies have investigated EEG parameters beyond the first three days after birth. Moreover, the automated extraction of features from the EEG is limited to the computation of SAT-related parameters, while many other maturational EEG features have proven useful in the quantification of early brain function [50, 142]. At last, all previous studies were evaluated on a relatively small cohort of preterm infants.

The current study has two primary aims. First, the main goal of this study is to investigate the relationship between maturational features extracted from serial aEEG measurements during the first weeks after extremely premature birth and features derived from MRI recordings measured around 30 weeks PMA and at TEA. In contrast with the previously described studies, the brain function is not only assessed during the perinatal adaptation period the first 72 h after birth, but with serial aEEG recordings up to five weeks after birth. Moreover, in this study different aspects of the aEEG trace will be quantified (e.g. complexity, spectral content, etc.). A secondary objective is a multicenter validation of the maturational EEG features. More specifically, we want to verify whether previously developed maturational EEG features also show a strong relationship with postmenstrual age on 2-channel EEG measurements of a younger patient population acquired using a different monitor.

The remaining part of this chapter proceeds as follows. First, the acquisition and characteristics of the dataset consisting of both aEEG and MRI recordings are described. Afterwards, the preprocessing and feature extraction of both modalities are discussed. The next section is concerned with the methodology used for this study. On the one hand, it explains how the relationship between the postmenstrual age and the maturational features is assessed. On the other hand, the correlation and regression analysis used to investigate the relationship

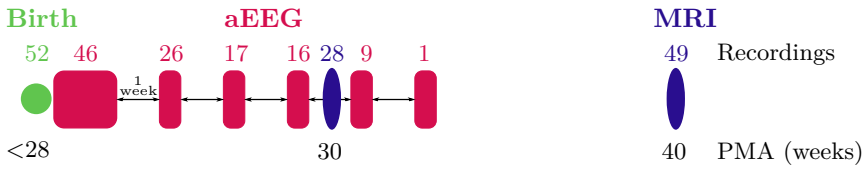


Figure 9.1: Visualization of the dataset used for the analysis. All babies are born before 28 weeks of gestation. A first 72h aEEG recording is measured immediately after birth, followed by weekly sequential recordings. Note that not all neonates are born at the same GA, hence the first recording of one patient can be measured at the same PMA of the third recording of another patient. In 28 patients, an MRI is recorded as soon as possible after birth, around 30 weeks PMA. Additionally, an MRI is performed at term equivalent age (40 weeks).

between the EEG and MRI features are described. Finally, the results of the study are presented and discussed.

9.2 Materials and methods

9.2.1 Database

The analysis was performed on aEEG and MRI recordings measured as part of the NEOBRAIN study between May 2008 and October 2010 at the Neonatal Intensive Care Unit of the Wilhelmina Children Hospital (Utrecht, The Netherlands). In total 106 infants were enrolled in this study, all born extremely preterm, at a gestational age below 28 weeks. Patients with congenital disorders or syndromes were excluded from the study. The medical ethical review committee approved the use of these data for research purposes. A schematic overview of the different aspects of the dataset is given in Figure 9.1.

aEEG acquisition

Bedside aEEG monitoring of each infant was started as soon as possible after birth for the first three days of life (72 hours). In addition, a sequential aEEG recording with a duration between 2 and 4 hours is performed each week for the first 4 up to even 5 consecutive weeks after birth. However, some weekly aEEG registrations could not be performed due to the unstable condition of the neonate.
The aEEG time series were measured using four subcutaneous needle electrodes

positioned on the parietal and frontal lobes (F3-P3 and F4-P4) according to the international 10–20 system. Additionally, a reference electrode measuring the impedance is placed on the central lobe. The aEEG was acquired at a sampling frequency of 256 Hz using either the BRM2 or BRM3 BrainZ monitor (BRM2/BRM3, BrainZ; Natus, Seattle, WA). The BRM2 and BRM3 monitor apply a highpass filter with a cutoff frequency at 2 Hz and 0.5 Hz, respectively. Note that aEEG was monitored because it requires fewer electrodes (easier placement and less invasive for the infant) and allows interpretation at the bedside, however the acquired raw EEG signals will be used in further analysis.

MRI acquisition

MRI was performed at term equivalent age (PMA: 40 - 44 weeks) in all infants born below 28 weeks of gestation as standard of care at the Wilhelmina Children's Hospital (WKZ). Moreover, some patients also had an MRI around 30 weeks PMA. The MRI was performed on a 3T MR system (Achieva, Philips Medical Systems, Best, the Netherlands) using a SENSE head coil. In order to reduce movement artifacts, the infants were positioned in a vacuum pillow. Moreover, all infants received double-layer hearing protection using Minimuffs (Natus Medical Incorporated, San Carlos, CA) and Earmuffs (EM's 4 Kids, Brisbane, Australia) to reduce noise exposure.

The scanning protocol included coronal T2-weighted imaging (turbo spin echo, at 30 weeks: repetition time 10085 ms; echo time 120ms; slice thickness 2 mm, in-plane spatial resolution 0.35 x 0.35 mm^2; at TEA: repetition time 4847-6293 ms; echo time 120-150 ms; slice thickness 1.2 mm, in-plane spatial resolution 0.35 x 0.35 mm^2, full brain coverage) for the calculation of brain volumes. A neonatologist or physician assistant, with experience in neonatal neurology, was present during the whole examination to check the patient's condition and quality of the MRI images. Oxygen saturation, respiratory and heart rate of the infants were continuously monitored. All infants were sedated using oral chloral hydrate 50 – 60 mg/kg.

9.2.2 Preprocessing and feature extraction

This section describes the preprocessing performed on the EEG and MRI recordings. The maturational features extracted from the preprocessed EEG epochs and metrics related to brain development or injury derived from the segmented MR images are then discussed. An overview of all the features computed from on the one hand the EEG recordings, and on the other hand the MR images is given in Table 9.1.

Table 9.1: Overview of maturational features extracted from EEG recordings and developmental and injury measures extracted from MR images. The two rightmost columns indicate which features were extracted from the first (PMA: 30 weeks) and/or the second (PMA: 40 weeks) MRI recording.

Modality		Feature	Rec 1	Rec 2
EEG		• Relative power in frequency subbands (delta, theta, alpha, beta)		
		• Spectral edge frequency 75%/90%		
		• SAT rate complete recording/5min		
		• Median IBI duration		
		• Complexity index		
		• Mean slope small scales (1-5)		
		• Maximum value MSE curve		
MRI	**Development**	• Cerebellar volume		✓
		• Cortical grey matter volume		✓
		• Total brain volume		✓
		• Cerebellar height (mid-sagittal)	✓	✓
		• Cerebellar width (mid-sagittal)	✓	✓
		• Cerebellar width (coronal)	✓	✓
	Injury	• Total white matter injury score		✓
		• White matter classes		✓
		• Grey matter classes		✓
		• Ventricle left	✓	✓
		• Ventricle right	✓	✓
		• Interhemispheric fissure	✓	✓

EEG preprocessing

Automated artifact detection based on features proposed by Stevenson et al. [180], i.e. the amplitude, standard deviation, peak frequency and spectral distortion of the EEG, was applied on each 2-channel EEG recording. This analysis revealed that many EEG recordings are heavily distorted by artifacts. Therefore, the quality of the EEG recordings was visually evaluated in order to exclude severe abnormalities, artifacts or caretaking events. During this manual data preselection a representative one hour EEG epoch is selected taking into account the sleep-wake cycling of the infant.
From each of the three consecutive days of the first 72 hour EEG recording, the best quality EEG epoch is selected. In four patients, the first EEG recording did not span the complete period of three days and only two 1h EEG epochs

could be chosen. If the quality of the subsequent weekly recording allowed it, a 1 hour epoch of clean EEG was visually selected. Thus, some poor quality EEG recordings were discarded from the database and all further processing is performed on EEG epochs with a duration of one hour. In total 115 EEG recordings from 52 neonates are retained after visual assessment of the quality. In 46 of these neonates the first recording of 72 h as soon as possible after birth is available, 26 recordings are measured approximately one week after birth, 17 were recorded around two weeks after birth, 16 recordings are measured in the subsequent week and respectively 9 and 1 recording are registered at 4 and 5 weeks after birth. Figure 9.1 illustrates how many EEG recordings are retained for further analyis at each time point.

Each of the manually preselected 1h EEG epochs are then bandpass filtered to preserve the electrocortical activity of interest, while reducing the contribution of low- and high frequency artifacts. The bandwidth of a normal preterm EEG signal ranges from around 0.5 Hz up to 40 Hz. However, the BRM2 BrainZ monitor has a built-in highpass filter with cutoff frequency 2 Hz. Therefore, all signals had to be filtered with a highpass filter with 2 Hz cutoff frequency in order to compose a homogeneous database. Thus, all EEG time series are filtered between 2 and 40 Hz using a FIR highpass filter with cutoff frequency 2 Hz followed by FIR lowpass filter with cutoff frequency 40 Hz. In addition, a notch filter at 50 Hz is used to remove any remaining powerline interference.

EEG feature extraction

After preprocessing, a set of features are extracted from the EEG time series. These features assess three characteristics of the changing EEG patterns during maturation: 1) the spectral content, 2) the continuity and 3) the complexity of the EEG time series.

First of all, a set of commonly used spectral features, the relative power in the delta (δ: 2 – 4 Hz), theta (θ: 4 – 8 Hz), alpha (α: 8 – 12 Hz) and beta (β: 12 – 30 Hz) frequency band, are extracted. Moreover, the spectral edge frequency (SEF) with the edge at 75% and 90% are computed. These spectral features are derived from the power spectral density estimated using Welch's method, computed in nonoverlapping 25 s EEG segments for each channel.

The second type of EEG features, are event-based EEG measures related to the continuity of the signal: the burst percentage or SAT% and the median interburst interval (IBI) duration. The bursts are detected using the algorithm proposed by Palmu et al. [153]. This algorithm applies the nonlinear energy operator (NLEO) to the bandpass filtered EEG time series. The smoothed absolute value of the output of NLEO is then thresholded to identify the location of the SATs. Subsequently, the SAT percentage among the complete recording,

the average SAT percentage in nonoverlapping windows of 5 minutes and the median IBI length among the complete EEG recording are computed. Figure 9.2 illustrates the NLEO-based detection of SAT events and interburst intervals on a 5 min EEG segment.
At last, the complexity of each nonoverlapping 100 s EEG segment is quantified using multiscale entropy [38, 50]. The multiscale entropy curve, representing the sample entropy as a function of the scale τ, is constructed per channel for scales ranging from 1 to 20. Three features are derived from this multiscale entropy curve: the complexity index which is computed as the area under the curve, the average slope in the small scales (τ: 1 – 5) and the maximum value of the curve.

For all above-mentioned maturational features, the grand average among all EEG segments within the 1 hour selected clean EEG epoch is computed in order to obtain one feature value per recording (for each channel). In case of the first long recording, the average among the three clean 1 hour EEG epochs is computed. Moreover, all features are averaged across the two channels. This averaging procedure relies on the fact that similar characteristics are expected in both hemispheres and will reduce the impact of short-duration distortions of the signal. As a result, one robust feature is obtained for each 2-channel EEG recording.

MRI preprocessing

MR image quality was checked by two neonatologists with more than 20 years experience in neonatal neuroimaging. Only in 58 out of 106 infants the MRI was considered of appropriate quality to perform image postprocessing and compute brain volume measures.

MRI feature extraction

In all images of appropriate quality, automatic MR image segmentation was performed with the atlas-free method of Gui et al. [76]. This yields segmentations of the following regions: cortical grey matter, unmyelinated white matter, myelinated white matter, subcortical grey matter, cerebrospinal fluid (CSF), cerebellum and brain stem. These segmentations are used to extract the following tissue volumes: cortical grey matter (GM), unmyelinated white matter (UWM), myelinated white matter (My), subcortical grey matter (SGM), cerebral spinal fluid (CSF), brain stem and cerebellar volumes. Total brain volume (TBV) was calculated as the sum of all brain region volumes, including the CSF. An example of a segmented MRI is presented in Figure 9.3,

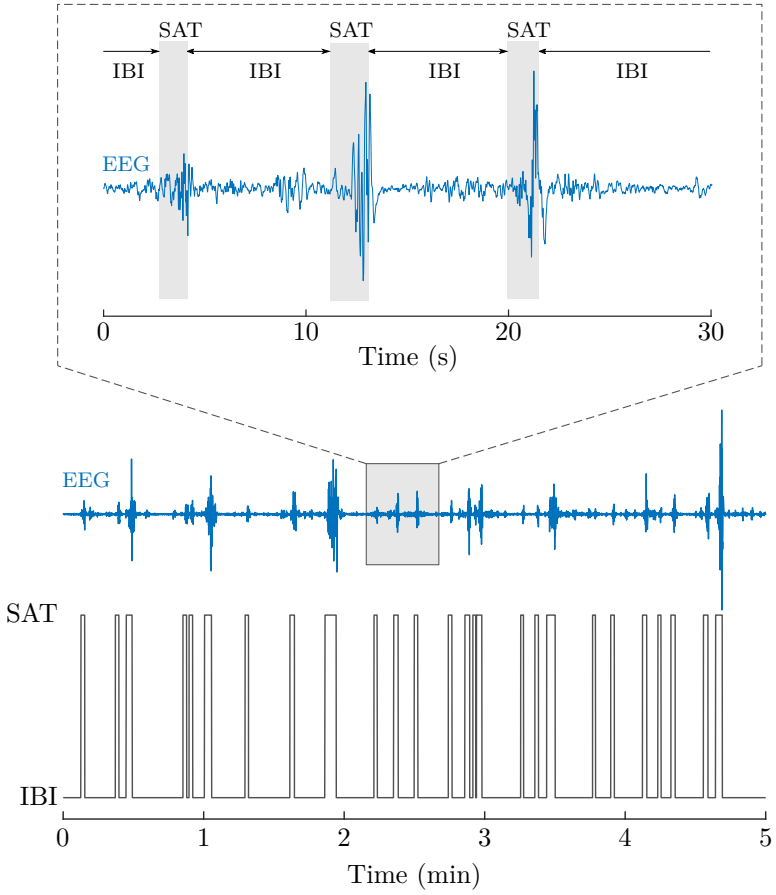
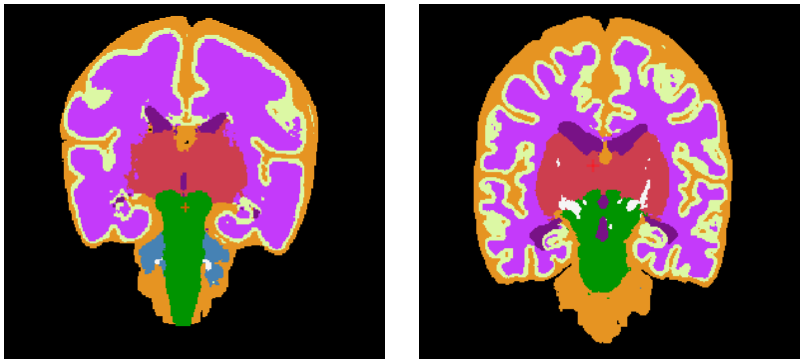


Figure 9.2: Illustration of detection of spontaneous activity transient and interburst intervals in the preterm EEG. At the bottom, a 5 min EEG segment and the output of the NLEO-based SAT detection algorithm are presented. On top, a 30 s excerpt of the 5 min EEG segment is shown with labelled SAT events and IBI periods.



(a) MRI at 30 weeks PMA (b) MRI at 40 weeks PMA

Figure 9.3: Examples of an automatically segmented MRI at (a) 30 weeks PMA and (b) 40 weeks PMA. Volumes are indicated according to the following colour scheme: red: deep grey matter; light purple: unmyelinated white matter; dark purple: ventricles; yellow: cortical grey matter; orange: cerebral spinal fluid; white: myelinated white matter; green: brain stem.

where the colours indicate different tissue volumes. Because the MRI recordings were not all measured at the same postmenstrual age, brain volumes were adjusted for postmenstrual age at the moment of the scan according to [93]. The brain volumes expected to be the most associated to early brain function (cerebellar volume, cortical grey matter volume, total brain volume) are retained for analysis as listed in Table 9.1.
In addition to volumetric measurements, the MRIs were also scored for brain injury by two experienced neonatologists according to Kidokoro et al. [98]. The Kidokoro MRI score assesses both brain injury and impaired brain growth. The MRI metrics derived from this assessment are the linear measurements of the cerebellar size in the coronal and mid-sagittal plane and all the brain injury metrics mentioned in Table 9.1.

9.2.3 Brain maturation quantification using EEG features

A number of EEG features have been developed to track brain maturation in preterm infants. Many researchers have utilised characteristics related to the continuity of the EEG, because this is one of the main criteria in the visual assessment of neonatal EEG. However, more recently also spectral and

complexity features have been proposed. Since these features have not all been developed or validated on 2-channel EEG recordings of such a young preterm cohort, the first step in the analysis consists of investigating the relationship between each of the extracted maturational EEG feature and the infant's postmenstrual age (PMA) at the moment of the recording.
In total 115 EEG recordings from 52 neonates measured at a postmenstrual age ranging from 24 weeks and 5 days to 31 weeks and 6 days are used for the analysis. The Pearson correlation coefficient ρ is computed between each of the EEG metrics and the PMA. Subsequently, a regression model is built to investigate the relationship between these two variables. As multiple EEG recordings are measured from the same neonate, these recordings cannot be regarded as independent from each other. Therefore, a linear mixed-effects model accounting for the repeated measurements is implemented. A random intercept and slope were introduced to correct for inter-patient variability with the subject as grouping variable. For all statistical tests, the significance level was set at $p < 0.05$.

9.2.4 Relationship between early brain function and structure

Early EEG metrics versus structural brain development and brain injury

After establishing which EEG features are suitable to track brain maturation in this group of preterm neonates, the relationship between the maturational EEG features and each of the MRI measures is investigated. In this first step, a similar approach to the existing literature [16] is taken. This means that the EEG features extracted from the first aEEG recording, during the first 72 hours after birth, are associated to MRI recordings performed at 30 and/or 40 weeks PMA. This analysis can be performed on all 43 infants for which a clean first EEG recording and a high quality MRI recording is available.
In addition, we will examine whether neonates with increased early brain activity are characterized by increased brain growth between approximately 30 and 40 weeks PMA [187]. This relationship can only be explored for the subset of 25 patients that underwent two serial MRI recordings and for the 3 developmental brain metrics that have been extracted from both MRI recordings. Therefore, only the change in size of the cerebellum in the mid-sagittal and coronal plane can be studied (see Table 9.1). The structural brain growth (increase/decrease per week) is computed as the difference between the feature at the second and first MRI recording divided by the number of weeks in between the recordings.

In the same way as described in the previous section 9.2.3, the relationship between the early brain function and structure is evaluated using a correlation and regression analysis. The neonates included in this analysis are born at

different gestational ages (GA: 24 weeks 5 days - 27 weeks 6 days). Moreover, 26 out of 43 infants received morphine as sedation on mechanical ventilation. These variables might affect the found association. Therefore, the partial correlation between each of the EEG features and each of the MRI parameters, controlling for the gestational age and morphine administration, are computed. Since only the first EEG measurement is used, all observations are independent and there is no need to include random effects in the regression model. However, gestational age and morphine administration are included as additional independent variables in the model in order to control for confounding factors [132]. This results in a multiple regression model with an EEG feature, gestational age and morphine administration as independent variables, and the MRI metric as dependent variable. Thus, for each possible combination of EEG and MRI (growth) metrics, the partial correlation is computed and a multiple regression model is built.

Change in EEG metrics versus structural brain development and injury at term equivalent age

In the second part of exploring the relationship between brain function and structure, the serial EEG recordings will be exploited. As most of the neonates' electrocortical activity is recorded at multiple time points, namely immediately after birth and in the subsequent weeks, a trajectory of the maturational EEG feature can be constructed per neonate. For the purpose of this analysis, all patients with multiple EEG recordings are selected and the slope of these patient-specific maturation trajectories are estimated using a linear regression model. The hypothesis that will be tested is that faster increase of early brain activity, a greater slope of the maturation trajectory, will be reflected in increased structural brain development on the MRI recording.
In total 32 neonates had multiple EEG measurements and an MRI recording at term equivalent age, but for only 17 of these patients the early MRI was also available. Because of this small sample size, the relationship with features extracted from the early MRI measurements were not investigated.
For each of these 32 patients a linear regression model is used to estimate the slope of the maturational feature as a function of PMA, expressed as increase/decrease per week. The association between the change in early brain activity and each of the MRI metrics derived from the term recording is then explored by means of partial correlation and multiple linear regression accounting for the confounding factors GA and morphine administration.

9.3 Results

9.3.1 Brain maturation quantification using EEG features

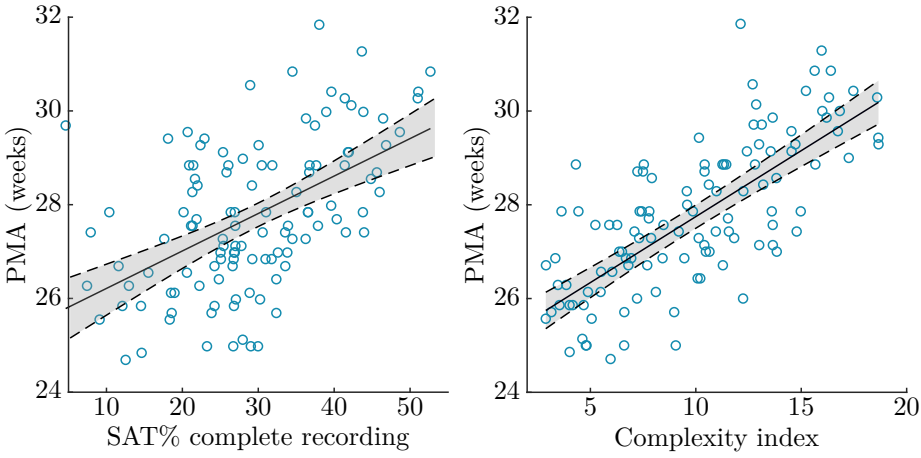


Figure 9.4: The fixed effects of the regression model fitting the relationship between SAT% and PMA are shown on the left. The regression model on the right shows the association between the complexity index and the PMA. Each blue circle represents a data point (corresponding to 1 EEG recording), while the black solid line shows the regression line. The 95% confidence bounds are indicated by the dashed lines and the corresponding confidence interval is highlighted in light grey.

The results of the correlational analysis, in particular the correlation coefficient ρ and the corresponding p-value, between each of the maturational EEG features and the PMA are presented in the left half of Table 9.2. The right half of the table shows the characteristics of the mixed-effects regression model, consisting of the coefficient of determination or R^2, the regression coefficient b and its 95% confidence interval and p-value.
Among the spectral features, the relative power in the beta band and the spectral edge frequency 90% are significantly correlated with the postmenstrual age of the neonate. Moreover, the regression coefficients of these predictors are significantly different from zero. However, the coefficient of determination is for both models equal to 0.07. This low R-squared value indicates that these models are not reliable in predicting the age of the infant.
As expected the SAT percentage computed in the complete recording or averaged within 5 minute windows has a significant positive correlation with PMA, while the median IBI duration decreases with increasing age. These trends reflect

the gradual increase of EEG continuity with maturation. The null-hypothesis that the regression coefficient is equal to zero was rejected and the R-squared value is equal to 0.27, 0.26 and 0.20, respectively. The left graph in Figure 9.4 shows the fixed effects of the regression model fitting the relationship between the SAT% and the PMA.
Finally, all maturational features related to the complexity of the EEG signal are strongly correlated with PMA with a correlation coefficient between 0.70 and 0.73. Moreover, the fitted regression models have a significant positive slope and a coefficient of determination equal to 0.71 or 0.72. This strong positive correlation is also illustrated in Figure 9.4 where the fixed effects of the regression between the complexity index and the PMA are visualized. It is apparent from Table 9.2 that the complexity features derived from the multiscale entropy curve are the most associated with the age of the baby, followed by the SAT-related features. Based on these results, we expect that these features will show the strongest relationship with the brain structure metrics derived from the MRI.

Table 9.2: The results of the correlation and regression analysis performed to assess the relationship between the maturational EEG features and the postmenstrual age at the moment of the recording. The table reports the Pearson correlation coefficient ρ and its p-value. Moreover, the coefficient of determination R^2, the regression coefficient b and its 95% confidence interval (ci_{lower},ci_{upper}) and p-value are set out.

EEG Feature	**Correlation**		**Regression**		
	ρ	p(ρ)	R^2	b(ci_{lower},ci_{upper})	p(b)
Relative power δ band	-0.12	0.21	0.04	-3.29(-9.2;2.62)	0.27
Relative power θ band	-0.02	0.82	0.05	-2.6(-10.45;5.24)	0.51
Relative power α band	0.15	0.12	0.08	13.57(-3.85;30.98)	0.13
Relative power β band	0.19	0.04	0.07	13.17(0.52;25.83)	0.04
Spectral edge frequency 75%	0.13	0.15	0.05	0.29(-0.11;0.69)	0.15
Spectral edge frequency 90%	0.26	0.00	0.07	0.34(0.11;0.57)	0.00
SAT% complete recording	0.52	0.00	0.27	0.08(0.06;0.10)	0.00
Mean SAT% 5 min	0.51	0.00	0.26	0.08(0.05;0.10)	0.00
Median IBI duration	-0.29	0.00	0.20	-0.24(-0.36;-0.12)	0.00
Complexity index	0.73	0.00	0.72	0.28(0.24;0.33)	0.00
Mean slope small scales (1-5)	0.70	0.00	0.71	35.43(29.59;41.26)	0.00
Maximum value MSE curve	0.72	0.00	0.71	4.52(3.79;5.24)	0.00

9.3.2 Relationship between early brain function and structure

Early EEG metrics versus structural brain development and brain injury

Since there are a tremendous amount of possible combinations of EEG and MRI features, only the EEG and MRI metrics which were significantly related will be presented. The results of the correlation and regression analysis for significantly correlated features are set out in Table 9.4. It can be seen that mainly features related to the size of the cerebellum at either 30 weeks or 40 weeks PMA, or the cerebellar growth are strongly associated with early brain activity. The complexity features are positively correlated with the height of the cerebellum early after birth, while the SAT-related features are more associated to the cerebellar size measured at term age. The relative power in the higher frequency bands and the spectral edge frequency 90 are negatively related to the size of the cerebellum at TEA. In addition, the relative power in the beta band is also inversely correlated with the volume of the cortical grey matter and the total brain volume.
The last three rows of Table 9.4 show the metrics for brain injury that were significantly correlated with early brain activity. The relative power of the EEG in the beta band is positively correlated with the total white matter injury score, while two complexity features are negatively correlated with the size of the right ventricle on the early MRI.

Change in EEG metrics versus structural brain development and injury at term equivalent age

The results of the exploration of the association between the change in EEG metrics and the term MRI recording are presented in Table 9.3. Among the MRI metrics, mainly the brain injury metrics, especially the size of the ventricles, are closely related to the change in EEG metrics. A greater increase of the maximum value of the multiscale entropy curve is associated with a smaller left and right ventricle size. Surprisingly, the slope of spectral features in the lower frequency ranges showed significant trends with the interhemispheric fissure (IHF), ventricular sizes and the cerebellar volume. The change in the relative power in the delta and theta frequency band is respectively negatively and positively associated with the ventricular size.

Table 9.3: The relationship between the change in maturational EEG feature and the significantly correlated MRI metrics at term equivalent age. The two leftmost columns show the MRI and EEG features of interest, respectively. The results of the correlation and regression analysis are presented at the right.

MRI Feature	**EEG Feature**	**Correlation**		**Regression**		
		ρ	p(ρ)	R^2	b(ci_{lower},ci_{upper})	p(b)
Cerebellar volume	Relative power delta band	-0.4	0.04	0.18	-73.43(-142.34;-4.51)	0.04
IHF	Relative power theta band	0.39	0.03	0.2	21.92(1.75;42.1)	0.03
Ventricle left	Relative power delta band	-0.4	0.04	0.15	-25.84(-49.99;-1.7)	0.04
	Relative power theta band	0.45	0.01	0.2	41.6(9.8;73.41)	0.01
	Maximum value MSE curve	-0.4	0.04	0.14	-5.64(-11.13;-0.14)	0.04
Ventricle right	Relative power delta band	-0.4	0.01	0.21	-33.03(-58.92;-7.15)	0.01
	Relative power theta band	0.68	0	0.47	69.06(40.14;97.98)	0
	Maximum value MSE curve	-0.4	0.03	0.17	-6.65(-12.66;-0.65)	0.03

Table 9.4: The relationship between early brain activity and features extracted from subsequent MRI recordings. The left column shows the MRI feature and the recording it was derived from (recording 1: around 30 weeks PMA, recording 2: around 40 weeks PMA). The next column lists the significantly associated EEG features and the right part represents the results of the correlational and regression analysis. The correlation coefficient ρ and its corresponding p-value p(ρ), the coefficient of determination R^2 and the regression coefficient b, expressing the change in the MRI feature per unit increase of the EEG feature, and its p-value p(b) are provided.

MRI Feature	**Rec**		**EEG Feature**	**Correlation**		**Regression**		
	1	2		ρ	p(ρ)	R^2	b(ci_{lower},ci_{upper})	p(b)
Cerebellar height (mid-sagittal)	✓		Complexity index	0.44	0.04	0.31	0.27(0.01;0.54)	0.04
	✓		Mean slope small scales	0.47	0.03	0.34	37.08(4.35;69.8)	0.03
Cerebellar width (mid-sagittal)		✓	Relative power alpha band	-0.45	0	0.23	-50.34(-82.75;-17.92)	0
		✓	SAT%	0.55	0	0.33	0.13(0.07;0.19)	0
		✓	Mean SAT% 5min	0.55	0	0.33	0.13(0.06;0.19)	0
		✓	Median IBI duration	-0.42	0.01	0.2	-0.23(-0.38;-0.07)	0.01
Growth cerebellar width (mid-sagittal)	✓	✓	SAT%	0.49	0.02	0.25	0.01(0;0.02)	0.02
	✓	✓	Mean SAT% 5min	0.49	0.02	0.25	0.01(0;0.02)	0.02
Cerebellar width (coronal)		✓	Relative power beta band	-0.56	0	0.43	-115.08(-171.7;-58.47)	0
		✓	SEF90	-0.37	0.02	0.29	-1.27(-2.31;-0.23)	0.02
Cerebellar volume		✓	SAT%	0.34	0.03	0.13	0.22(0.02;0.42)	0.03
		✓	Mean SAT% 5min	0.34	0.03	0.13	0.22(0.02;0.42)	0.03
		✓	Median IBI duration	-0.48	0	0.25	-0.73(-1.16;-0.3)	0
Cortical grey matter volume		✓	Relative power beta band	-0.37	0.02	0.15	-550.68(-1004.25;-97.11)	0.02
Total brain volume		✓	Relative power beta band	-0.32	0.05	0.12	-1065.85(-2131.15;-0.55)	0.05
Total white matter injury score		✓	Relative power beta band	0.31	0.05	0.12	32.8(0.1;65.49)	0.05
Ventricle right	✓		Complexity index	-0.43	0.05	0.22	-0.41(-0.82;-0.01)	0.05
	✓		Mean slope small scales	-0.47	0.03	0.25	-57.42(-107.8;-7.04)	0.03

9.4 Discussion

In this study, the relationship between the function and structure of the developing brain of preterm infants is investigated. It was hypothesized that increased brain function during the first weeks after birth will be reflected in greater structural brain development. The link between brain function and structure was evaluated by quantifying the relationship between maturational features extracted from serial EEG recordings and the structural metrics extracted from MRI recordings at 30 and 40 weeks postmenstrual age. If the maturational features extracted from the EEG during the first couple of weeks after birth could provide information about the subsequent structural development of the brain, earlier intervention would be possible and neonates at risk of abnormal brain growth or brain injury can be monitored more closely.

The first research question, however, was to validate the EEG features previously proposed to track early brain maturation in this patient group. For this purpose a correlation analysis and linear mixed-effects model, assessing the direction and strength of the relationship between each of the EEG features and the postmenstrual age of the patient, are used. Consistent with the literature, this study has found that the continuity of the neonatal EEG increases progressively during brain maturation. This is characterized by an increase of the burst percentage and a concomitant decrease of the median interburst interval duration. The SAT% computed over the complete 1 hour epoch of EEG and the average for consecutive 5 minute segments are very similar, which is an indication of clean brain activity with little distortion due to artifacts.
In contrast to earlier findings [142, 149], no maturational trend in the relative power of the delta, theta, alpha frequency band or spectral edge frequency 75 could be observed. This result may be explained by the preprocessing of the EEG signals. Due to the different filter settings of the two BrainZ monitors (BRM2 and BRM3), all EEG time series had to be filtered with a highpass filter with a cutoff frequency of 2 Hz, removing an important part of the delta frequency band. Consequently, all very low frequency information of the EEG (0.5 – 2Hz), which contains essential information in the early preterm infant, is lost. Only the relative power in the beta band (12 – 30 Hz) and spectral edge frequency 90 show a significant positive correlation with PMA. However, the coefficient of determination of the corresponding regression models is only equal to 0.07, indicating a poor fit.
The last set of EEG metrics assessed the complexity of the time series using features extracted from the multiscale entropy curve. These complexity features show the strongest relationship with PMA, with a correlation coefficient around 72% and a coefficient of determination of approximately 71%. These results are consistent with the findings of Chapter 8, hence they serve as multicenter

validation on a younger patient group with only 2-channel EEG recordings. As the continuity and complexity features are the strongest related to the PMA of the neonate, we expect the evolution of these features to be the most related to the structural development of the brain.

The main objective of the study was to determine whether early functional and structural development are intertwined. First, the association between brain function during the first three days after birth was compared to measures of both MRI recordings as well as to the growth of the cerebellum between the two measurements.
Two out of three complexity features are strongly positively correlated with the height of the cerebellum measured in the mid-sagittal plane at 30 weeks PMA. These complexity features are also negatively correlated with the size of the right ventricle on the early MRI. Hence, more complex early EEG patterns seem to be related to the size of the cerebellum, while reduced EEG complexity is related to ventricular enlargement, indicating brain injury.
The width of the cerebellum in the mid-sagittal plane at 40 weeks PMA is strongly associated with the decrease of EEG discontinuity. Moreover, increased burst percentage and shorter interburst intervals shortly after birth are associated to greater cerebellar volume at TEA. At last, a higher SAT percentage is also positively correlated with the subsequent increase of the cerebellar width in the mid-sagittal plane. This finding is in line with the results of Tataranno et al. [187] who found that the number of SATs correlated with the cerebellar growth between 30 and 40 weeks PMA. SATs are the main hallmark of the preterm EEG and previous studies have suggested their pivotal role in structural brain development [16, 201] and in the development of cortical circuits. The cerebellar volume almost triples during the last trimester of gestation, being one of the structures with the largest relative growth during this period. Moreover, Limperopoulos et al. demonstrated that injury to the cerebellum is associated with long-term cognitive impairment in extremely preterm infants [119].
Interestingly, these significant relationships between the complexity and continuity of the early EEG and the cerebellar size are observed in the mid-sagittal plane, but not in the coronal plane. Moreover, the EEG complexity features are related to the early MRI, while the EEG continuity features have significant associations with the MRI recording at TEA.
The most striking result to emerge from the data in Table 9.4 are the trends found for the spectral features. During maturation, the frequency of the preterm EEG increases slowly, which is captured by an increase of the spectral edge frequency and the relative power of the higher frequency bands, while the relative power in the low frequency bands decreases. However, the cerebellar width, cortical grey matter volume and total brain volume at term age are inversely correlated with the relative power in the higher frequency bands (alpha and beta band). Moreover, the relative power in the beta band is related to the

total white matter injury score. These unexpected findings indicate that higher relative power in the beta band might be associated with worse structural brain development.

Second, the increase/decrease of the EEG feature per week PMA are compared to the features of the term MRI recordings. A greater increase of the complexity during the first weeks after birth is related to smaller left and right ventricles. This demonstrates that EEG complexity might be suitable to predict brain injury. The change in the relative power in the delta and theta band are also associated to the ventricular size at TEA.

The key strength of the current study is that different aspects of the EEG time series are characterized. In contrast to earlier studies, we have not only looked at EEG features related to burst activity, but also assessed the spectral content and complexity of the time series. Moreover, a broad range of MRI metrics quantifying either the development or the injury of the brain have been considered. On top of that, the relation between the function and structure of the developing brain is extensively analysed in a systematic way. All possible interactions between EEG and MRI features have been examined. Furthermore, both early and term MRI measures, and the growth between the two recordings, have been explored. In addition, the brain function is not only quantified during the first postnatal days of the neonate. This is the first research where the maturational trend computed over serial EEG recordings is compared to the structural brain development.
However, a number of important limitations should be considered. Even though the current database consists of a larger patient group compared to previous research, there are some important drawbacks related to the dataset. Visual preselection of 1h EEG epochs was required because many of the EEG recordings are highly contaminated by artifacts. Due to this, a large proportion of each EEG recording was discarded and the processing is not completely automated, making it infeasible for use in clinical practice. Many of the weekly recordings were not even considered for analysis, as the poor quality did not allow the selection of 1 hour of clean data. As a result of the small sample size of subsequent weekly recordings, these recordings could not be separately related to each of the MRI features and the estimation of the slope of the maturation trajectory is less reliable. In addition, the fact that different BrainZ monitors were used to record the signals, led to the loss of important spectral information from the delta frequency band. Besides, the study is limited by the lack of clinical annotations of sleep-wake cycling. Because of this, the analysis had to be performed for the complete EEG epoch, rather than on each sleep stage separately. At last, the number of early MR images and the number of volumetric measures computed from these images is limited (e.g. total brain volume, cerebellar volume and cortical grey matter volume are only extracted at TEA). Future studies on the

current topic are therefore recommended.

9.5 Conclusion

In this study, the relationship between two complementary modalities EEG and MRI, providing information about the brain function and structure respectively, is investigated. First of all, we investigated which EEG features can be used to predict the postmenstrual age of the neonate, and in this way quantify the brain maturation. Both SAT percentage and EEG complexity were significantly positively correlated with PMA, while the median IBI duration was inversely associated with the age. This provides further support for the fact that with ageing of the infant, the continuity and complexity of the preterm EEG steadily increases.
A correlational analysis and multiple regression models are used to explore the relationship between early brain function and structural brain development. This analysis revealed that complexity features extracted from an EEG recording during the first three days after birth have a significant positive correlation with the cerebellar size around 30 weeks PMA, while the event-based measures are related to the cerebellar size at term equivalent age.
The current study provided additional insight into the early brain development, however further studies are needed to unravel the dynamics between the early functional and structural development of the brain in preterm infants.

Chapter 10

Measurement of thyroid hormone action in the preterm infants' brain using EEG

The previous chapters have proven that automated analysis of the EEG is an objective method to assess brain maturation in preterm infants. In this chapter, established maturational EEG features will be used to study the effect of thyroid hormone levels on early brain development in prematurely born infants. Thyroid hormones are indispensable for brain development and neonates born at an extremely low gestational age (ELGAN) often suffer from transient hypothyroxinemia of prematurity (THOP). The effect of this condition is a controversial and much disputed topic. In this chapter we hypothesized that in case of low circulating thyroid hormone levels, brain maturation will be affected, which will be reflected by the maturational EEG features. In order to assess the thyroid function, circulating free thyroxine (fT4) levels on day 0 (time 1) and at the end of the first week of life (time 2) were collected and delta (Δ) fT4 ($= fT4_{time2} - fT4_{time1}$) was calculated in 63 ELGANs. Brain maturation was quantified using complexity and spectral features extracted from the EEG around term equivalent age (TEA). Linear regression models are then used to investigate the relationship between the change in fT4 level during the first week of life and the brain maturation assessed via EEG features at TEA.

10.1 Introduction

Thyroid hormones (THs) are important developmental hormones in different kind of species [110]. In humans, they play a vital role in fetal and neonatal developmental processes in general, and in particular in the developing brain [44, 70]. This is illustrated in Figure 10.1. By activation of numerous thyroid responsive genes, they regulate several neurodevelopmental processes, such as neurogenesis, myelination, dendrite proliferation and synapse formation [18]. These actions occur in specific time windows and are initiated by binding of the active hormone T3 to nuclear receptors encoded by the thyroid hormone receptor genes (THRA, THRB). Most T3 is produced by deiodination of T4 in peripheral tissues [206]. Also in the human fetus, cerebral T3 availability is primarily generated by local T4 deiodination [94]. Fetal TH metabolism starts progressively from mid-gestation. Up to that point, the fetus is completely dependent of trans-placental maternal TH supply, which remains present until birth. Only then, the infant's thyroid hormone system acts completely autonomously [68]. However, in preterm birth, the trans-placental maternal TH supply is suddenly interrupted while the infant's thyroid system is still immature. These elements are contributing to the development of transient hypothyroxinemia of prematurity (THOP). THOP is characterized by low circulating total and free T4 concentrations, without the expected increase in pituitary thyroid-stimulating hormone (TSH) secretion. This phenomenon is more severe in patients born at lower gestational age (GA) [68]. For more than three decades, THOP has been a debatable research topic. Heterogeneity in THOP definitions, difficulties with TH assessment, identifying patients at risk and a lack of sufficiently powered studies add to the controversy [62]. Studying the trends in circulating TH levels in the first week of life has been proposed recently as a new approach to THOP. The difference between TH levels at the end of the first week of life (time 2) and cord blood (time 1) is calculated, the so called Δ, and it has been shown that immaturity is the most important contributing factor to a negative Δ [63].
Since THs play a critical role in the early development of the brain, we expect alterations in the brain maturation of ELGANs with THOP. The EEG is the most common tool for continuous bedside monitoring of the cerebral function of infants in the neonatal intensive care unit. The fast brain development during the first weeks after birth is reflected in a continuously changing EEG. Automated analysis of these rapidly changing EEG patterns can be used to quantify the brain maturation in order to monitor the brain development and to predict the neurodevelopmental outcome of the infant. Various characteristics of the neonatal EEG have been investigated, such as the continuity of the EEG [54, 141] the (relative) power in specific frequency bands [91, 142, 150], the spatial organization of the EEG [112, 133] and its complexity [50]. To the

best of our knowledge, the impact of THOP on brain maturation quantified using EEG features has never been assessed before. Nevertheless, diffuse slowing of background activity on EEGs in adults with hypothyroidism has been demonstrated [95]. In pediatric patients with congenital hypothyroidism, altered sleep state organization on polysomnographic studies at toddler age has been shown [167]. Finally, in a chicken-model of brain development, mild hypothyroidism was associated with a delay in developmental changes in basal EEG patterns [11]. Therefore, we hypothesized that temporary low circulating fT4 levels in preterm infants on a critical moment of their brain development might affect brain maturation, which might be reflected by alterations in electrocortical maturation on the EEG.
Previous research has established that spectral features and the complexity of the EEG are powerful characteristics to track brain maturation in preterm infants [50, 91]. Therefore, we decided to assess the brain maturation in this study by means of EEG complexity and frequency domain features. Features related to the continuity of the EEG were considered less appropriate since the EEG is recorded around term equivalent age. The aim of this study was to investigate whether a delayed brain maturation could be observed in ELGANs with lower TH levels in the first week of life.

10.2 Materials and Methods

10.2.1 Database

Patient selection

In this single-center retrospective study, quantitative EEG maturation at (near) term age (PMA: 35 weeks 2 days - 46 weeks 5 days) in 65 preterm infants born before 28 weeks of gestation (GA: 23 weeks and 6 days - 27 weeks 5 days), was analysed in relation to the evolution of their fT4 levels in the first week of life. The EEG signals were collected in the framework of the NeoGuard [57] and Resilience study (June 2012 – May 2017) at University Hospitals Leuven. Relevant clinical data were collected: gestational age (GA), birth weight, presence of intrauterine growth restriction (IUGR), CRIB score, moderate or severe bronchopulmonary dysplasia (BPD, need for supplemental oxygen/ventilation at 36 weeks GA), retinopathy of prematurity (ROP) with need for laser therapy, persistent ductus arteriosus (PDA), necrotizing enterocolitis (NEC, defined as Bell's stage II & III), sepsis with positive blood culture, cranial ultrasound abnormalities (grade II-IV IVH and localized infarct), levothyroxine treatment after the first week of life. TH supplementation

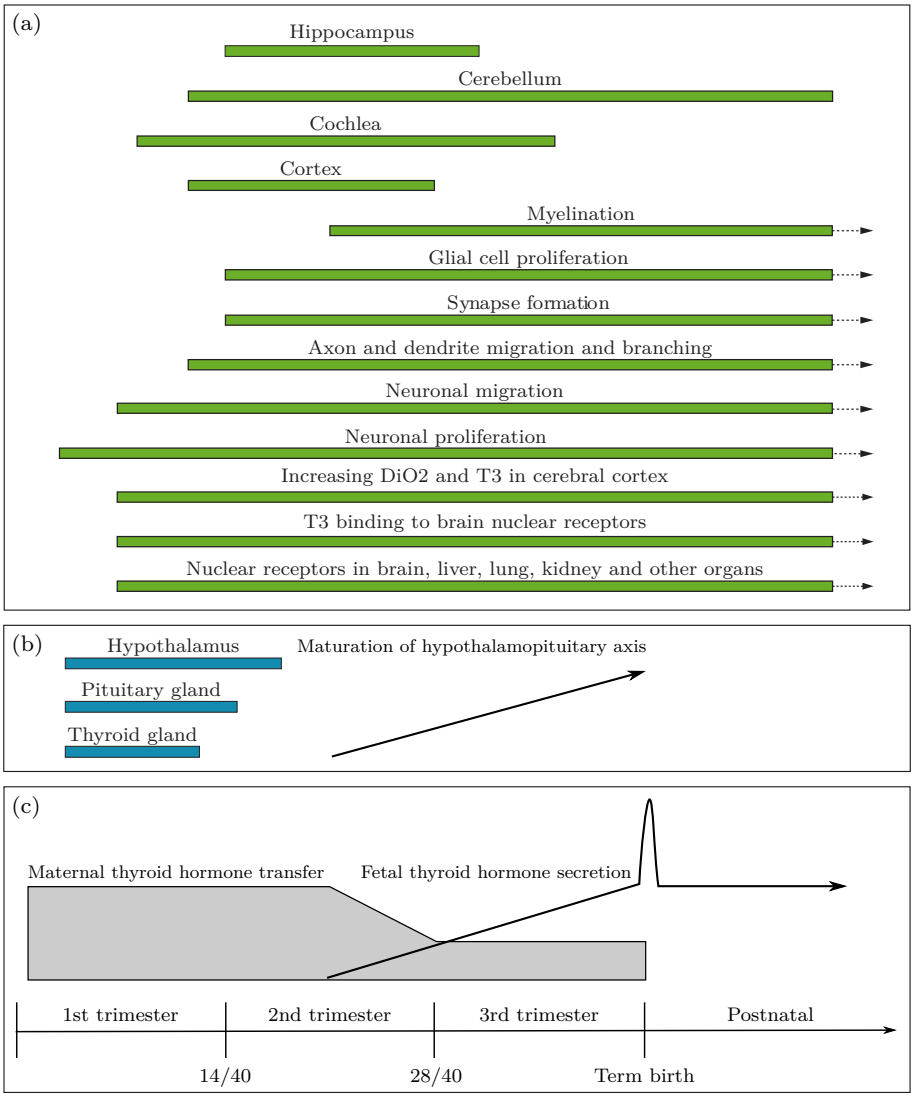


Figure 10.1: Fetal brain development in relation to maternal thyroid hormone supply and fetal thyroid hormone metabolism. (a) Embryological and fetal development of brain structures; (b) Embryological and fetal formation and maturation of thyroid hormone system structures; (c) Until mid-gestation, the fetus is completely dependent on maternal thyroid hormone supply. Thereafter, the fetal thyroid system starts to work, but it is only after birth that the infants' thyroid system functions completely autonomously. DIO2: type 2 deiodinase, T3: triiodothyronine. Adapted from [18].

(levothyroxine 10 μg/kg) for 14 days was initiated when fT4 levels were below 0.8 ng/dL at the end of the first week of life, according to the treatment regimen of the unit. Parental informed consent was obtained. The study was approved by the local ethical committee of the University Hospitals Leuven (S61028).

EEG monitoring

All infants had an overnight polysomnography recording with multichannel EEG, electrocardiogram, oxygen saturation, chin electromyogram, 2 electrooculograms, piezo-electric belts to measure abdominal and thoracic respiratory effort, and a nasal thermistor for airflow monitoring at a PMA between 35 weeks 2 days and 46 weeks 5 days. The EEG recordings were measured using 9 electrodes: Fp1, Fp2, C3, C4, T3, T4, O1, O2 and reference Cz, placed according to the international 10 – 20 EEG recording system (BrainRT, OSG equipment, Mechelen, Belgium). A monopolar EEG setup was used and the reference electrode Cz was not considered during the analysis. The EEG time series were sampled at 250 or 256 Hz. The duration of the EEG recording is in the range from 3 h 27 min to 15 h 24 min. The average recording length is equal to 9 h 55 min with a standard deviation of 1 h 15 min, which is sufficient to cover all sleep stages. No visual preselection of the data was performed.

10.2.2 Thyroid hormone function

As part of the clinical standard care protocol, fT4 levels on the first day of life (preferably cord blood) and at the end of the first week of life were determined by competitive immunoassay with ECL, Hitachi/Roche-Modular E. Blood samples were primarily taken through an arterial line, or by vena puncture, when no arterial line was available. The evolution of the fT4 levels through the first week of life was calculated as:

ΔfT4 = fT4 (at the end of the first week of life) - fT4 (at the day of birth).

Statistical analysis

Statistical analyses were done using IBM SPSS Statistics. Shapiro-Wilk was used for normality testing. Normally distributed data were analysed using the independent-samples t-test. The nonparametric Mann-Whitney U test was used when data were not normally distributed. Chi-square test of independence and Fisher's exact test were used for categorical variables. Data are expressed as

mean +/- SD or median with interquartile range. Statistical significance was assumed for a two-sided p-value < 0.05.

10.2.3 Automated EEG analysis

Preprocessing

The first step in the automated analysis of EEG is preprocessing of the data. The EEG time series is bandpass filtered between 0.5 and 40 Hz in order to remove low and high frequency noise. In addition, a notch filter at 50 Hz is applied to avoid distortion by powerline interference. After filtering the EEG, the signal is downsampled with a factor two in order to reduce the computational complexity during further analysis. Afterwards, an automated artifact detection step is used to identify the segments in the filtered EEG which could contain artifacts. For this automated artifact detection, the EEG is segmented into nonoverlapping windows of 100 s. If more than 5% of an EEG segment consists of missing values or has an absolute value of the amplitude higher than 150 μV, the segment is annotated as potential artifact and is not taken into account in further analysis. In addition, segments with a standard deviation above 50 μV or below 0.001 μV are discarded. This analysis is performed per channel and if an artifact is identified on more than half of the EEG channels, the segment is excluded from all channels. At last, the complete recording is considered to be of poor quality if for more than half of the channels (> 4) more than 50% of the segments are indicated as potential artifact. Because of this, two recordings were excluded from the analysis.

Feature extraction

After preprocessing the EEG time series, the maturational features are extracted from the EEG. On the one hand, spectral features are computed, on the other hand the complexity of the EEG is assessed.

Spectral features In order to compute the spectral features, the power spectral density of each nonoverlapping 25 s EEG segment is estimated. Welch's method was adopted using a 4 s Hamming windows with 50 % overlap. The relative power in the delta 1 (δ_1: 0.5 – 1 Hz), delta 2 (δ_2: 1 – 4 Hz), theta (θ: 4 – 8 Hz), alpha (α: 8 – 12 Hz) and beta (β: 12 – 30 Hz) band is then computed by integrating the modified periodograms in these frequency bands and dividing by the total power in the band from 0.5 to 30 Hz. In addition to the relative bandpower, the spectral edge frequency (SEF) 75 and 90 are also computed.

This corresponds to the frequency below which respectively 75% and 90% of the total power of the signal is located. In total 7 spectral features are extracted from each clean 25 s EEG segment, resulting in 56 (7 features $\times$ 8 channels) features per multichannel EEG segment. The average of each feature across all EEG segments within the recording is then computed, so that one value for each feature is obtained per EEG recording.

Complexity features Similarly to the previous studies, the complexity of each clean multichannel EEG segment of 100 s is assessed by computing the multiscale entropy. A multiscale entropy curve is then constructed for each EEG segment. Three features are extracted from the multiscale entropy curve. First of all, the complexity index (CI), estimated by computing the area under the multiscale entropy curve [25, 92, 125]. This feature provides insight in the regularity of the signal over all considered scales. Second, the average slope of the multiscale entropy in the small scales, τ: $1 - 5$, is computed. The last feature is the maximum value of the multiscale entropy curve. Thus, for each multichannel EEG segment three complexity features are extracted, resulting in 24 values for each 100 s segment (three for each EEG channel). Similarly to the frequency domain features, each of the complexity features is averaged among all 100 s segments within the EEG recording. Therefore, each EEG recording is characterized by 24 complexity features.

Regression analysis

A regression analysis is performed in order to investigate whether poor thyroid function is related to abnormal brain maturation reflected in EEG features around term equivalent age. First, the relationship between the delta fT4 and each of the maturational EEG features is investigated per channel. As the EEG recordings took place at different postmenstrual ages and an increase/decrease of the extracted maturational features with PMA is expected, this might lead to biased results. To account for this confounding variable, the postmenstrual age is included in the multiple regression model. Moreover, only a subset of the cohort was treated with levothyroxine and some of the neonates suffered from an intraventricular haemorrhage (IVH) as presented in Table 10.1. Both factors are expected to influence the subsequent brain development, and as a consequence also the appearance of the EEG around term equivalent age. That is why these variables are included as additional independent variables. The independent variables can be represented in a matrix $\mathbf{X} \in \mathbb{R}^{N \times d}$, with N the number of observations, corresponding to 63 EEG recordings, and d = 4 representing the number of independent variables (delta fT4, postmenstrual age, treatment and IVH). Note that ΔfT4 and postmenstrual age are continuous variables, while

treatment and IVH are dummy variables (either 0 or 1). Eventually, for each channel a multiple linear regression model is fitted between the independent variables and each of the maturational EEG features. Then, the significance of the regression coefficient of interest, the one indicating the relationship between ΔfT4 and the EEG feature holding all confounding variables constant, was tested. The significance level was set at 0.05.
While the multiple regression model described above can indicate which features are significantly related to the fT4 level evolution for a particular EEG channel, it suffers from the multiple comparisons problem as the significance for each feature for all 8 channels is tested separately [17]. In order to reduce the risk of incorrectly rejecting the null hypothesis, we have also developed a regression model in which all EEG channels are considered simultaneously. Since the different observations are no longer independent (8 channels from the same neonate), a mixed-effects model with a random intercept using the patient as grouping variable is adopted. Moreover, the EEG channel is included as additional independent variable, because channel-dependent differences are expected relying on the fact that not all brain regions develop equally fast. This will result in a mixed-effects regression model investigating the relationship between the change in delta fT4 and each EEG maturational feature considering 504 observations (63 EEG recordings $\times$ 8 channels). In the same way as the channel-wise analysis, the significance of the regression coefficient reflecting the association between ΔfT4 and the EEG feature is tested.

10.3 Results

10.3.1 Thyroid hormone function

Both EEG recordings at term equivalent age and fT4 levels at birth and during the first week of life were available in 65 ELGANs, but 2 patients were removed from the analysis during the preprocessing phase because of poor EEG quality. 27 (43 %) neonates had a positive ΔfT4 level in the first week of life, whereas 36 (57%) neonates had a negative ΔfT4 level. Patient characteristics are summarized in Table 10.1. Gestational age was significantly lower in the negative ΔfT4 group ($p < 0.01$), (Figure 10.2a). Moreover, infants with negative ΔfT4 had a significant higher rate of supplementation therapy with levothyroxine ($p < 0.01$). Although the number of patients with intracranial ultrasound abnormalities were not significantly different between both groups, we included IVH as a confounding factor in the analysis since these brain lesions can theoretically also affect the EEG appearance. Figure 10.2b shows the fT4 levels at day of birth, at the end of the first week of life and the change between

Table 10.1: Patient characteristics

	Positive ΔfT4 (n=27)	**Negative ΔfT4 (n=36)**	**p-value**
Median GA (w) (+/- IQR)	26 (1)	26 (2)	0.007
Mean birth weight (g) (+/- SD)	910 (213)	834 (170)	0.2
IUGR (n) (%)	4 (14)	3 (8)	0.4
Median CRIB score (range)	4 (1-15)	5 (1-12)	0.1
Treatment with levothyroxine (n) (%)	4 (14)	20 (55)	0.001
BPD (n) (%)	7 (25)	18 (50)	0.07
ROP (n) (%)	8 (28)	15 (41)	0.4
PDA (n) (%)	11 (40)	25 (69)	0.04
Sepsis (n) (%)	9 (32)	20 (55)	0.1
NEC (n) (%)	1(3)	5 (14)	0.2
Cerebral lesions (n) (%)	5 (18)	16 (43)	0.04
IVH & infarcts (n) (%)	4 (14)	9 (24)	0.4
Persistent flaring (n) (%)	1 (3)	7 (19)	0.1

GA: gestational age; w: weeks; IQR: interquartile range; g: gram; SD: standard deviation; n: number; IUGR: intrauterine growth restriction; CRIB score: Clinical Risk Index for Babies (birth weight $\leqslant$ 1500 g and or gestational age $<$ 31 weeks), range 0-23; BPD: bronchopulmonary dysplasia, defined as O2 and/or ventilation need at 36 weeks postmenstrual age; ROP: retinopathy of prematurity with need for lasertherapy; PDA: persistent ductus arteriosus; sepsis: positive blood culture; NEC: necrotizing enterocolitis, defined as Bell's stage II & III; cerebral lesions: grade II-IV intraventricular haemorrhage (IVH), persistent flaring $>$ 2 weeks and cerebral infarcts.

the measurements in patients with a positive and a negative ΔfT4. It can be seen that the difference in fT4 level at the end of the first week of life and the change in fT4 is significantly different between the group with positive ΔfT4 and negative ΔfT4.

10.3.2 Automated EEG analysis

Regression analysis was used to examine whether thyroid hormone evolution during the first week of life is associated with maturational features extracted from the EEG. First, this relationship is evaluated for each EEG channel separately. Only for one of the maturational EEG features, more specifically the spectral edge frequency 75 in EEG channel T4, a significant inverse relationship with the delta fT4 was found.
Second, a mixed-effects model, fitting the relationship between delta fT4 and a

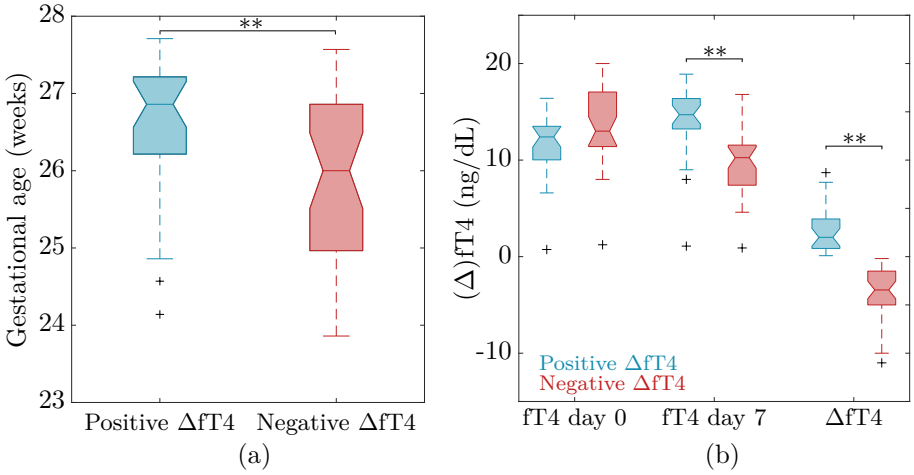


Figure 10.2: (a) Boxplot of the gestational age of infants with positive ΔfT4 and negative ΔfT4. The gestational age is significantly lower in patients with negative ΔfT4 as indicated by '**' ($p < 0.01$). (b) Comparison of fT4 levels at day of birth, at the end of the first week of life and the difference between the two measurements for the patients with positive ΔfT4 (blue) versus patients with negative ΔfT4 (red). Both fT4 level at day 7 and the change in the first week is significantly lower in patients with negative ΔfT4 ($p < 0.01$, marked as '**').

specific maturational EEG feature (including all channels), was also considered. The results of this analysis are presented in Table 10.2. As can be seen from the table, none of the maturational EEG features are significantly related to the change in free thyroxine concentrations during the first week of life ($p>0.05$).

10.4 Discussion

In this retrospective, observational study the relationship between changes in fT4 hormone during the first week of life and a set of maturational features extracted from the EEG is explored by means of a regression analysis. Only for the relationship between the delta fT4 and the spectral edge frequency 75 in the T4 channel a significant inverse association was observed. However, the p-value was close to significance level ($p=0.048$) and due to the many statistical tests, it is likely that this corresponds to a false positive. Actually, we would even expect a positive association between ΔfT4 and SEF, given that the dominant

Table 10.2: Results of the mixed-effects model. For each maturational EEG feature, the regression coefficient indicating the association between the EEG feature and delta fT4 is set out ($b_{\Delta fT4}$). Moreover, the lower and upper bounds of the confidence interval of the regression coefficient (CI(lower;upper)) and its p-value are presented.

EEG feature	$\mathbf{b_{\Delta fT4}}$	**CI(lower;upper)**	**p–value**
Complexity index	0.0423	(-0.0649;0.1495)	0.44
Average slope small scales	0.0001	(-0.0009;0.0012)	0.80
Maximum MSE curve	0.0001	(-0.0002;0.0003)	0.55
Relative power delta1	0.0029	(-0.0031;0.0088)	0.35
Relative power delta2	0.0008	(-0.0018;0.0033)	0.56
Relative power theta	0.0009	(-0.0018;0.0035)	0.52
Relative power alpha	0.0002	(-0.0019;0.0022)	0.88
Relative power beta	-0.0007	(-0.002;0.0005)	0.25
SEF75	-0.0003	(-0.0007;0.0001)	0.13
SEF90	0	(-0.0004;0.0005)	0.98

frequency of the EEG increases with ageing. Generally, the multiple linear regression model and mixed effects model, corrected for possible confounding by postmenstrual age, treatment or IVH, did not demonstrate a significant association between disturbed thyroid function and EEG features at term age.

As far as we know, this is the first study that investigates THOP in relation to functional brain maturation assessed via EEG. In contrast to our study, De Vries et al. described a relationship between hypothyroxinemia and neural maturation in very low birth weight preterm infants [48]. They demonstrated that low thyroxine concentrations are associated with a delay in progression of peripheral nerve conduction velocity in preterm infants at term equivalent age, possibly due to decreased central and peripheral myelin formation. In the TIPIT-trial, a randomized placebo-controlled trial investigating the role of prophylactic levothyroxine therapy in ELGANs, no advantage of prophylactic supplementation therapy could be demonstrated [140]. However, in a sub-study with MRI, the lowest fT4 levels were associated with markers of poorly organized brain microstructure [139]. The association between EEG and MRI findings in very preterm infants has already been demonstrated and contributed to the outcome prediction at 24 months [84]. It would therefore be of further interest to study THOP in relation to both EEG and MRI findings and in relation to the long term neurodevelopmental outcome.

Several limitations of the current study need to be considered. First of all, this was a retrospective study. Although fT4 levels at day 0 were usually obtained

in cord blood, to exclude the effect of the TH surge during the first hours of life [68], there is uncertainty whether this was the case in all infants. Albeit the TH surge in ELGANs is usually limited compared to term infants [68], increased TH levels in the first hours of life may have influenced the ΔfT4 results. An additional issue is related to the equipment used to measure the free thyroxine concentrations. Commercial kits, as used in this study, are known to underestimate the fT4 levels when there is low TH protein binding capacity. The golden standard for fT4 measurement is the expensive and labour-intensive equilibrium dialysis, nevertheless this is rarely used in the clinical setting. With this technique, depression of T4 levels and maintenance of fT4 levels at 2 weeks of age in ELGANs were described [214]. Information about whether blood samples were collected from an arterial line or by venipuncture was also not available. Heparin is used for patency of the arterial line. Heparin releases plasma lipases, leading to increased plasma free fatty acids. These compete with T4 for binding to plasma binding proteins, causing an increase in fT4 levels [90]. Therefore, these alterations might have influenced the obtained ΔfT4.
Besides, there are important differences in patient characteristics between the groups with negative and positive ΔfT4 levels. The group with negative ΔfT4 levels had a significantly lower GA. This is in accordance with the fact that immaturity is an important contributing factor to the development of THOP [198]. Although the infants in the negative ΔfT4 group were significantly younger than the infants in the positive ΔfT4 group, their CRIB scores, reflecting the severity of disease, were comparable. Moreover, as illustrated in Figure 10.3, the EEG signals are generally recorded at a higher age for neonates with a disturbed thyroid function. EEG recordings at a higher PMA are expected to reflect increased brain maturation. As a result, patients with a negative ΔfT4 evolution are biased towards better maturational EEG features. Therefore, correction for the postmenstrual age was deemed necessary. Although current evidence for THOP treatment is lacking, as is the case in several topics in neonatal medicine [213], infants with the lowest fT4 levels received levothyroxine treatment. As this might affect the patterns observed in the EEG, this is also considered as a confounding variable in the regression analysis. It is important to note that correction for confounding variables can affect the obtained results. In a previous analysis described in [61] a more strict age correction was used and a significant difference in EEG complexity between the patients with positive versus negative delta fT4 was observed in some EEG channels.
The current investigation was also limited by the lack of information on the sleep-wake cycling during the EEG recording. In this study the EEG complexity of the complete recording was considered, however it is generally known that the EEG complexity is dependent on the sleep state of the infant [50]. Hence, a different proportion of the sleep might have affected the average value of the maturational EEG feature. An additional uncontrolled factor is that part of the

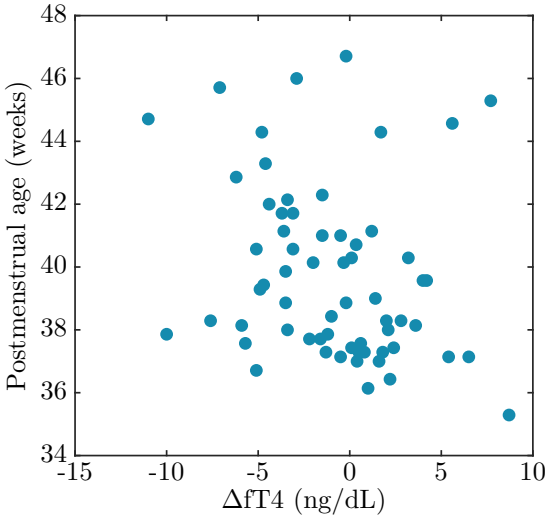


Figure 10.3: Postmenstrual age at which EEG time series were measured versus the change in free thyroxine concentration during the first week of life. It is clear that the electrocortical activity of patients with a decrease in fT4 level are measured at a later age, compared to neonates with a positive trend in fT4.

artifacts might not have been removed from the EEG during the preprocessing phase and influenced the subsequent analysis. Finally, the regression models used in this study assumed a linear relationship between the ΔfT4 levels and metrics to quantify brain maturation extracted from the EEG. However, the evolution of the maturational EEG features have been extensively studied up to around 40 weeks PMA, but little is known about the trends beyond term equivalent age. Therefore, it is suggested that further research investigates thyroid function in relation to functional brain maturation assessed at the same postmenstrual age.

Despite these limitations, this study is the first step towards a new, neurophysiological approach to obtain a better understanding of THOP. As many of the aforementioned limitations are related to the dataset, further prospective studies accounting for these weaknesses are recommended. These prospective studies exploring the impact of THOP on EEG maturation, in relation to the neurodevelopmental outcome are required to improve insights in the development and impact of THOP.

10.5 Conclusion

This study was set out to determine the relationship between the evolution of free thyroxine levels during the first week of life and functional brain maturation assessed via EEG features at term equivalent age. For this purpose, linear regression analyses accounting for possible confounding by PMA, levothyroxine treatment and IVH has been performed. No significant correlations between thyroid function and maturational EEG features were found.

Part IV

Conclusion

Chapter 11

Conclusions and Future directions

In this thesis, we focused on two important topics for automated monitoring of the brain development of preterm infants: automated EEG sleep staging and automated quantification of brain maturation. On the one hand, we proposed algorithms for automated sleep staging based on various machine learning approaches. On the other hand, we investigated biomarkers for brain maturation in preterm infants for different clinical applications. The most important findings and contributions of the work presented in this thesis will be summarized in this final chapter. To conclude, some open problems and recommendations for future work will be discussed.

11.1 Conclusions

In the introduction, the two main goals of this thesis have been defined as:

1. Development of algorithms for automated sleep stage classification
2. Automated analysis of the EEG to quantify brain maturation

In the next sections, we will summarize the findings of the presented research in relation to these objectives.

11.1.1 Automated EEG sleep staging

The first objective of this thesis was the development of algorithms for automated EEG sleep staging. In this thesis three different approaches to perform automated sleep stage classification in preterm infants have been proposed.

The first algorithm, proposed in Chapter 4, uses a set of EEG complexity features in combination with a fixed-size LS-SVM classifier to discriminate quiet sleep from non-quiet sleep. Performance evaluated on a large database of labelled preterm EEG recordings revealed that this method works well in neonates between 31 and 42 weeks PMA, but has a poor performance in neonates below 31 weeks postmenstrual age. This analysis confirmed that the EEG complexity is different depending on the behavioural state, therefore it is a highly useful feature in the context of neonatal sleep stage classification.

Chapter 5 aimed at identifying neonatal sleep stages in an unsupervised manner. The proposed algorithm uses tensor algebra, relying on the sleep state-dependent difference in EEG complexity described in the previous chapter. This novel approach based on the CPD of the multiscale entropy tensor allows to retrieve the sleep cycling of the infant in a data-driven way. However, the decision of the rank of the decomposition is still an open problem. As the postmenstrual age and/or the amount of artifacts increases, the optimal rank increases. As a consequence, a logical choice is to choose the rank dependent on the age of the neonate. Nevertheless, approaches estimating the amount of noise to define the optimal number of components could also be considered.

The last algorithm for automated sleep staging (Chapter 6) makes use of deep learning, a special field within machine learning which is omnipresent in (biomedical) signal processing nowadays. The implemented deep convolutional neural network outperforms the previously developed algorithms, while its performance is comparable to existing algorithms from the literature. The

presented CNN is considerably faster in terms of recall time, which is an important consideration for implementation in a neonatal brain monitor.

Finally, all proposed algorithms have been evaluated on the same test recordings and their performance and properties are thoroughly compared. The advantages and disadvantages of each of the proposed methods for sleep stage classification are listed and discussed in Chapter 7.

11.1.2 Automated brain maturation quantification

The second objective of this thesis is related to automated analysis of the neonatal EEG to quantify functional brain maturation. In Chapter 8 a set of EEG complexity features are proposed to track the evolution of brain development. Chapters 9 and 10 combine these complexity features with other maturational EEG features for two clinical applications.

While the complexity of the EEG was studied in relation to the vigilance states in Chapter 4, Chapter 8 examined whether features characterizing the complexity of the neonatal EEG signal can be used to track the maturity of the preterm brain. A significant positive correlation between the EEG complexity and the postmenstrual age of the infant was established. Moreover, a brain-age model based on complexity features could reliably estimate the age of the infant. As a consequence, the EEG complexity can be used as a biomarker for brain maturation.

Chapter 9 validated the results of Chapter 8 on a large dataset of aEEG recordings from a younger patient group. Moreover, the relationship between early brain function and structural brain development is investigated by means of regression models. This study revealed that the continuity and the complexity of the EEG during the first postnatal days is associated with the cerebellar size in the subsequent weeks. However, unexpected results regarding the relationship between the bandpower in frequency subbands and structural brain development were observed.

The aim of Chapter 10 is to gain insight into the effect of the thyroid function on early brain development in ELGANs. Multiple linear regression models were used to assess the relationship between the change in free thyroxine concentration during the first week of life and maturational features extracted from the EEG at TEA. These models could not identify a significant association between early thyroid hormone action and functional brain maturation at TEA.

11.2 Future directions

In this research important steps towards automated EEG sleep staging and quantification of brain maturation in preterm infants have been taken. Nevertheless, there is abundant room for further progress. In the next sections, possible refinements and extensions of the proposed algorithms will be discussed. Moreover, some interesting preliminary results which can initiate further research will be presented.

11.2.1 Automated EEG sleep staging

Neonatology is a relatively small field within medicine, hence monitoring of the neurological function in prematurely born infants is still heavily relying on visual interpretation of the EEG. Only in the last decades, there has been an increasing interest in algorithms which analyse the EEG patterns automatically. As a result, a wide variety of algorithms to perform automated sleep stage classification in adults exist, while only a handful of methods are capable of discriminating sleep stages in preterm infants, dealing with the changing patterns of the neonatal EEG. The sleep stage classifiers presented in this dissertation can successfully discriminate quiet sleep from non-quiet sleep in a wide PMA range. However, the proposed algorithms can be further improved by extending the dataset, by exploiting the temporal dependencies or by incorporating additional EEG features or other modalities. These different aspects, which can be investigated in future studies, will be thoroughly discussed in the next sections.

Dataset

All algorithms were developed using the same dataset consisting of 97 EEG recordings from prematurely born infants. Only quiet sleep periods were clinically annotated in this database, as this is the only neonatal sleep stage that can be reliably discriminated in very young neonates. As a result, our algorithms could only be trained on the detection of quiet sleep periods. However, not only quiet sleep, but also active sleep, are believed to be crucial during early brain development in preterm infants (e.g. during memory consolidation processes and synaptic remodelling). To develop a clinically relevant sleep stage classifier, future studies should focus on further differentiation of the sleep stages within non-quiet sleep (wake, active sleep and indeterminate sleep).
The 97 EEG recordings were acquired between 27 and 42 weeks postmenstrual age. However, most of the EEG registrations took place between 30 and 38 weeks PMA. As a result the presented methods could not be properly trained

and tested for neonates below 30 or above 42 weeks PMA. Therefore, additional data from very young preemies would be useful to validate the sleep stage classifiers. Besides, if a considerable amount of EEG recordings measured above 38 weeks PMA are available, we could further examine whether the proposed algorithms are able to deal with the changing sleep architecture and the transition towards four distinct sleep stages. The extension of the proposed methods to the classification of four sleep stages in term neonates can be an interesting topic for further research. This has already been done for the CNN-based method, where the network architecture was further improved and successfully applied for two- and four-stage sleep stage classification in both preterm and term infants [10]. At last, the methods should also be validated on datasets from other clinical centers in order to test the generalizability of the presented sleep stage classifiers.

Temporal dependencies

Based on the characteristics of sleep-wake cycling in preterm infants, the proposed algorithms can be further refined. First of all, sleep states do not change instantly (in the current database the minimal length of quiet sleep periods is set at 3 consecutive minutes). Moreover, the state of the current EEG segment provides information about the state probability of the next EEG segment. For example, the longer the neonate has spent in a quiet sleep, the higher the probability of transition to non-quiet sleep. Likewise, shortly after the sleep state transition, the sleep stage of the next EEG segment is most probable the same as the current EEG segment. In term infants, where four distinct sleep stages can be identified, specific sleep-state transitions are physiologically more likely. A typical sleep cycle of a term neonate consists of the following transitions: AS1/M $\rightarrow$HVS $\rightarrow$TA $\rightarrow$AS2/LVI $\rightarrow$wake [56]. Pillay et al. adopted hidden markov models to incorporate these transition probabilities in a four-state sleep classifier for term infants [159].

Two of the proposed algorithms accounted for these temporal patterns by adopting a moving average filter to smooth the obtained sleep state probabilities. Although this simple postprocessing step reduced the number of misclassifications, more advanced strategies could be used to further improve the results. In the case of deep learning approaches, this could be accomplished by using recurrent neural networks, especially long-short term memory (LSTM) seem promising to incorporate these temporal dependencies in the network [156]. For the tensor-based approach, we have explored the usage of an additional regularization term to the objective function of the CPD. In this way, the smoothing operation is imposed during the tensor decomposition. However, the resulting smoothed temporal signature was very similar to the one obtained using a simple postprocessing step, while the regularization parameter was more

difficult to tune. Therefore, further research to other approaches fully exploiting the temporal dependencies in tensor decompositions should be carried out.

Other characteristics of the EEG patterns

The algorithms presented in Chapter 4 and 5 are solely based on the sleep state-dependent difference in EEG complexity to identify neonatal sleep stages. However, previous research has identified a wide range of EEG features with high discriminative power (e.g. spectral features, features related to EEG discontinuity, nonlinear features, etc.) [103, 162, 169]. Therefore, a logical extension of the fixed-size LS-SVM classifier is to combine all sleep state-related features, perform feature selection and use the selected, most discriminative features to develop the final classifier. In case of the tensor-based sleep stage classifier, the temporal signature of interest can be combined with other features to perform either clustering or to train a supervised classifier. Moreover, other tensorizations of the EEG time series can be explored to retrieve information about the sleep cycling. In the next paragraph we will present preliminary results on the use of tensor decompositions to extract sleep stage information from the brain connectivity of EEG signals.

Brain connectivity for sleep staging Recent research has suggested that the functional brain connectivity is fundamentally different depending on the sleep state [191, 192]. As the interactions between different brain regions can naturally be represented in a tensor, we have explored decompositions of brain connectivity tensors to obtain information about the neonate's sleep cycling. First of all, the EEG time series is bandpass filtered in order to remove the most severe artifacts. Each filtered 8-channel EEG recording is then segmented into windows of 30 seconds. For each of these windows the coherency between every possible channel combination is computed. By using the imaginary part of the coherency, we avoid false interactions due to volume conduction which is a common problem in brain connectivity studies [143]. Finally, each EEG recording can be tensorized into a fourth-order tensor with modes: channels, channels, frequency and segments. So each frontal slice will reflect how strongly each EEG channel is interacting with any other channel (in graph theory this is referred to as the adjacency matrix) at a specific frequency in a specific time segment. As the imaginary part of the coherency is a measure of functional brain connectivity and is therefore undirected, each frontal slice will be symmetric. This prior information can be incorporated during the tensor decomposition. Similarly to the approach proposed in Chapter 5, the nonnegative rank-1 CPD of the brain connectivity tensor is computed. Assuming that the brain connectivity is significantly different in quiet sleep compared to non-quiet sleep, we expect

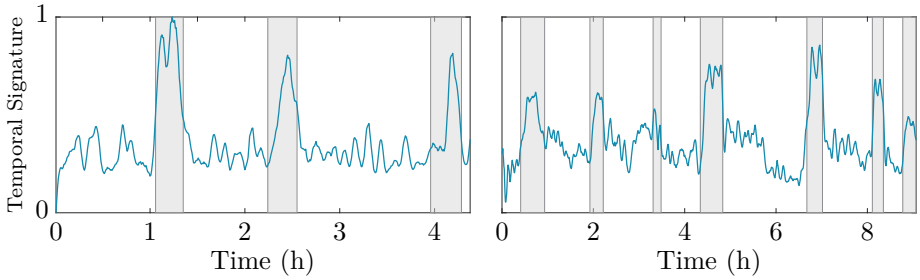


Figure 11.1: Example of two smoothed temporal signatures obtained by the rank-1 CPD of the brain connectivity tensor of two EEG recordings (left: PMA = 32 weeks, right: PMA = 39 weeks). The temporal signatures are clearly increased during quiet sleep periods highlighted in light grey.

that the factor vector in the fourth mode (temporal signature) will provide information about the neonate's sleep cycles. The smoothed temporal signature of two EEG recordings are shown in blue, while the clinically labelled quiet sleep periods are highlighted in light grey in Figure 11.1. From these plots it can be seen that the temporal signature is clearly increased during quiet sleep. The left graph is obtained from the EEG recording measured at 32 weeks, while the right plot corresponds to a measurement at 39 weeks PMA. This illustrates that the brain connectivity can also be a useful feature for sleep stage classification at different postmenstrual ages.

These preliminary results seem promising to perform neonatal sleep staging, however in some EEG recordings the obtained temporal signature was not related to the sleep cycling. This is illustrated in the example in Figure 11.2, where the smoothed temporal signature obtained from the rank-1 CPD of a EEG recording at 36 weeks (left) and 40 weeks (right) PMA is represented. Consequently, further studies on this novel approach are recommended. As in the proposed method for sleep stage identification based on the CPD of the multiscale entropy tensor, the main challenge in this approach will be reliably estimating the number of components. So far, we have only considered a rank-1 CPD of the brain connectivity tensor. However, in future work higher ranks and other tensor decompositions, allowing more flexibility, should be considered.

Other modalities

In this thesis, we focused on the use of EEG in order to classify the sleep stages in preterm infants. However, when a newborn is admitted to the NICU it is

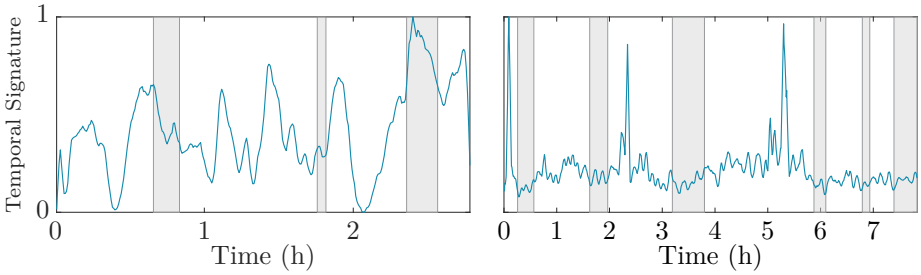


Figure 11.2: Example of two smoothed temporal signatures obtained by the rank-1 CPD of the brain connectivity tensor of two EEG recordings (left: PMA = 36 weeks, right: PMA = 40 weeks). The temporal signatures do not show a clear correlation with the clinically labelled quiet sleep periods indicated in light grey.

common to also monitor other vital parameters, such as ECG, respiration, chin muscle EMG activity in addition to the EEG. During visual labelling of the sleep-wake cycling, these other modalities are also considered by the clinician. As has been mentioned before, quiet sleep is characterized by more regular cardiorespiratory patterns and less body movements [56]. As a consequence, in future studies these additional modalities can be incorporated in the sleep stage classifiers to further improve the performance. In the next paragraph, we will briefly demonstrate that the respiration signal can provide information about the sleep cycling in preterm neonates.

Regularity of respiration for sleep staging Based on a proposal of W. Deburchgraeve [52], we have investigated whether the regularity of the respiration is a discriminative feature to distinguish quiet sleep from non-quiet sleep in preterm infants. The regularity of the respiration signal was assessed in sliding windows of 30 s by computing the area under its normalized autocorrelation. The regularity of each of the 97 EEG recordings measured between 27 and 42 weeks was assessed and compared to clinically labelled quiet sleep periods. As stated above, the regularity of the respiration is expected to be higher during quiet sleep compared to non-quiet sleep segments. Figure 11.3 shows the regularity of the respiration of two EEG recordings in blue, while the quiet sleep periods are highlighted in light grey. In the left graph, the quiet sleep periods are characterized by an increase of the respiration regularity. However, this expected pattern is not observed in the EEG recording visualized on the right. To evaluate how well the regularity of the respiration is related to the sleep stages across the complete dataset, ROC curves are constructed based on the

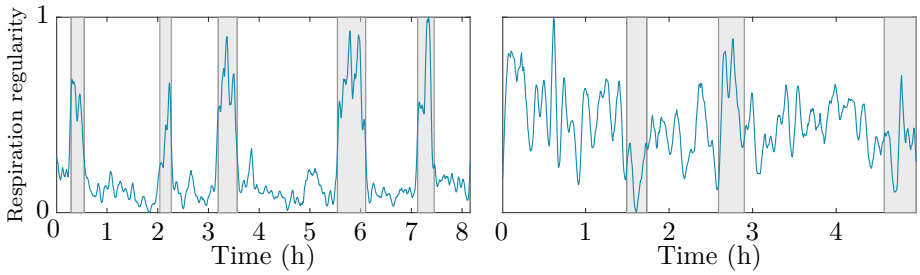


Figure 11.3: The regularity of the respiration of two EEG recordings. The clinically labelled quiet sleep periods are highlighted in light grey. In the left graph, the quiet sleep periods are characterized by an increased regularity of the respiration. In the EEG recording shown on the right, no clear trend in the respiration regularity with regard to the sleep staging can be observed.

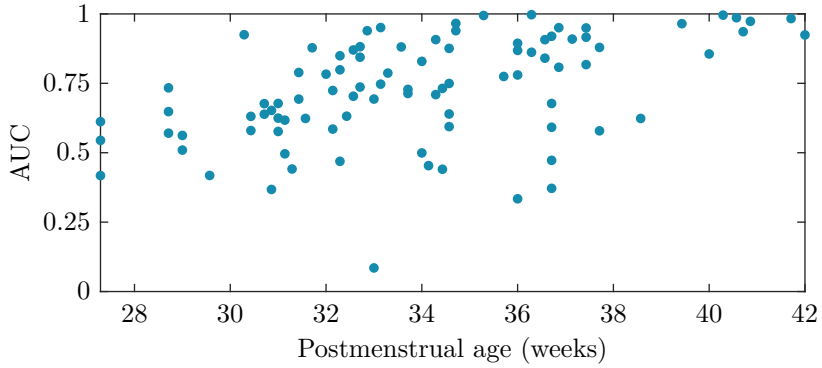


Figure 11.4: The area under the ROC curves constructed based on the regularity of the respiration and the clinical sleep labels as a function of PMA. A nearly perfect sleep state identification is observed near term age, while the respiration regularity is less discriminating in the very young neonates.

respiration regularity and the annotated quiet sleep segments. The AUC of each of the 97 recordings is represented in function of PMA in Figure 11.4. It can be seen from Figure 11.4 that a nearly perfect AUC is obtained for EEG signals recorded near term age, while the performance is significantly lower in very young neonates. It is likely that these younger neonates received respiratory support which diminishes the difference in respiration regularity between the sleep stages. Therefore, we can conclude that this modality can mainly add to the classification performance in older neonates without ventilation.

11.2.2 Automated brain maturation quantification

The second problem dealt with in this thesis, is the automated quantification of brain maturation in preterm infants. EEG complexity features are proposed as a biomarker of brain maturation in preterm infants. Based on these complexity features, a brain-age model which can accurately estimate the postmenstrual age of the neonate has been developed. Similarly to the sleep stage algorithms, a natural progression of this work is to include other state-of-the-art maturational EEG features in this regression model. The most important limitation of this research lies in the fact that the proposed model was only evaluated on neonates with good neurodevelopmental outcome.
The studies presented in Chapter 8 and Chapter 9 investigated a set of maturational EEG features (including the above-mentioned complexity features) in relation to structural brain development and thyroid hormone function during the first week of life, respectively. The main weakness of these studies is the quality of the EEG recordings and the heterogeneity of the dataset.
In the next sections, suggestions for further research will be given starting from these limitations.

Dataset

One of the main struggles in biomedical data processing, especially in automated processing of EEG signals, are data-related. EEG is characterized by a poor signal-to-noise ratio and is known to be sensitive to artifacts. As a result, preprocessing of the data is an important step in the algorithmic pipeline. Especially in Chapter 9 and 10 EEG recordings were of poor quality and automated or visual preselection of data was deemed necessary. However, this preselection of clean EEG segments leads to a drastic reduction of the sample size. Therefore, automated methods to remove the source of the artifacts without loss of data (e.g. independent component analysis) should be considered in further work. Another major issue encountered in these studies is the heterogeneity of the dataset. In Chapter 9 two EEG monitors with different filter settings were used, some of the patients received morphine, the neonates included in the study were not all born at the same gestational age and there is variation in the postmenstrual age at which the EEGs and MRIs were recorded. Although we have tried to correct for the most important confounding variables, the variability within the dataset hampers the analysis. In Chapter 10 similar problems arose as the patients with low ΔfT4 levels were born at a lower gestational age and their EEG was recorded at a higher PMA. Moreover, levothyroxine was administered to a subset of the patients and some infants suffered from IVH. Therefore,

prospective studies with more homogeneous datasets should be carried out to confirm the findings reported in these studies.

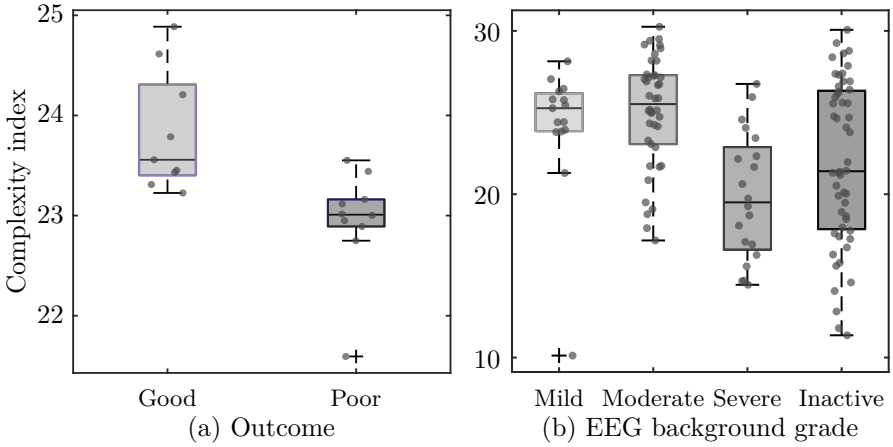
Automated prognosis of neurodevelopmental outcome

The idea behind the brain-age model based on multiscale entropy features is that a large deviation between the estimated and the real postmenstrual age of the neonate might indicate abnormal brain development. However, as the database of serial EEG recordings only contained patients with good neurodevelopmental outcome, we could not confirm that neonates with poor outcome are characterized by a different maturational trajectory. Hence, this is an interesting research line for future studies. An exploratory analysis performed in neonates with HIE will be described below.

Complexity index to assess severity of EEG background and outcome in neonates with HIE In order to explore whether the complexity of the neonatal EEG is reduced in patients with poor outcome, we have computed the multiscale entropy on EEG recordings of 19 neonates with hypoxic ischaemic encephalopathy (HIE) (GA: 36 – 41 weeks) [55]. The multichannel EEG recordings were measured during therapeutic hypothermia and rewarming, and the background activity of each epoch of six hours of EEG was labelled from grade 0 (normal EEG) to 4 (inactive EEG) according to Murphy et al. [138]. Neurodevelopmental outcome was assessed based on the Bayley Scales of Infant Development cognitive and motor score at 12 and 24 months. Ten patients had poor outcome and nine of them died following redirection of care. The remaining nine patients had good neurodevelopmental outcome assessed at 24 months of age.

The complexity index was computed for each EEG recording in windows of 100 s. The average complexity index for each patient is then computed by taking the mean across all EEG segments and channels. The boxplot in Figure 11.5a compares this average complexity index between the neonates with favourable outcome and adverse outcome. Even though it is a very small sample size, a clear difference between the two patient groups can be observed. As expected, the EEG activity of the patients with good outcome is more complex compared to infants with poor outcome.

To establish whether the EEG complexity is different depending on the grade of background abnormality, the average complexity index is also computed for each 6 h epoch of EEG. The average complexity index for each of the EEG grades is compared in the boxplots of Figure 11.5b. The complexity index is clearly higher for EEG windows labelled as mild or moderate compared to the



(a) Outcome (b) EEG background grade

Figure 11.5: (a) The boxplot shows the average complexity index among the complete recording length and all channels for HIE neonates with good and poor outcome. It is clear that the complexity index is greater in patient with good neurodevelopmental outcome. (b) The boxplots illustrate the difference in complexity index across different severity grades of background activity [138]. EEG segments with mild and moderate abnormalities have a higher complexity index averaged among the channels compared to segments with severe abnormalities or inactive EEG.

segments with severe abnormalities or even inactive EEG. Note that none of the EEG segments were scored as normal.

These findings, while preliminary, suggest that the proposed complexity features can also be useful in the automated prognosis of the neurodevelopmental outcome of neonates with HIE. Moreover, based on the observations in Figure 11.5b, it is probable that the complexity index can be used to track the evolution of the EEG background activity in HIE patients. However, discrimination between mild and moderate abnormalities or severe abnormalities and inactive EEG seems challenging.

The results of this small study support the idea that the complexity features can also be used to track recovery or deterioration of a neonate's brain activity and detect patients at risk of adverse neurodevelopmental outcome. Therefore, future studies evaluating the proposed brain-age model using complexity features on serial EEG recordings of patients with poor neurodevelopmental outcome are recommended.

Appendix A

Performance of algorithms for automated EEG sleep staging in preterm infants

This chapter gives a detailed overview of the performance of each proposed algorithm for automated sleep staging and two algorithms described in the literature. Tables A.1 – A.6 list the performance for each EEG recording of the test set separately.

Table A.1: Performance of the complexity feature-based algorithm on each of the test recordings.

Patient	**GA** (wks;days)	**PMA** (wks;days)	**Acc**	**Sens**	**Spec**	**Kappa**	**AUC**
19	30;0	32;0	0.72	0.98	0.58	0.47	0.9
19	30;0	34;4	0.91	0.87	0.92	0.74	0.97
22	25;3	28;5	0.7	0.38	0.79	0.16	0.59
22	25;3	34;2	0.85	0.83	0.85	0.53	0.92
22	25;3	37;5	0.91	0.59	0.96	0.57	0.91
35	30;3	31;1	0.73	0.96	0.61	0.49	0.92
35	30;3	34;1	0.85	0.81	0.87	0.63	0.92
35	30;3	35;5	0.93	0.78	0.97	0.77	0.98
35	30;3	36;4	0.93	0.73	0.99	0.77	0.96
40	30;2	31;3	0.89	0.88	0.9	0.67	0.93
40	30;2	36;0	0.91	0.74	0.96	0.74	0.96
56	30;4	30;6	0.43	0.44	0.43	-0.11	0.45
56	30;4	33;1	0.92	0.87	0.94	0.8	0.96
56	30;4	36;6	0.88	0.55	0.96	0.56	0.94
57	30;4	30;6	0.74	0.84	0.71	0.45	0.85
57	30;4	33;1	0.8	0.94	0.75	0.57	0.96
57	30;4	36;6	0.8	0.37	0.91	0.31	0.82
60	31;6	32;2	0.85	0.9	0.82	0.69	0.93
60	31;6	34;0	0.87	0.7	0.93	0.64	0.92
60	31;6	36;2	0.93	0.93	0.93	0.8	0.98
72	30;0	30;5	0.64	0.72	0.58	0.29	0.68
72	30;0	32;2	0.88	0.46	0.98	0.54	0.85
72	30;0	34;0	0.87	0.05	1	0.08	0.73
72	30;0	36;0	0.84	0.22	1	0.31	0.93
72	30;0	37;3	0.84	0.23	0.98	0.29	0.84
75	28;6	31;2	0.85	0.81	0.86	0.66	0.89
75	28;6	33;2	0.93	0.97	0.91	0.81	0.99
75	28;6	36;2	0.83	0.57	0.9	0.48	0.82
78	29;6	30;2	0.75	0.67	0.8	0.47	0.82
78	29;6	32;5	0.74	0.86	0.68	0.47	0.86
78	29;6	34;4	0.89	0.55	0.99	0.63	0.96
78	29;6	36;5	0.84	0.38	0.95	0.39	0.84
106	26;2	31;5	0.88	0.72	0.93	0.68	0.93
106	26;2	33;4	0.89	0.69	0.97	0.71	0.97
106	26;2	36;5	0.94	0.89	0.95	0.78	0.97
159	29;4	32;4	0.86	0.7	0.95	0.68	0.92
159	29;4	34;3	0.8	0.73	0.81	0.33	0.88
159	29;4	36;4	0.78	0.38	0.94	0.38	0.83
159	29;4	37;3	0.89	0.62	0.96	0.64	0.93
159	29;4	41;5	0.83	0.79	0.85	0.59	0.88
160	29;4	32;5	0.72	0.42	0.78	0.17	0.64
160	29;4	36;5	0.87	0.63	1	0.69	0.96
160	29;4	40;6	0.9	0.87	0.91	0.75	0.95

Table A.2: Performance of the tensor-based algorithm on each of the test recordings.

Patient	**GA** (wks;days)	**PMA** (wks;days)	**Acc**	**Sens**	**Spec**	**Kappa**	**AUC**
19	30;0	32;0	0.59	0.85	0.45	0.25	0.76
19	30;0	34;4	0.91	0.97	0.9	0.75	0.97
22	25;3	28;5	0.66	0.82	0.61	0.3	0.8
22	25;3	34;2	0.83	0.89	0.83	0.51	0.91
22	25;3	37;5	0.78	1	0.75	0.42	0.97
35	30;3	31;1	0.83	0.81	0.85	0.64	0.94
35	30;3	34;1	0.89	0.97	0.86	0.72	0.99
35	30;3	35;5	0.93	0.72	0.98	0.76	0.89
35	30;3	36;4	0.91	0.87	0.93	0.77	0.97
40	30;2	31;3	0.85	0.79	0.86	0.55	0.91
40	30;2	36;0	0.83	0.81	0.83	0.58	0.9
56	30;4	30;6	0.62	0.81	0.53	0.28	0.67
56	30;4	33;1	0.94	0.87	0.97	0.86	0.99
56	30;4	36;6	0.41	0.02	0.49	-0.33	0.22
57	30;4	30;6	0.75	0.58	0.81	0.38	0.79
57	30;4	33;1	0.94	0.9	0.96	0.85	0.98
57	30;4	36;6	0.6	0.25	0.68	-0.05	0.58
60	31;6	32;2	0.82	0.84	0.81	0.63	0.91
60	31;6	34;0	0.87	0.61	0.97	0.64	0.94
60	31;6	36;2	0.56	0.81	0.5	0.18	0.79
72	30;0	30;5	0.73	0.43	0.99	0.44	0.81
72	30;0	32;2	0.81	0.77	0.81	0.48	0.91
72	30;0	34;0	0.56	1	0.5	0.21	0.84
72	30;0	36;0	0.9	0.67	0.96	0.67	0.93
72	30;0	37;3	0.47	1	0.34	0.17	0.92
75	28;6	31;2	0.9	0.95	0.87	0.77	0.98
75	28;6	33;2	0.97	0.92	0.98	0.91	0.99
75	28;6	36;2	0.58	0.3	0.66	-0.04	0.57
78	29;6	30;2	0.83	0.89	0.79	0.65	0.88
78	29;6	32;5	0.88	0.79	0.92	0.72	0.92
78	29;6	34;4	0.93	0.73	0.99	0.78	0.95
78	29;6	36;5	0.75	0.57	0.79	0.31	0.72
106	26;2	31;5	0.93	0.8	0.98	0.81	0.94
106	26;2	33;4	0.9	0.75	0.96	0.74	0.96
106	26;2	36;5	0.94	0.86	0.95	0.77	0.96
159	29;4	32;4	0.93	0.83	0.99	0.85	0.99
159	29;4	34;3	0.84	0.87	0.84	0.45	0.94
159	29;4	36;4	0.69	0	0.97	-0.04	0.74
159	29;4	37;3	0.87	0.83	0.88	0.64	0.9
159	29;4	41;5	0.73	1	0.64	0.48	0.97
160	29;4	32;5	0.75	0.3	0.84	0.14	0.78
160	29;4	36;5	0.8	0.74	0.83	0.56	0.83
160	29;4	40;6	0.68	1	0.57	0.41	0.95

Table A.3: Performance of the CNN sleep stage classifier on each of the test recordings.

Patient	**GA** (wks;days)	**PMA** (wks;days)	**Acc**	**Sens**	**Spec**	**Kappa**	**AUC**
19	30;0	32;0	0.85	0.98	0.79	0.7	0.93
19	30;0	34;4	0.91	0.97	0.89	0.74	0.98
22	25;3	28;5	0.77	0.34	0.88	0.24	0.64
22	25;3	34;2	0.94	0.91	0.94	0.77	0.99
22	25;3	37;5	0.89	0.28	0.97	0.31	0.79
35	30;3	31;1	0.89	1	0.83	0.76	0.98
35	30;3	34;1	0.89	0.76	0.93	0.69	0.95
35	30;3	35;5	0.96	0.88	0.98	0.88	0.99
35	30;3	36;4	0.96	0.83	1	0.88	0.99
40	30;2	31;3	0.97	0.82	1	0.88	0.99
40	30;2	36;0	0.92	0.69	0.99	0.75	0.96
56	30;4	30;6	0.54	0.24	0.68	-0.08	0.36
56	30;4	33;1	0.95	0.9	0.97	0.88	0.99
56	30;4	36;6	0.88	0.4	0.99	0.5	0.96
57	30;4	30;6	0.88	0.78	0.91	0.68	0.95
57	30;4	33;1	0.87	0.99	0.82	0.7	0.99
57	30;4	36;6	0.9	0.54	0.99	0.64	0.95
60	31;6	32;2	0.89	0.96	0.84	0.77	0.98
60	31;6	34;0	0.86	0.97	0.82	0.69	0.98
60	31;6	36;2	0.92	0.9	0.92	0.75	0.98
72	30;0	30;5	0.62	0.83	0.44	0.27	0.74
72	30;0	32;2	0.58	0.9	0.51	0.22	0.87
72	30;0	34;0	0.81	0.5	0.86	0.3	0.81
72	30;0	36;0	0.92	0.7	0.98	0.74	0.9
72	30;0	37;3	0.92	0.6	0.99	0.69	0.91
75	28;6	31;2	0.79	1	0.69	0.59	0.98
75	28;6	33;2	0.92	0.97	0.9	0.79	0.99
75	28;6	36;2	0.89	0.98	0.86	0.72	0.98
78	29;6	30;2	0.9	0.99	0.85	0.8	0.96
78	29;6	32;5	0.89	0.68	0.99	0.73	0.96
78	29;6	34;4	0.93	0.96	0.91	0.81	0.99
78	29;6	36;5	0.86	0.27	1	0.37	0.93
106	26;2	31;5	0.93	0.92	0.94	0.83	0.98
106	26;2	33;4	0.94	1	0.92	0.86	0.99
106	26;2	36;5	0.95	0.88	0.97	0.83	0.98
159	29;4	32;4	0.96	0.93	0.97	0.9	0.99
159	29;4	34;3	0.98	0.87	0.99	0.89	1
159	29;4	36;4	0.96	0.9	0.98	0.9	0.99
159	29;4	37;3	0.94	0.77	0.98	0.8	0.98
159	29;4	41;5	0.92	0.73	0.98	0.76	0.93
160	29;4	32;5	0.87	0.81	0.88	0.59	0.87
160	29;4	36;5	0.95	0.91	0.97	0.89	0.99
160	29;4	40;6	0.86	1	0.81	0.69	0.96

Table A.4: Performance of the cluster-based adaptive sleep staging (CLASS) algorithm on each of the test recordings.

Patient	**GA** (wks;days)	**PMA** (wks;days)	**Acc**	**Sens**	**Spec**	**Kappa**	**AUC**
19	30;0	32;0	0.81	0.58	0.93	0.54	0.84
19	30;0	34;4	0.89	0.83	0.91	0.69	0.96
22	25;3	28;5	0.72	0.18	0.86	0.05	0.49
22	25;3	34;2	0.95	0.98	0.94	0.8	0.99
22	25;3	37;5	0.93	0.82	0.94	0.68	0.97
35	30;3	31;1	0.87	0.65	0.98	0.69	0.97
35	30;3	34;1	0.91	0.71	0.97	0.73	0.95
35	30;3	35;5	0.97	0.85	1	0.89	0.98
35	30;3	36;4	0.95	0.8	0.99	0.85	0.99
40	30;2	31;3	0.9	0.84	0.91	0.68	0.97
40	30;2	36;0	0.93	0.73	0.99	0.78	0.99
56	30;4	30;6	0.54	0.12	0.73	-0.16	0.3
56	30;4	33;1	0.91	0.72	0.98	0.76	0.98
56	30;4	36;6	0.94	0.77	0.98	0.79	0.97
57	30;4	30;6	0.86	0.59	0.95	0.59	0.94
57	30;4	33;1	0.92	0.78	0.97	0.79	0.98
57	30;4	36;6	0.93	0.67	0.99	0.75	0.98
60	31;6	32;2	0.73	0.39	0.95	0.37	0.76
60	31;6	34;0	0.89	0.62	0.99	0.69	0.98
60	31;6	36;2	0.94	0.79	0.97	0.8	0.97
72	30;0	30;5	0.72	0.42	0.97	0.41	0.85
72	30;0	32;2	0.81	0.58	0.86	0.42	0.87
72	30;0	34;0	0.87	0.92	0.86	0.59	0.94
72	30;0	36;0	0.94	0.72	0.99	0.79	0.99
72	30;0	37;3	0.9	0.74	0.94	0.67	0.95
75	28;6	31;2	0.91	0.72	1	0.78	0.98
75	28;6	33;2	0.98	0.92	1	0.94	1
75	28;6	36;2	0.75	0.3	0.88	0.2	0.79
78	29;6	30;2	0.62	0.24	0.83	0.08	0.61
78	29;6	32;5	0.91	0.73	0.99	0.78	0.94
78	29;6	34;4	0.96	0.85	1	0.89	0.99
78	29;6	36;5	0.96	0.93	0.97	0.88	1
106	26;2	31;5	0.93	0.78	0.98	0.81	0.99
106	26;2	33;4	0.9	0.7	0.98	0.73	0.99
106	26;2	36;5	0.97	0.95	0.98	0.9	1
159	29;4	32;4	0.9	0.71	0.99	0.75	0.99
159	29;4	34;3	0.96	0.89	0.97	0.8	1
159	29;4	36;4	0.82	0.62	0.9	0.54	0.94
159	29;4	37;3	0.95	0.82	0.98	0.84	0.99
159	29;4	41;5	0.92	0.73	0.98	0.76	0.96
160	29;4	32;5	0.9	0.71	0.93	0.64	0.94
160	29;4	36;5	0.89	0.7	1	0.75	1
160	29;4	40;6	0.87	0.65	0.96	0.65	0.96

Table A.5: Performance of the spectral feature-based algorithm without postprocessing on each of the test recordings.

Patient	**GA** (wks;days)	**PMA** (wks;days)	**Acc**	**Sens**	**Spec**	**Kappa**	**AUC**
19	30;0	32;0	0.77	0.82	0.75	0.53	0.86
19	30;0	34;4	0.86	0.95	0.84	0.64	0.95
22	25;3	28;5	0.76	0.3	0.88	0.2	0.6
22	25;3	34;2	0.9	0.71	0.93	0.61	0.93
22	25;3	37;5	0.85	0.63	0.88	0.42	0.86
35	30;3	31;1	0.68	0.85	0.6	0.39	0.82
35	30;3	34;1	0.71	0.85	0.66	0.39	0.82
35	30;3	35;5	0.86	0.89	0.86	0.65	0.95
35	30;3	36;4	0.92	0.92	0.92	0.78	0.97
40	30;2	31;3	0.85	0.76	0.87	0.54	0.86
40	30;2	36;0	0.92	0.78	0.96	0.77	0.94
56	30;4	30;6	0.52	0.45	0.54	0	0.5
56	30;4	33;1	0.88	0.7	0.96	0.69	0.89
56	30;4	36;6	0.86	0.84	0.87	0.61	0.92
57	30;4	30;6	0.7	0.71	0.7	0.34	0.74
57	30;4	33;1	0.78	0.81	0.77	0.5	0.86
57	30;4	36;6	0.86	0.84	0.86	0.61	0.92
60	31;6	32;2	0.71	0.43	0.88	0.33	0.77
60	31;6	34;0	0.85	0.71	0.9	0.61	0.88
60	31;6	36;2	0.92	0.87	0.93	0.76	0.96
72	30;0	30;5	0.67	0.57	0.75	0.32	0.72
72	30;0	32;2	0.73	0.51	0.78	0.24	0.73
72	30;0	34;0	0.87	0.61	0.91	0.49	0.83
72	30;0	36;0	0.87	0.57	0.95	0.57	0.89
72	30;0	37;3	0.9	0.66	0.95	0.64	0.9
75	28;6	31;2	0.81	0.68	0.87	0.55	0.86
75	28;6	33;2	0.89	0.87	0.89	0.71	0.93
75	28;6	36;2	0.84	0.88	0.83	0.61	0.92
78	29;6	30;2	0.66	0.84	0.55	0.35	0.74
78	29;6	32;5	0.66	0.87	0.57	0.35	0.83
78	29;6	34;4	0.85	0.64	0.92	0.57	0.84
78	29;6	36;5	0.88	0.75	0.91	0.62	0.91
106	26;2	31;5	0.83	0.72	0.86	0.55	0.85
106	26;2	33;4	0.81	0.45	0.95	0.46	0.84
106	26;2	36;5	0.89	0.67	0.93	0.59	0.9
159	29;4	32;4	0.82	0.74	0.86	0.6	0.88
159	29;4	34;3	0.93	0.74	0.95	0.66	0.94
159	29;4	36;4	0.85	0.73	0.9	0.64	0.89
159	29;4	37;3	0.88	0.86	0.88	0.67	0.94
159	29;4	41;5	0.83	0.95	0.79	0.63	0.96
160	29;4	32;5	0.8	0.54	0.85	0.35	0.77
160	29;4	36;5	0.82	0.79	0.83	0.61	0.9
160	29;4	40;6	0.84	0.97	0.79	0.65	0.96

Table A.6: Performance of the spectral feature-based algorithm with postprocessing on each of the test recordings.

Patient	**GA** (wks;days)	**PMA** (wks;days)	**Acc**	**Sens**	**Spec**	**Kappa**	**AUC**
19	30;0	32;0	0.88	0.91	0.87	0.74	0.95
19	30;0	34;4	0.92	0.99	0.9	0.77	0.99
22	25;3	28;5	0.79	0.3	0.92	0.25	0.63
22	25;3	34;2	0.93	0.78	0.96	0.72	0.98
22	25;3	37;5	0.93	0.65	0.97	0.64	0.94
35	30;3	31;1	0.8	0.99	0.71	0.62	0.95
35	30;3	34;1	0.8	0.92	0.76	0.55	0.95
35	30;3	35;5	0.95	0.95	0.95	0.85	0.99
35	30;3	36;4	0.98	0.97	0.98	0.93	0.99
40	30;2	31;3	0.93	0.78	0.96	0.75	0.97
40	30;2	36;0	0.96	0.87	0.99	0.89	0.98
56	30;4	30;6	0.54	0.29	0.64	-0.06	0.48
56	30;4	33;1	0.92	0.76	0.99	0.81	0.97
56	30;4	36;6	0.96	0.92	0.97	0.87	0.99
57	30;4	30;6	0.85	0.83	0.86	0.63	0.9
57	30;4	33;1	0.87	0.86	0.87	0.68	0.96
57	30;4	36;6	0.96	0.88	0.98	0.86	0.99
60	31;6	32;2	0.75	0.37	0.99	0.41	0.93
60	31;6	34;0	0.93	0.78	0.98	0.81	0.97
60	31;6	36;2	0.98	0.94	0.99	0.94	1
72	30;0	30;5	0.73	0.49	0.93	0.43	0.87
72	30;0	32;2	0.9	0.55	0.98	0.62	0.9
72	30;0	34;0	0.93	0.61	0.98	0.66	0.92
72	30;0	36;0	0.9	0.51	1	0.62	0.99
72	30;0	37;3	0.92	0.66	0.99	0.72	0.98
75	28;6	31;2	0.92	0.8	0.99	0.82	0.98
75	28;6	33;2	0.97	0.88	0.99	0.9	0.98
75	28;6	36;2	0.93	0.96	0.92	0.81	0.99
78	29;6	30;2	0.65	0.98	0.47	0.38	0.91
78	29;6	32;5	0.75	0.97	0.65	0.51	0.97
78	29;6	34;4	0.92	0.65	1	0.74	0.93
78	29;6	36;5	0.95	0.85	0.97	0.83	0.98
106	26;2	31;5	0.93	0.82	0.97	0.81	0.96
106	26;2	33;4	0.82	0.37	0.99	0.45	0.95
106	26;2	36;5	0.95	0.75	0.98	0.79	0.98
159	29;4	32;4	0.93	0.87	0.96	0.84	0.98
159	29;4	34;3	0.98	0.8	1	0.86	1
159	29;4	36;4	0.91	0.77	0.96	0.77	0.97
159	29;4	37;3	0.95	0.98	0.94	0.87	0.99
159	29;4	41;5	0.95	1	0.94	0.88	1
160	29;4	32;5	0.85	0.46	0.93	0.42	0.84
160	29;4	36;5	0.93	0.87	0.97	0.86	0.98
160	29;4	40;6	0.91	1	0.88	0.79	0.99

Bibliography

[1] Abdel-Hamid, O., Mohamed, A.-r., Jiang, H., and Penn, G. Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2012), IEEE, pp. 4277–4280.

[2] Acar, E., and Yener, B. Unsupervised Multiway Data Analysis: A Literature Survey. *IEEE Transactions on Knowledge and Data Engineering 21*, 1 (2009), 6–20.

[3] Acharya, U. R., Fujita, H., Oh, S. L., Hagiwara, Y., Tan, J. H., and Adam, M. Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Information Sciences 415-416* (2017), 190–198.

[4] Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., and Adeli, H. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Computers in Biology and Medicine 100* (2017), 270–278.

[5] Alexander, D. L., Tropsha, A., and Winkler, D. A. Beware of R2: Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models. *Journal of Chemical Information and Modeling 55*, 7 (2015), 1316–1322.

[6] Allen, K. A. Promoting and Protecting Infant Sleep. *Advances in neonatal care 12*, 5 (2012), 288–291.

[7] André, M., Lamblin, M.-D., D'Allest, A., Curzi-Dascalova, L., Moussalli-Salefranque, F., Nguyen The Tich, S., Vecchierini-Blineau, M.-F., Wallois, F., Walls-Esquivel, E., and Plouin, P. Electroencephalography in premature and full-term infants. Developmental features and glossary. *Neurophysiologie Clinique/Clinical Neurophysiology 40*, 2 (2010), 59–124.

[8] Ansari, A. H., Cherian, P. J., Caicedo, A., Naulaers, G., De Vos, M., and Van Huffel, S. Neonatal Seizure Detection Using Deep Convolutional Neural Networks. *International Journal of Neural Systems 29*, 4 (2019), 1850011.

[9] Ansari, A. H., De Wel, O., Lavanga, M., Caicedo, A., Dereymaeker, A., Jansen, K., Vervisch, J., De Vos, M., Naulaers, G., and Van Huffel, S. Quiet sleep detection in preterm infants using deep convolutional neural networks. *Journal of Neural Engineering 15*, 6 (2018), 066006.

[10] Ansari, A. H., De Wel, O., Pillay, K., Dereymaeker, A., Jansen, K., Van Huffel, S., Naulaers, G., and De Vos, M. A convolutional neural network outperforming state-of-the-art sleep staging algorithms for both preterm and term infants. *Accepted for publication in Journal of Neural Engineering* (2019).

[11] Atkinson, R., Migues, P. V., Cammarota, M., Kavanagh, J., Hunter, M., and Rostas, J. A. P. Biochemical, behavioural and electrophysiological investigations of brain maturation in chickens. *Brain Research Bulletin 76*, 3 (2008), 217–223.

[12] Barbeau, D. Y., and Weiss, M. D. Sleep Disturbances in Newborns. *Children 4*, 10 (2017), 90.

[13] Barlow, J. S. Computer characterization of tracé alternant and REM sleep patterns in the neonatal EEG by adaptive segmentation–an exploratory study. *Electroencephalography and clinical neurophysiology 60*, 2 (1985), 163–173.

[14] Battin, M., and Rutherford, M. A. Magnetic resonance imaging of the brain in preterm infants: 24 weeks' gestation to term. In *MRI of the neonatal brain*, M. A. Rutherford, Ed. W.B. Saunders Ltd., 2001, ch. 3.

[15] Batuwita, R., and Palade, V. Class Imbalance Learning Methods for Support Vector Machines. In *Imbalanced Learning: Foundations, Algorithms, Applications*, H. He and Y. Ma, Eds. Wiley-IEEE Press, 2013, ch. 6, pp. 101–150.

[16] Benders, M. J., Palmu, K., Menache, C., Borradori-Tolsa, C., Lazeyras, F., Sizonenko, S., Dubois, J., Vanhatalo, S., and Hüppi, P. S. Early brain activity relates to subsequent brain growth in premature infants. *Cerebral Cortex 25*, 9 (2014), 3014–3024.

[17] Benjamini, Y. Simultaneous and selective inference: Current successes and future challenges. *Biometrical journal. Biometrische Zeitschrift 52*, 6 (2010), 708–21.

[18] Bernal, J. Thyroid hormone receptors in brain development and function. *Nature Clinical Practice Endocrinology & Metabolism 3*, 3 (2007), 249–259.

[19] Bishop, C. M. Sparse Kernel Machines. In *Pattern Recognition and Machine Learning*, M. Jordan, J. Kleinberg, and B. Schölkopf, Eds. Springer, 2006, ch. 7, pp. 325–358.

[20] Blencowe, H., Cousens, S., Oestergaard, M. Z., Chou, D., Moller, A.-B., Narwal, R., Adler, A., Vera Garcia, C., Rohde, S., Say, L., and Lawn, J. E. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. *The Lancet 379*, 9832 (2012), 2162–2172.

[21] Britton, J. W., Frey, L. C., Hopp, J. L., Korb, P., Koubeissi, M. Z., Lievens, W. E., Pestana-Knight, E. M., and Louis, E. K. S. *Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants.* American Epilepsy Society, 2016.

[22] Bro, R. PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems 38*, 2 (1997), 149–171.

[23] Bro, R., and Kiers, H. A. L. A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics 17*, 5 (2003), 274–286.

[24] Burdjalov, V. F., Baumgart, S., and Spitzer, A. R. Cerebral function monitoring: a new scoring system for the evaluation of brain maturation in neonates. *Pediatrics 112*, 4 (2003), 855–861.

[25] Busa, M. A., and van Emmerik, R. E. Multiscale entropy: A tool for understanding the complexity of postural control. *Journal of Sport and Health Science 5*, 1 (2016), 44–51.

[26] Caicedo, A., Wel, O. D., Vandecappelle, M., Thewissen, L., Smits, A., Allegaert, K., Lathauwer, L. D., Naulaers, G., and Huffel, S. V. Monitoring of Brain Hemodynamics Coupling in Neonates using Updated Tensor Decompositions. In *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Berlin, Germany, 2019), IEEE, pp. 660–663.

[27] Calciolari, G., and Montirosso, R. The sleep protection in the preterm infants. *The Journal of Maternal-Fetal & Neonatal Medicine 24*, suppl. 1 (2011), 12–14.

[28] Carrozzi, M., Accardo, A., and Bouquet, F. Analysis of sleep-stage characteristics in full-term newborns by means of spectral and fractal parameters. *Sleep 27*, 7 (2004), 1384–1393.

[29] Chambon, S., Galtier, M., Arnal, P., Wainrib, G., and Gramfort, A. A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. *IEEE Transactions on Neural Systems and Rehabilitation Engineering 26*, 4 (2018), 758–769.

[30] Cherian, P. J., Swarte, R. M., and Visser, G. H. Technical standards for recording and interpretation of neonatal electroencephalogram in clinical practice. *Annals of Indian Academy of Neurology 12*, 1 (2009), 58–70.

[31] Cichocki, A., Mandic, D., De Lathauwer, L., Zhou, G., Zhao, Q., Caiafa, C., and Phan, H. A. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE Signal Processing Magazine 32*, 2 (2015), 145–163.

[32] Cichocki, A., Zdunek, R., Phan, A. H., and Amari, S. I. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation.* John Wiley and Sons, Ltd, 2009.

[33] Ciresan, D. C., Meier, U., Masci, J., Maria Gambardella, L., and Schmidhuber, J. Flexible, high performance convolutional neural networks for image classification. In *IJCAI Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence* (2011), vol. 2, Barcelona, Spain, pp. 1237–1242.

[34] Clouchoux, C., Guizard, N., Evans, A. C., du Plessis, A. J., and Limperopoulos, C. Normative fetal brain growth by quantitative in vivo magnetic resonance imaging. *American Journal of Obstetrics and Gynecology 206* (2012), 173.e1–173.e8.

[35] Cohen, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement 20*, 1 (1960), 37–46.

[36] Colvin, M., McGuire, W., and Fowlie, P. W. Neurodevelopmental outcomes after preterm birth. *BMJ 329*, 7479 (2004), 1390–1393.

[37] Cong, F., Lin, Q. H., Kuang, L. D., Gong, X. F., Astikainen, P., and Ristaniemi, T. Tensor decomposition of EEG signals: A brief review. *Journal of Neuroscience Methods 248* (2015), 59–69.

[38] Costa, M., Goldberger, A. L., and Peng, C.-K. Multiscale Entropy Analysis of Complex Physiologic Time Series. *Physical Review Letters 89*, 6 (2002), 068102.

[39] Costa, M., Goldberger, A. L., and Peng, C.-K. Multiscale entropy analysis of biological signals. *Physical Review E 71* (2005), 021906.

[40] Costa, M., Peng, C.-K., Goldberger, A. L., and Hausdorff, J. M. Multiscale entropy analysis of human gait dynamics. *Physica A 330* (2003), 53–60.

[41] Counsell, S. J. Magnetic resonance imaging of preterm brain injury. *Archives of Disease in Childhood - Fetal and Neonatal Edition 88*, 4 (2003), F269–F274.

[42] Covington, P., and Adams, J. Deep Neural Networks for YouTube Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, USA, 2016), pp. 191–198.

[43] De Brabanter, K., Karsmakers, P., Ojeda, F., Alzate, C., De Brabanter, J., Pelckmans, K., De Moor, B., Vandewalle, J., and Suykens, J. LS-SVM Toolbox User's Guide version 1.8. Tech. rep., ESAT-SISTA, KU Leuven, 2010.

[44] De Groot, L. J., Chrousos, G., Dungan, K., Feingold, K. R., Grossman, A., Hershman, J. M., Koch, C., Korbonits, M., McLachlan, R., New, M., and Others. Thyroid Hormones in Brain Development and Function–Endotext. In *Endotext*. MDText. com, Inc., 2000.

[45] De Lathauwer, L. A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization. *SIAM Journal on Matrix Analysis and Applications 28*, 3 (2006), 642–666.

[46] De Lathauwer, L. Decompositions of a Higher-Order Tensor in Block Terms—Part I: Lemmas for Partitioned Matrices. *SIAM Journal on Matrix Analysis and Applications 30*, 3 (2008), 1022–1032.

[47] De Lathauwer, L., De Moor, B., and Vandewalle, J. A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications 21*, 4 (2000), 1253–1278.

[48] De Vries, L. S., Heckmatt, J. Z., Burrin, J. M., Dubowitz, L. M., and Dubowitz, V. Low serum thyroxine concentrations and neural maturation in preterm infants. *Archives of Disease in Childhood 61*, 9 (1986), 862–866.

[49] De Vries, L. S., and Hellström-Westas, L. Role of cerebral function monitoring in the newborn. *Archives of Disease in Childhood: Fetal and Neonatal Edition 90*, 3 (2005), 201–207.

[50] De Wel, O., Lavanga, M., Caicedo, A., Jansen, K., Dereymaeker, A., Naulaers, G., and Van Huffel, S. Complexity Analysis of Neonatal EEG Using Multiscale Entropy : Applications in Brain Maturation and Sleep Stage Classification. *Entropy 19*, 10 (2017), 516.

[51] De Wel, O., Lavanga, M., Caicedo, A., Jansen, K., Naulaers, G., and Van Huffel, S. Decomposition of a Multiscale Entropy Tensor for Sleep Stage Identification in Preterm Infants. *Entropy 21*, 10 (2019), 936.

[52] Deburchgraeve, W. *Development of an automated neonatal EEG seizure monitor*. PhD thesis, KU Leuven, 2010.

[53] Deburchgraeve, W., Cherian, P. J., De Vos, M., Swarte, R. M., Blok, J. H., Visser, G. H., Govaert, P., and Van Huffel, S. Neonatal seizure localization using PARAFAC decomposition. *Clinical Neurophysiology 120*, 10 (2009), 1787–1796.

[54] Dereymaeker, A., Koolen, N., Jansen, K., Vervisch, J., Ortibus, E., De Vos, M., Van Huffel, S., and Naulaers, G. The suppression curve as a quantitative approach for measuring brain maturation in preterm infants. *Clinical Neurophysiology 127*, 8 (2016), 2760–2765.

[55] Dereymaeker, A., Matic, V., Vervisch, J., Cherian, P. J., Ansari, A. H., De Wel, O., Govaert, P., De Vos, M., Van Huffel, S., Naulaers, G., and Jansen, K. Automated EEG background analysis to identify neonates with hypoxic-ischemic encephalopathy treated with hypothermia at risk for adverse outcome: A pilot study. *Pediatrics and Neonatology 60*, 1 (2019), 50–58.

[56] Dereymaeker, A., Pillay, K., Vervisch, J., De Vos, M., Van Huffel, S., Jansen, K., and Naulaers, G. Review of sleep-EEG in preterm and term neonates. *Early Human Development 113* (2017), 87–103.

[57] Dereymaeker, A., Pillay, K., Vervisch, J., Van Huffel, S., Naulaers, G., Jansen, K., and De Vos, M. An Automated Quiet Sleep Detection Approach in Preterm Infants as a Gateway to Assess Brain Maturation. *International Journal of Neural Systems 27*, 6 (2017), 1750023.

[58] Dix, L. M. L., van Bel, F., and Lemmers, P. M. A. Monitoring Cerebral Oxygenation in Neonates: An Update. *Frontiers in pediatrics 5*, 46 (2017).

[59] Domanov, I., and De Lathauwer, L. On the Uniqueness of the Canonical Polyadic Decomposition of Third-Order Tensors—Part I: Basic results and uniquensess of one factor matrix and part II: Uniqueness of the overall decomposition. *SIAM Journal on Matrix Analysis and Applications 34*, 3 (2013), 855–903.

[60] Eckart, C., and Young, G. The approximation of one matrix by another of lower rank. *Psychometrika 1*, 3 (1936), 211–218.

[61] Eerdekens, A. *New Insights in Thyroid Function in Preterm Infants*. PhD thesis, KU Leuven, 2019.

[62] Eerdekens, A., Langouche, L., den Berghe, G. V., Verhaeghe, J., Naulaers, G., and Vanhole, C. Review shows that thyroid hormone substitution could benefit transient hypothyroxinaemia of prematurity but treatment strategies need to be clarified. *Acta Paediatrica 108*, 5 (2019), 792–805.

[63] Eerdekens, A., Naulaers, G., Ortibus, E., Verhaeghe, J., Langouche, L., and Vanhole, C. Evolution of circulating thyroid hormone levels in preterm infants during the first week of life: Perinatal influences and impact on neurodevelopment. *Journal of Pediatric Endocrinology and Metabolism 32*, 6 (2019), 597–606.

[64] Engle, W. Age Terminology During the Perinatal Period. *PEDIATRICS 114*, 5 (2004), 1362–1364.

[65] Espinoza, M., Suykens, J., and Moor, B. Fixed-size Least Squares Support Vector Machines: A Large Scale Application in Electrical Load Forecasting. *Computational Management Science 3*, 2 (2006), 113–129.

[66] Fawcett, T. An introduction to ROC analysis. *Pattern Recognition Letters 27*, 8 (2006), 861–874.

[67] Figueiredo, M. A. Adaptive sparseness for supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence 25*, 9 (2003), 1150–1159.

[68] Fisher, D. A. Thyroid System Immaturities in Very Low Birth Weight Premature Infants. *Seminars in Perinatology 32*, 6 (2008), 387–397.

[69] Foreman, S. W., Thomas, K. A., and Blackburn, S. T. Preterm Infant State Development: Individual and Gender Differences Matter. *Journal of obstetric, gynecologic, and neonatal nursing 37*, 6 (2008), 657–665.

[70] Forhead, A. J., and Fowden, A. L. Thyroid hormones in fetal growth and prepartum maturation. *Journal of Endocrinology 221*, 3 (2014), R87—-R103.

[71] Fraiwan, L., Lweesy, K., Khasawneh, N., Fraiwan, M., Wenz, H., and Dickhaus, H. Time frequency analysis for automated sleep stage identification in fullterm and preterm neonates. *Journal of Medical Systems 35*, 4 (2011), 693–702.

[72] Gibson, A., and Patterson, J. *Deep Learning*. O'Reilly Media, Inc., 2017.

[73] Goodfellow, I., Bengio, Y., and Courville, A. *Deep Learning*. MIT Press, 2016.

[74] Graven, S. Sleep and Brain Development. *Clinics in Perinatology 33*, 3 (2006), 693–706.

[75] Graven, S. N., and Browne, J. V. Sleep and Brain Development. The Critical Role of Sleep in Fetal and Early Neonatal Brain Development. *Newborn and Infant Nursing Reviews 8*, 4 (2008), 173–179.

[76] Gui, L., Lisowski, R., Faundez, T., Hüppi, P. S., Lazeyras, F., and Kocher, M. Morphology-driven automatic segmentation of MR images of the neonatal brain. *Medical image analysis 16*, 8 (2012), 1565–1579.

[77] Gupta, P., Sodhi, K. S., Saxena, A. K., Khandelwal, N., and Singhi, P. Neonatal cranial sonography: A concise review for clinicians. *Journal of Pediatric Neurosciences 11*, 1 (2016), 7–13.

[78] Hellström-Westas, L., Klette, H., Thorngren-Jerneck, K., and Rosén, I. Early prediction of outcome with aEEG in preterm infants with large intraventricular hemorrhages. *Neuropediatrics 32*, 6 (2001), 319–24.

[79] Hellström-Westas, L., and Rosén, I. Continuous brain-function monitoring: State of the art in clinical practice. *Seminars in Fetal and Neonatal Medicine 11*, 6 (2006), 503–511.

[80] Hellström-Westas, L., Rosén, I., de Vries, L. S., and Greisen, G. Amplitude-integrated EEG Classification and Interpretation in Preterm and Term Infants. *NeoReviews 7*, 2 (2006), e76–e87.

[81] Herculano-Houzel, S. The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in human neuroscience 3*, 31 (2009), 1–11.

[82] Houston, S. M., Herting, M. M., and Sowell, E. R. The neurobiology of childhood structural brain development: conception through adulthood. *Current topics in behavioral neurosciences 16* (2014), 3–17.

[83] Humeau-Heurtier, A. The Multiscale Entropy Algorithm and Its Variants: A Review. *Entropy 17*, 5 (2015), 3110–3123.

[84] Hüning, B., Storbeck, T., Bruns, N., Dransfeld, F., Hobrecht, J., Karpienski, J., Sirin, S., Schweiger, B., Weiss, C., Felderhoff-Müser, U., and Müller, H. Relationship between brain function (aEEG) and brain structure (MRI) and their predictive value for neurodevelopmental outcome of preterm infants. *European Journal of Pediatrics 177*, 8 (2018), 1181–1189.

[85] Hunyadi, B., Camps, D., Sorber, L., Paesschen, W. V., Vos, M. D., Huffel, S. V., and Lathauwer, L. D. Block term decomposition for modelling epileptic seizures. *EURASIP Journal on Advances in Signal Processing*, 139 (2014), 1–19.

[86] Hunyadi, B., Dupont, P., Van Paesschen, W., and Van Huffel, S. Tensor decompositions and data fusion in epileptic electroencephalography and functional magnetic resonance imaging data. *WIREs Data Mining and Knowledge Discovery 7*, e1197 (2017), 1–15.

[87] Hunyadi, B. H., Van Huffel, S. V. H., and De Vos, M. D. V. The power of tensor decompositions in biomedical applications. In *Machine Learning for Healthcare Technologies*. Institution of Engineering and Technology, 2016, ch. 5, pp. 83–109.

[88] Janjarasjitt, S., Scher, M. S., and Loparo, K. A. Nonlinear dynamical analysis of the neonatal EEG time series : The relationship between neurodevelopment and complexity. *Clinical Neurophysiology 119*, 4 (2008), 822–836.

[89] Janjarasjitt, S., Scher, M. S., and Loparo, K. A. Nonlinear dynamical analysis of the neonatal EEG time series : The relationship between sleep state and complexity. *Clinical Neurophysiology 119*, 8 (2008), 1812–1823.

[90] Jaume, J. C., Mendel, C. M., Frost, P. H., Greenspan, F. S., and Laughton, C. W. Extremely Low Doses of Heparin Release Lipase Activity into the Plasma and Can Thereby Cause Artifactual Elevations

in the Serum-Free Thyroxine Concentration as Measured by Equilibrium Dialysis. *Thyroid 6*, 2 (1996), 79–83.

[91] Jennekens, W., Niemarkt, H. J., Engels, M., Pasman, J. W., van Pul, C., and Andriessen, P. Topography of maturational changes in EEG burst spectral power of the preterm infant with a normal follow-up at 2 years of age. *Clinical Neurophysiology 123*, 11 (2012), 2130–2138.

[92] Kang, H. G., and Dingwell, J. B. Differential Changes with Age in Multiscale Entropy of Electromyography Signals from Leg Muscles during Treadmill Walking. *PLOS ONE 11*, 8 (2016), e0162034.

[93] Kersbergen, K. J., Makropoulos, A., Aljabar, P., Groenendaal, F., De Vries, L. S., Counsell, S. J., and Benders, M. J. N. L. Longitudinal Regional Brain Development and Clinical Risk Factors in Extremely Preterm Infants. *The Journal of Pediatrics 178* (2016), 93–100.e6.

[94] Kester, M. H. A., de Mena, R. M., Obregon, M. J., Marinkovic, D., Howatson, A., Visser, T. J., Hume, R., and de Escobar, G. M. Iodothyronine Levels in the Human Developing Brain: Major Regulatory Roles of Iodothyronine Deiodinases in Different Areas. *The Journal of Clinical Endocrinology & Metabolism 89*, 7 (2004), 3117–3128.

[95] Khedr, E. M., Toony, L. F. E., Tarkhan, M. N., and Abdella, G. Peripheral and Central Nervous System Alterations in Hypothyroidism: Electrophysiological Findings. *Neuropsychobiology 41*, 2 (2000), 88–94.

[96] Kidokoro, H., Inder, T., Okumura, A., and Watanabe, K. What does cyclicity on amplitude-integrated EEG mean. *Journal of Perinatology 32*, 8 (2012), 565–569.

[97] Kidokoro, H., Kubota, T., Hayashi, N., Hayakawa, M., Takemoto, K., Kato, Y., and Okumura, A. Absent cyclicity on aEEG within the first 24 h is associated with brain damage in preterm infants. *Neuropediatrics 41*, 6 (2010), 241–245.

[98] Kidokoro, H., Neil, J. J., and Inder, T. E. New MR imaging assessment tool to define brain abnormalities in very preterm infants at term. *American Journal of Neuroradiology 34*, 11 (2013), 2208–2214.

[99] Klebermass, K., Olischar, M., Waldhoer, T., Fuiko, R., Pollak, A., and Weninger, M. Amplitude-integrated EEG pattern predicts further outcome in preterm infants. *Pediatric research 70*, 1 (2011), 102–108.

[100] Kobayashi, H., Kakihana, W., and Kimura, T. Combined effects of age and gender on gait symmetry and regularity assessed by autocorrelation of trunk acceleration. *Journal of Neuroengineering and Rehabilitation 11*, 109 (2014).

[101] Kolda, T. G., and Bader, B. W. Tensor Decompositions and Applications. *SIAM Review 51*, 3 (2009), 455–500.

[102] Koolen, N., Jansen, K., Vervisch, J., Matic, V., De Vos, M., Naulaers, G., and Van Huffel, S. Line length as a robust method to detect high-activity events: automated burst detection in premature EEG recordings. *Clinical neurophysiology 125*, 10 (2014), 1985–1994.

[103] Koolen, N., Oberdorfer, L., Rona, Z., Giordano, V., Werther, T., Klebermass-Schrehof, K., Stevenson, N., and Vanhatalo, S. Automated classification of neonatal sleep states using EEG. *Clinical Neurophysiology 128*, 6 (2017), 1100–1108.

[104] Korotchikova, I., Stevenson, N. J., Livingstone, V., Ryan, C. A., and Boylan, G. B. Sleep-wake cycle of the healthy term newborn infant in the immediate postnatal period. *Clinical neurophysiology 127*, 4 (2016), 2095–2101.

[105] Krajča, V., Petránek, S., Mohylová, J., Paul, K., Gerla, V., and Lhotská, L. Neonatal EEG Sleep Stages Modelling by Temporal Profiles. In *Moreno Díaz R., Pichler F., Quesada Arencibia A. (eds) Computer Aided Systems Theory – EUROCAST 2007. EUROCAST 2007. Lecture Notes in Computer Science* (Berlin, 2007), vol. 4739, Springer, Berlin, Heidelberg, pp. 195–201.

[106] Krajča, V., Petránek, S., Paul, K., Matousek, M., Mohylová, J., and Lhotská, L. Automatic Detection of Sleep Stages in Neonatal EEG Using the Structural Time Profiles. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference* (Shanghai, China, 2005), IEEE, pp. 6014–6016.

[107] Krijnen, W. P., Dijkstra, T. K., and Stegeman, A. On the non-existence of optimal solutions and the occurrence of "degeneracy" in the CANDECOMP/PARAFAC model. *Psychometrika 73*, 3 (2008), 431–439.

[108] Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (Lake Tahoe, Nevada, 2012), pp. 1097–1105.

[109] Laoprasert, P. Artifacts. In *Atlas of pediatric eeg*. The McGraw-Hill Companies, Inc., 2011, ch. 2, pp. 125–200.

[110] Laudet, V. The Origins and Evolution of Vertebrate Metamorphosis. *Current Biology 21*, 18 (2011), R726–R737.

[111] Lavanga, M., De Wel, O., Caicedo, A., Heremans, E., Jansen, K., Dereymaeker, A., Naulaers, G., and Van Huffel, S. Automatic quiet sleep detection based on multifractality in preterm neonates: effects of maturation. In *Proc. 39th Annual International Conference of the IEEE Engineering in Medicine & Biology Society* (Seogwipo, South Korea, 2017).

[112] Lavanga, M., De Wel, O., Caicedo, A., Jansen, K., Dereymaeker, A., Naulaers, G., and Huffel, S. V. A brain-age model for preterm infants based on functional connectivity. *Physiological Measurement 39*, 4 (2018), 44006.

[113] Lavanga, M., De Wel, O., Caicedo, A., Jansen, K., Dereymaeker, A., Naulaers, G., and Van Huffel, S. Linear and nonlinear functional connectivity methods to predict brain maturation in preterm babies. In *Proc. of the 8th international workshop on biosignal interpretation.* (2016), pp. 1–4.

[114] Lavanga, M., De Wel, O., Caicedo, A., Jansen, K., Dereymaeker, A., Naulaers, G., and Van Huffel, S. Monitoring effective connectivity in the preterm brain: a graph approach to study maturation. *Complexity* (2017), 1–13.

[115] Lawrence, S., Giles, C. L., Tsoi, A. C., and Back, A. D. Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks 8*, 1 (1997), 98–113.

[116] LeCun, Y., and Bengio, Y. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks 3361*, 10 (1995).

[117] LeCun, Y., Bengio, Y., and Hinton, G. Deep learning. *Nature 521*, 7553 (2015), 436–444.

[118] Libenson, M. H. The electroencephalogram of the newborn. In *Practical Approach to Electroencephalography*. Elsevier Health Sciences, 2010, ch. 13, pp. 301–328.

[119] Limperopoulos, C., Bassan, H., Gauvreau, K., Robertson, R. L., Sullivan, N. R., Benson, C. B., Avery, L., Stewart, J., Soul,

J. S., Ringer, S. A., Volpe, J. J., and DuPlessis, A. J. Does cerebellar injury in premature infants contribute to the high prevalence of long-term cognitive, learning, and behavioral disability in survivors? *Pediatrics 120*, 3 (2007), 584–593.

[120] Lipsitz, L. A., and Goldberger, A. L. Loss of 'Complexity' and Aging. *Jama 267*, 13 (1992), 1806–1809.

[121] Liu, K., da Costa, J. P. C., So, H. C., Huang, L., and Ye, J. Detection of number of components in CANDECOMP/PARAFAC models via minimum description length. *Digital Signal Processing 51* (2016), 110–123.

[122] Liu, L., Oza, S., Hogan, D., Chu, Y., Perin, J., Zhu, J., Lawn, J. E., Cousens, S., Mathers, C., and Black, R. E. Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the Sustainable Development Goals. *Lancet 388* (2016), 3027–3035.

[123] Lloyd, R., Goulding, R., Filan, P., and Boylan, G. Overcoming the practical challenges of electroencephalography for very preterm infants in the neonatal intensive care unit. *Acta paediatrica 104*, 2 (2015), 152–157.

[124] Lorenzo-Seva, U., and Ten Berge, J. M. F. Tucker's Congruence Coefficient as a Meaningful Index of Factor Similarity. *Methodology 2*, 2 (2006), 57–64.

[125] Lu, W. Y., Chen, J. Y., Chang, C. F., Weng, W. C., Lee, W. T., and Shieh, J. S. Multiscale entropy of electroencephalogram as a potential predictor for the prognosis of neonatal seizures. *PLoS ONE 10*, 12 (2015), e0144732.

[126] Malmivuo, J., and Plonsey, R. *Bioelectromagnetism Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford University Press, New York, 1995.

[127] Manor, B., and Lipsitz, L. A. Physiologic complexity and aging: Implications for physical function and rehabilitation. *Progress in Neuro-Psychopharmacology and Biological Psychiatry 45* (2013), 287–293.

[128] Mariani, S., Borges, A. F. T., Henriques, T., Goldberger, A. L., and Costa, M. D. Use of multiscale entropy to facilitate artifact detection in electroencephalographic signals. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2015), NIH Public Access, pp. 7869–7872.

[129] Matic, V., Cherian, P. J., Koolen, N., Naulaers, G., Swarte, R. M., Govaert, P., Van Huffel, S., and De Vos, M. Holistic approach for automated background EEG assessment in asphyxiated full-term infants. *Journal of Neural Engineering 11* (2014), 066007.

[130] Maynard, D., Prior, P. F., and Scott, D. F. Device for continuous monitoring of cerebral activity in resuscitated patients. *British medical journal 4*, 5682 (1969), 545–6.

[131] McHugh, M. L. Interrater reliability: The kappa statistic. *Biochemia Medica 22*, 3 (2012), 276–282.

[132] McNamee, R. Regression modelling and other methods to control confounding. *Occupational and Environmental Medicine 62*, 7 (2005), 500–506.

[133] Meijer, E. J., Hermans, K. H. M., Zwanenburg, A., Jennekens, W., Niemarkt, H. J., Cluitmans, P. J. M., Van Pul, C., Wijn, P. F. F., and Andriessen, P. Functional connectivity in preterm infants derived from EEG coherence analysis. *European Journal of Paediatric Neurology 18*, 6 (2014), 780–789.

[134] Mikkelsen, K., and De Vos, M. Personalizing deep learning models for automatic sleep staging. *arXiv:1801.02645.* (2018).

[135] Miwakeichi, F., Martínez-Montes, E., Valdés-Sosa, P. A., Nishiyama, N., Mizuhara, H., and Yamaguchi, Y. Decomposing EEG data into space-time-frequency components using Parallel Factor Analysis. *NeuroImage 22*, 3 (2004), 1035–1045.

[136] Mizuno, T., Takahashi, T., Cho, R. Y., Kikuchi, M., Murata, T., Takahashi, K., and Wada, Y. Assessment of EEG dynamical complexity in Alzheimer's disease using multiscale entropy. *Clinical Neurophysiology 121*, 9 (2010), 1438–1446.

[137] Morabito, F. C., Labate, D., La Foresta, F., Bramanti, A., Morabito, G., and Palamara, I. Multivariate multi-scale permutation entropy for complexity analysis of Alzheimer's disease EEG. *Entropy 14*, 7 (2012), 1186–1202.

[138] Murray, D. M., Boylan, G. B., Ryan, C. A., and Connolly, S. Early EEG findings in hypoxic-ischemic encephalopathy predict outcomes at 2 years. *Pediatrics 124*, 3 (2009).

[139] Ng, S. M., Turner, M. A., Gamble, C., Didi, M., Victor, S., Atkinson, J., Sluming, V., Parkes, L. M., Tietze, A., Abernethy,

L. J., and Weindling, A. M. Effect of thyroxine on brain microstructure in extremely premature babies: magnetic resonance imaging findings in the TIPIT study. *Pediatric Radiology 44*, 8 (2014), 987–996.

[140] Ng, S. M., Turner, M. A., Gamble, C., Didi, M., Victor, S., Manning, D., Settle, P., Gupta, R., Newland, P., and Weindling, A. M. An explanatory randomised placebo controlled trial of levothyroxine supplementation for babies born. *Trials 14*, 1 (2013), 211.

[141] Niemarkt, H. J., Andriessen, P., Peters, C. H. L., Pasman, J. W., Zimmermann, L. J., and Bambang Oetomo, S. Quantitative analysis of maturational changes in EEG background activity in very preterm infants with a normal neurodevelopment at 1year of age. *Early Human Development 86*, 4 (2010), 219–224.

[142] Niemarkt, H. J., Jennekens, W., Pasman, J. W., Katgert, T., van Pul, C., Gavilanes, A. W. D., Kramer, B. W., Zimmermann, L. J., Bambang Oetomo, S., and Andriessen, P. Maturational Changes in Automated EEG Spectral Power Analysis in Preterm Infants. *Pediatric Research 70*, 5 (2011), 529–534.

[143] Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., and Hallett, M. Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clinical Neurophysiology 115*, 10 (2004), 2292–2307.

[144] Nosarti, C., Nam, K. W., Walshe, M., Murray, R. M., Cuddy, M., Rifkin, L., and Allin, M. P. G. Preterm birth and structural brain alterations in early adulthood. *NeuroImage: Clinical 13*, 6 (2014), 180–191.

[145] Oates, J., Karmiloff-Smith, A., and Johnson, M. Developing Brains. *Early Childhood in Focus 7* (2012).

[146] Ortinau, C., and Neil, J. The neuroanatomy of prematurity: Normal brain development and the impact of preterm birth. *Clinical Anatomy 28*, 2 (2015), 168–183.

[147] Oseledets, I. V. Tensor-Train Decomposition. *SIAM Journal on Scientific Computing 33*, 5 (2011), 2295–2317.

[148] Osredkar, D., Toet, M. C., van Rooij, L. G. M., van Huffelen, A. C., Groenendaal, F., and de Vries, L. S. Sleep-Wake Cycling on Amplitude-Integrated Electroencephalography in Term Newborns With Hypoxic-Ischemic Encephalopathy. *Pediactrics 115*, 2 (2005), 327–332.

[149] O'Toole, J. M., and Boylan, G. B. Quantitative Preterm EEG Analysis: The Need for Caution in Using Modern Data Science Techniques. *Frontiers in Pediatrics 7*, 174 (2019).

[150] O'Toole, J. M., Boylan, G. B., Vanhatalo, S., and Stevenson, N. J. Estimating functional brain maturity in very and extremely preterm neonates using automated analysis of the electroencephalogram. *Clinical Neurophysiology 127*, 8 (2016), 2910–2918.

[151] Padhy, S., Goovaerts, G., Boussé, M., De Lathauwer, L., and Van Huffel, S. The power of tensor-based approaches in cardiac applications. In *Biomedical Signal Processing - Advances in Theory, Algorithms and Applications*. Springer, 2019.

[152] Palmu, K., Kirjavainen, T., Stjerna, S., Salokivi, T., and Vanhatalo, S. Sleep wake cycling in early preterm infants: Comparison of polysomnographic recordings with a novel EEG-based index. *Clinical Neurophysiology 124*, 9 (2013), 1807–1814.

[153] Palmu, K., Stevenson, N., Wikström, S., Hellström-Westas, L., Vanhatalo, S., and Palva, J. M. Optimization of an NLEO-based algorithm for automated detection of spontaneous activity transients in early preterm EEG. *Physiological Measurement 31*, 11 (2010), N85–N93.

[154] Parmelee, A. H., Schulte, F. J., Akiyama, Y., Wenner, W. H., Schultz, M. A., And, E., and Stern, M. A. Maturation of EEG activity during sleep in premature infants. *Electroencephalography and clinical neurophysiology 24*, 4 (1968), 319–329.

[155] Pavlidis, E., Lloyd, R. O., Mathieson, S., and Boylan, G. B. A review of important electroencephalogram features for the assessment of brain maturation in premature infants. *Acta Paediatrica, International Journal of Paediatrics 106*, 9 (2017), 1394–1408.

[156] Phan, H., Andreotti, F., Cooray, N., Chén, O. Y., and De Vos, M. SeqSleepNet: End-to-End Hierarchical Recurrent Neural Network for Sequence-to-Sequence Automatic Sleep Staging. *IEEE Transactions on Neural Systems and Rehabilitation Engineering 27*, 3 (2019), 400–410.

[157] Picchietti, D. L., Sheldon, S. H., and Iber, C. The Visual Scoring of Sleep and Arousal in Infants and Children. *Journal of Clinical Sleep Medicine 3*, 1 (2007), 201–240.

[158] Pickler, R. H., McGrath, J. M., Reyna, B. A., McCain, N., Lewis, M., Cone, S., Wetzel, P., and Best, A. A Model of Neurodevelopmental Risk and Protection for Preterm Infants. *The Journal of Perinatal & Neonatal Nursing 24*, 4 (2010), 356–365.

[159] Pillay, K., Dereymaeker, A., Jansen, K., Naulaers, G., Van Huffel, S., and De Vos, M. Automated EEG sleep staging in the term-age baby using a generative modelling approach. *Journal of Neural Engineering 15*, 3 (2018), 036004.

[160] Pincus, S. M. Approximate entropy as a measure of system complexity. *Proc. Nati. Acad. Sci. USA 88*, 6 (1991), 2297–2301.

[161] Pincus, S. M., Gladstone, I. M., and Ehrenkranz, R. A. A regularity statistic for medical data analysis. *Journal of Clinical Monitoring 7*, 4 (1991), 335–345.

[162] Piryatinska, A., Terdik, G., Woyczynskic, W. A., Loparo, K. A., Scher, M. S., and Zlotnikf, A. Automated detection of neonate EEG sleep stages. *Computer Methods and Programs in Biomedicine 95*, 1 (2009), 31–46.

[163] Ramenghi, L. A., Rutherford, M., Fumagalli, M., Bassi, L., Messner, H., Counsell, S., and Mosca, F. Neonatal neuroimaging: Going beyond the pictures. *Early Human Development 85*, 10 (2009), S75–S77.

[164] Rawlings, J. O., Pantula, S. G., and Dickey, D. A. *Applied Regression Analysis: A Research Tool*. Springer Science & Business Media, 2001.

[165] Richman, J. S., and Moorman, J. R. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology: Heart and Circulatory Physiology 278*, 6 (2000), H2039–H2049.

[166] Saigal, S., and Doyle, L. W. An overview of mortality and sequelae of preterm birth from infancy to adulthood. *The Lancet 371*, 9608 (2008), 261–269.

[167] Santana-Miranda, R., Murata, C., Bruni, O., Rosa, A., Ruiz, G. A. A., Montoya, C. R. C., Rojas-Zamorano, J. Á., Esqueda-León, E., Dominguez-Salazar, E., Poblano, A., and Velazquez-Moctezuma, J. Cyclic alternating pattern in infants with congenital hypothyroidism. *Brain and Development 41*, 1 (2019), 66–71.

[168] Scher, M. S. Ontogeny of EEG-sleep from neonatal through infancy periods. *Sleep Medicine 9*, 6 (2008), 615–636.

[169] Scher, M. S., Waisanen, H., Loparo, K. A., and Johnson, M. W. Prediction of Neonatal State and Maturational Change Using Dimensional Analysis. *Journal of Clinical Neurophysiology 22*, 3 (2005), 159–165.

[170] Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Networks 61* (2015), 85–117.

[171] Schneider, A., Hommel, G., and Blettner, M. Linear Regression Analysis - Part 14 of a Series on Evaluation of Scientific Publications. *Dtsch Arztebl Int 107*, 44 (2010), 776–82.

[172] Shellhaas, R. A., Burns, J. W., Hassan, F., Carlson, M. D., Barks, J. D. E., and Chervin, R. D. Neonatal Sleep–Wake Analyses Predict 18-month Neurodevelopmental Outcomes. *Sleep 40*, 11 (2017).

[173] Shen, D., Wu, G., and Suk, H.-I. Deep Learning in Medical Image Analysis. *Annual review of biomedical engineering 19* (2017), 221–248.

[174] Sidiropoulos, N. D., De Lathauwer, L., Fu, X., Huang, K., Papalexakis, E. E., and Faloutsos, C. Tensor Decomposition for Signal Processing and Machine Learning. *IEEE Transactions on Signal Processing 65*, 13 (2017), 3551–3582.

[175] Sim, J., and Wright, C. C. The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy 85*, 3 (2005), 257–268.

[176] Simonyan, K., and Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[177] Sorber, L., Van Barel, M., and De Lathauwer, L. Optimization-based algorithms for tensor decompositions: Canonical polyadic decomposition, decomposition in rank-(Lr, Lr, 1) terms, and a new generalization. *SIAM Journal on Optimization 23*, 2 (2013), 695–720.

[178] Sors, A., Bonnet, S., Mirek, S., Vercueil, L., and Payen, J. F. A convolutional neural network for sleep stage scoring from raw single-channel EEG. *Biomedical Signal Processing and Control 42* (2018), 107–114.

[179] Stevenson, N. J., Oberdorfer, L., Koolen, N., O'Toole, J. M., Werther, T., Klebermass-Schrehof, K., and Vanhatalo, S. Functional maturation in preterm infants measured by serial recording of cortical activity. *Scientific Reports 7*, 1 (2017), 12969.

[180] Stevenson, N. J., O'Toole, J. M., Korotchikova, I., and Boylan, G. B. Artefact detection in neonatal EEG. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2014), IEEE, pp. 926–929.

[181] Stevenson, N. J., Palmu, K., Wikström, S., Hellström-Westas, L., and Vanhatalo, S. Measuring brain activity cycling (BAC) in long term EEG monitoring of preterm babies. *Physiological Measurement 35*, 7 (2014), 1493–1508.

[182] Stiles, J., and Jernigan, T. L. The Basics of Brain Development. *Neuropsychology review 20*, 4 (2010), 327–48.

[183] Stockart-Pope, J., Werner, S., and Bickford, R. *Atlas of Neonatal Electroencephalography*, 2 ed. New York: Raven Press, 1992.

[184] Supratak, A., Dong, H., Wu, C., and Guo, Y. DeepSleepNet: a Model for Automatic Sleep Stage Scoring based on Raw Single-Channel EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering 25*, 11 (2017), 1998–2008.

[185] Suykens, J. A., and Vandewalle, J. Least squares support vector machine classifiers. *Neural Processing Letters 9*, 3 (1999), 293–300.

[186] Suykens, J. A. K., Van Gestel, T., De Brabanter, J., De Moor, B., and Vandewalle, J. *Least Squares Support Vector Machines*. World Scientific, 2002.

[187] Tataranno, M. L., Claessens, N. H. P., Moeskops, P., Toet, M. C., Kersbergen, K. J., Buonocore, G., Išgum, I., Leemans, A., Counsell, S., Groenendaal, F., de Vries, L. S., and Benders, M. J. N. L. Changes in brain morphology and microstructure in relation to early brain activity in extremely preterm infants. *Pediatric Research 83*, 4 (2017), 834–842.

[188] Tau, G. Z., and Peterson, B. S. Normal Development of Brain Circuits. *Neuropsychopharmacology 35*, 1 (2010), 147–168.

[189] Tibshirani, R. Regression Selection and Shrinkage via the Lasso. *Journal of the Royal Statistical Society B 58*, 1 (1996), 267–288.

[190] Timmerman, M. E., and Kiers, H. A. L. Three-mode principal components analysis: Choosing the numbers of components and sensitivity to local optima. *British Journal of Mathematical and Statistical Psychology 53*, 1 (2000), 1–16.

[191] Tokariev, A., Roberts, J. A., Zalesky, A., Zhao, X., Vanhatalo, S., Breakspear, M., and Cocchi, L. Large-scale brain modes reorganize between infant sleep states and carry prognostic information for preterms. *Nature Communications 10*, 1 (2019), 2619.

[192] Tokariev, A., Videman, M., Palva, J. M., and Vanhatalo, S. Functional Brain Connectivity Develops Rapidly Around Term Age and Changes Between Vigilance States in the Human Newborn. *Cerebral Cortex 26*, 12 (2016), 4540–4550.

[193] Tomasi, G., and Bro, R. A comparison of algorithms for fitting the PARAFAC model. *Computational Statistics & Data Analysis 50*, 7 (2006), 1700–1734.

[194] Tsinalis, O., Matthews, P. M., Guo, Y., and Zafeiriou, S. Automatic Sleep Stage Scoring with Single-Channel EEG Using Convolutional Neural Networks. *arXiv preprint arXiv:1610.01683* (2016).

[195] Valencia, F. J., Porta, A., Vallverdù, M., Clarià, F., Baranowski, R., Orłowska-Baranowska, E., and Caminal, P. Refined Multiscale Entropy : Application to 24-h Holter Recordings of Heart Period Variability in Healthy and Aortic Stenosis Subjects. *IEEE Transactions on Biomedical Engineering 56*, 9 (2009), 2202–2213.

[196] van den Hoogen, A., Teunis, C. J., Shellhaas, R. A., Pillen, S., Benders, M., and Dudink, J. How to improve sleep in a neonatal intensive care unit: A systematic review. *Early Human Development 113* (2017), 78–86.

[197] Van Eyndhoven, S., Vervliet, N., De Lathauwer, L., and Van Huffel, S. Identifying Stable Components of Matrix/Tensor Factorizations via Low-Rank Approximation of Inter-Factorization Similarity. In *Proc. of the 27th European Signal Processing Conference (EUSIPCO)* (A Coruna, Spain, 2019), pp. 2–6.

[198] van Wassenaer, A. G., Kok, J. H., de Vijlder, J. J. M., Briët, J. M., Smit, B. J., Tamminga, P., van Baar, A., Dekker, F. W., and Vulsma, T. Effects of Thyroxine Supplementation on Neurologic Development in Infants Born at Less Than 30 Weeks' Gestation. *New England Journal of Medicine 336*, 1 (1997), 21–26.

[199] van Wezel-Meijler, G., Steggerda, S. J., and Leijser, L. M. Cranial Ultrasonography in Neonates: Role and Limitations. *Seminars in Perinatology 34*, 1 (2010), 28–38.

[200] Vandecappelle, M., Vervliet, N., and De Lathauwer, L. Nonlinear least squares updating of the canonical polyadic decomposition. In *2017 25th European Signal Processing Conference (EUSIPCO)* (2017), IEEE, pp. 663–667.

[201] Vanhatalo, S., and Kaila, K. Development of neonatal EEG activity: from phenomenology to physiology. *Seminars in fetal & neonatal medicine 11*, 6 (2006), 471–8.

[202] Vapnik, V. *The nature of statistical learning theory.* Springer Science & Business Media, New York, 1995.

[203] Varon, C., Alzate, C., Suykens, J. a. K., and Member, S. Noise Level Estimation for Model Selection in Kernel PCA Denoising. *IEEE Transactions on Neural Networks and Learning Systems 26*, 11 (2015), 2650–2663.

[204] Vedaldi, A., and Lenc, K. MatConvNet: Convolutional Neural Networks for MATLAB. In *Proceedings of the 23rd ACM international conference on Multimedia* (2015), ACM, pp. 689–692.

[205] Vervliet, N., Debals, O., Sorber, L., Van Barel, M., and De Lathauwer, L. *Tensorlab 3.0*, 2016.

[206] Visser, T. J. Thyroid hormone transport across the placenta. *Annales d'Endocrinologie 77*, 6 (2016), 680–683.

[207] Volpe, J. J. *Neurology of the Newborn*, 5 ed. Saunders, 2008.

[208] Wallois, F. Synopsis of maturation of specific features in EEG of premature neonates. *Neurophysiology 40*, 2 (2010), 125–126.

[209] Watanabe, E., Kiyono, K., Hayano, J., Yamamoto, Y., Inamasu, J., Yamamoto, M., Ichikawa, T., Sobue, Y., Harada, M., and Ozaki, Y. Multiscale entropy of the heart rate variability for the prediction of an ischemic stroke in patients with permanent atrial fibrillation. *PLoS ONE 10*, 9 (2015), e0137144.

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

[211] Whitehead, K., Pressler, R., and Fabrizi, L. Characteristics and clinical significance of delta brushes in the EEG of premature infants. *Clinical neurophysiology practice 2* (2017), 12–18.

[212] Wikström, S., Pupp, I. H., Rosén, I., Norman, E., Fellman, V., Ley, D., and Hellström-Westas, L. Early single-channel aEEG/EEG predicts outcome in very preterm infants. *Acta paediatrica 101*, 7 (2012), 719–726.

[213] Willhelm, C., Girisch, W., Gottschling, S., Gräber, S., Wahl, H., and Meyer, S. Systematic Cochrane Reviews in Neonatology: A Critical Appraisal. *Pediatrics & Neonatology 54*, 4 (2013), 261–266.

[214] Williams, F. L. R., Simpson, J., Delahunty, C., Ogston, S. A., Bongers-Schokking, J. J., Murphy, N., van Toor, H., Wu, S.-Y., Visser, T. J., and Hume, R. Developmental Trends in Cord and Postpartum Serum Thyroid Hormones in Preterm Infants. *The Journal of Clinical Endocrinology & Metabolism 89*, 11 (2004), 5314–5320.

[215] Wolf, M., Naulaers, G., Van Bel, F., Kleiser, S., and Greisen, G. A review of near infrared spectroscopy for term and preterm newborns. *Journal of Near Infrared Spectroscopy 20*, 1 (2012), 43–55.

[216] Yang, A. C., Hong, C.-J., Liou, Y.-J., Huang, K.-L., Huang, C.-C., Liu, M.-E., Lo, M.-T., Huang, N. E., Peng, C.-K., Lin, C.-P., and Tsai, S.-J. Decreased Resting-State Brain Activity Complexity in Schizophrenia Characterized by Both Increased Regularity and Randomness. *Human Brain Mapping 36*, 6 (2015), 2174–2186.

[217] Zanin, M., Zunino, L., Rosso, O. A., and Papo, D. Permutation entropy and its main biomedical and econophysics applications: A review. *Entropy 14*, 8 (2012), 1553–1577.

[218] Zhang, D., Ding, H., Liu, Y., Zhou, C., Ding, H., and Ye, D. Neurodevelopment in newborns: a sample entropy analysis of electroencephalogram. *Physiological measurement 30*, 5 (2009), 491–504.

[219] Zhang, X., Chen, X., Barkhaus, P. E., and Zhou, P. Multiscale Entropy Analysis of Different Spontaneous Motor Unit Discharge Patterns. *IEEE Journal of Biomedical and Health Informatics 17*, 2 (2013), 470–476.

[220] Zhou, G., Zhao, Q., Zhang, Y., Adali, T., Xie, S., and Cichocki, A. Linked Component Analysis From Matrices to High-Order Tensors: Applications to Biomedical Data. *Proceedings of the IEEE 104*, 2 (2016), 310–331.

Curriculum vitae

Ofelie De Wel was born in Bruges, Belgium on October 13, 1992.
In September 2010 she started studying engineering at KU Leuven. During the first year of the master "Biomedical Engineering" she studied abroad at Trinity College Dublin in the framework of an Erasmus programme. During her master thesis entitled *'Tensor-based classification of irregular heartbeats'* she explored multilinear discriminant analysis to perform irregular heartbeat classification. She received a M.Sc. degree in Biomedical Engineering (magna cum laude) at KU Leuven, Belgium, in June 2015.
In October 2015 she joined STADIUS, Department of Electrical Engineering (ESAT), KU Leuven as a PhD student under the supervision of Prof. dr. ir. Sabine Van Huffel and Prof. dr. Gunnar Naulaers. She incorporated tensor approaches in neonatal EEG signal processing as part of the Biotensors project. In close collaboration with the University Hospitals Leuven, she developed automated algorithms to track the sleep-wake cycling and brain maturation of preterm infants.

List of publications

Articles in internationally reviewed journals

1. Ansari A. H., **De Wel O.**, Pillay K., Dereymaeker A., Jansen K., Van Huffel S., De Vos M. (2019). A convolutional neural network outperforming state-of-the-art sleep staging algorithms for both preterm and term infants. Accepted for publication in *Journal of Neural Engineering*.

2. **De Wel O.**, Lavanga M., Caicedo A., Jansen K., Naulaers G., Van Huffel S. (2019). Decomposition of a Multiscale Entropy Tensor for Sleep Stage Identification in Preterm Infants. *Entropy*, 21(10), 936.

3. Hendrikx D., Smits A., Lavanga M., **De Wel O.**, Thewissen L., Jansen K., Naulaers G. (2019). Measurement of neurovascular coupling in neonates. *Frontiers in Physiology*, 10(65).

4. Dereymaeker A., Matic V., Vervisch J., Cherian P. J., Ansari A. H., **De Wel O.**, Jansen K. (2019). Automated EEG background analysis to identify neonates with hypoxic-ischemic encephalopathy treated with hypothermia at risk for adverse outcome: A pilot study. *Pediatrics and Neonatology*, 60(1), 50–58.

5. (Ansari A. H., **De Wel O.**)*, Lavanga M., Caicedo A., Dereymaeker A., Jansen K., Vervisch J., De Vos M., Naulaers G., Van Huffel S. (2018). Quiet sleep detection in preterm infants using deep convolutional neural networks. *Journal of Neural Engineering*, 15 (6), 066006.
(*: joint first author)

6. Lavanga M., **De Wel O.**, Caicedo A., Jansen K., Dereymaeker A., Naulaers G., Van Huffel S. (2018). A brain-age model for preterm infants based on functional connectivity. *Physiological Measurement*, 39(4), 44006.

7. Lavanga M., **De Wel O.**, Caicedo A., Jansen K., Dereymaeker A., Naulaers G., Van Huffel S. (2017). Monitoring effective connectivity in the preterm brain: a graph approach to study maturation. *Complexity*.

8. **De Wel O.**, Lavanga M., Caicedo A., Jansen K., Dereymaeker A., Naulaers G., Van Huffel S. (2017). Complexity Analysis of Neonatal EEG Using Multiscale Entropy: Applications in Brain Maturation and Sleep Stage Classification. *Entropy*, 19 (10), 516.

Articles in preparation

1. **De Wel O.**, Van Huffel S., Lavanga M., Isgum I., Dudink J., Groenendaal F., de Vries L., Dereymaeker A., Jansen K., Naulaers G., Benders M., Tataranno M.L. (2020). Relationship between early functional and structural brain development in preterm infants.

2. (**De Wel O.**, Eerdekens A.)*, Lavanga M., Caicedo A., Jansen K., Vanhole C., Van Huffel S., Naulaers G. (2020). Measurement of thyroid hormone action in the preterm infants' brain using EEG.
(*: joint first author)

Papers in proceedings of international conferences

1. Lavanga M., **De Wel O.**, Caicedo A., Deviaene M., Moeyersons J., Varon C., Bollen B., Jansen K., Ortibus E., Naulaers G., Van Huffel S. (2019). The implementation of an apnea-based perinatal stress calculator. In *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 6000-6003). Berlin, Germany.

2. Caicedo A., **De Wel O.**, Vandecappelle M., Thewissen L., Smits A., Allegaert K., Van Huffel S. (2019). Monitoring of Brain Hemodynamics Coupling in Neonates using Updated Tensor Decompositions. In *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 660–663). Berlin, Germany.

3. Lavanga M., **De Wel O.**, Caicedo A., Heremans E., Jansen K., Dereymaeker A., Van Huffel S. (2017). Automatic quiet sleep detection based on multifractality in preterm neonates: effects of maturation. In *Proc. 39th Annual International Conference of the IEEE Engineering in Medicine* & Biology Society. Seogwipo, South Korea.

4. Lavanga M., **De Wel O.**, Caicedo A., Jansen K., Dereymaeker A., Naulaers G., Van Huffel S. (2016). Linear and nonlinear functional connectivity methods to predict brain maturation in preterm babies. In *Proc. of the 8th international workshop on biosignal interpretation.* (pp. 1–4). Osaka, Japan.

Abstracts in proceedings of (inter)national conferences

1. **De Wel O.**, Lavanga M., Caicedo A., Jansen K., Naulaers G., Van Huffel S. (2019). Sleep stage identification in preterm infants based on the decomposition of a multiscale entropy tensor. In *The 11th International Conference on Brain Monitoring and Neuroprotection in the Newborn.* Clearwater Beach, Florida, USA.

2. **De Wel O.**, Lavanga M., Caicedo A., Jansen K., Naulaers G., Van Huffel S. (2019). Decomposition of a multiscale entropy tensor for sleep stage identification in preterm neonates. In *7th Dutch Bio-Medical Engineering Conference.* Egmond aan Zee, The Netherlands.

3. **De Wel O.**, Ansari A. H., Lavanga M., Caicedo A., Dereymaeker A., Jansen K., De Vos M., Naulaers G., Van Huffel S. (2018). Automated EEG sleep staging in preterm infants using deep learning. In *17th National Day on Biomedical Engineering.* Brussels, Belgium.

4. **De Wel O.**, Lavanga M., Caicedo A., Jansen K., Naulaers G., Van Huffel S. (2019). Decomposition of a multiscale entropy tensor for sleep stage identification in preterm neonates. In *EURASIP Summer School, Tensor-Based Signal Processing.* Leuven, Belgium.

5. **De Wel O.**, Lavanga M., Caicedo A., Jansen K., Dereymaeker A., Naulaers G., Van Huffel S. (2017). The Evolution of EEG Complexity during Brain Maturation in Preterm Infants using Multiscale Entropy. In *16th National Day on Biomedical Engineering.* Brussels, Belgium.

6. **De Wel O.**, Lavanga M., Caicedo A., Jansen K., Dereymaeker A., Naulaers G., Van Huffel S. (2017). The Evolution of EEG Complexity during Brain Maturation in Preterm Infants using Multiscale Entropy. In *The 10th International Conference on Brain Monitoring and Neuroprotection in the Newborn.* Killarney, Ireland.

7. **De Wel O.**, Lavanga M., Caicedo A., Jansen K., Naulaers G., Van Huffel S. (2017) Canonical Polyadic Decomposition of a Multiscale Entropy

Tensor for Sleep Stage Identification in Preterm Neonates. In *Coupling and Causality in Complex Systems*. Cologne, Germany.



FACULTY OF ENGINEERING SCIENCE
DEPARTMENT OF ELECTRICAL ENGINEERING
STADIUS CENTER FOR DYNAMICAL SYSTEMS, SIGNAL PROCESSING AND DATA ANALYTICS
Kasteelpark Arenberg 10 box 2446
B-3001 Leuven
ofelie.dewel@kuleuven.be
http://www.esat.kuleuven.be/stadius/