Computational features of neonatal EEG monitoring after asphyxia

Julia Jaatela

School of Science

Thesis submitted for examination for the degree of Master of Science in Technology.

Helsinki 23.1.2017

Thesis supervisor:

Prof. Mikko Sams

Thesis advisors:

PhD. Anton Tokariev

PhD. Nathan Stevenson

Aalto University
School of Science

AALTO UNIVERSITY
SCHOOL OF SCIENCE

ABSTRACT OF THE
MASTER'S THESIS

Author: Julia Jaatela

Title: Computational features of neonatal EEG monitoring after asphyxia

Date: 23.1.2017 Language: English Number of pages: 7+64

Department of Neuroscience and Biomedical Engineering

Professorship: Human Neuroscience and -technology

Supervisor: Prof. Mikko Sams

Advisors: PhD. Anton Tokariev, PhD. Nathan Stevenson

The aim of this thesis was to find out if computational electroencephalography (EEG) features can be used in the automated monitoring of newborns after asphyxia. EEG is already widely used in the neonatal intensive care units but there is a need for quantitative measures that can be obtained without the presence of a clinical expert. One of the biggest challenges in the treatment of newborns with asphyxia is also to correctly estimate the severity of the resulting neurological problems.
Eight different feature classes were computed for 42 full-term babies from periods of quiet and active sleep. These feature classes measured correlations of amplitude and phase, interhemispheric synchrony, multifractality and spectral properties. We then studied the ability of these features to distinguish between different severity groups and also tested a classification algorithm to predict the outcome of the babies. Quiet sleep was noted to be more sensitive when separating groups with different grades of severity and most of the used feature classes showed significant results in statistical testing between the groups. The babies with the normal outcome were classified more accurately with the EEG based classification algorithm, than with only the clinical estimation.

Keywords: EEG, HIE, SVM, Neonate

AALTO-YLIOPISTO
PERUSTIETEIDEN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Julia Jaatela

Työn nimi: Laskennalliset muuttujat vastasyntyneen aivomonitoroinnin arvioinnissa

Päivämäärä: 23.1.2017 Kieli: Englanti Sivumäärä: 7+64

Neurotieteen ja lääketieteellisen tekniikan laitos

Professuuri: Neurotiede ja -teknologia

Työn valvoja: Prof. Mikko Sams

Työn ohjaajat: PhD. Anton Tokariev, PhD. Nathan Stevenson

Työn tarkoituksena oli selvittää, onko aivosähkökäyrästä (EEG) laskettuja parametreja mahdollista käyttää happivajeesta kärsineiden vastasyntyneiden automaattisessa monitoroinnissa. EEG on jo nyt yleisesti käytössä vastasyntyneiden teho-osastoilla, mutta tarve kvantitatiivisille mittareille, joiden tulkintaan ei tarvita lääketieteen asiantuntijaa, on suuri. Lisäksi yksi suurimmista haasteista on pystyä arvioimaan tarkasti, kuinka vakaviin neurologisiin ongelmiin happivaje johtaa.
Työssä laskettiin kahdeksan erilaista muuttujajoukkoa 42 täysiaikaiselle vauvalle sekä hiljaisen että aktiivisen unen aikana. Nämä muuttujat mittasivat amplitudin ja vaiheen korrelaatioita, aivopuoliskojen välistä synkroniaa, multifraktaalisuutta sekä taajuusjakaumaa. Tämän jälkeen tutkittiin muuttujien kykyä erotella eri vakavuusasteisia ryhmiä ja testattiin luokittelualgoritmia vauvojen tulevan terveydentilan ennustamiseen. Hiljaisen unen huomattiin olevan herkempi havaitsemaan eroja eri vakavuusasteisten ryhmien välillä ja tilastollisen testauksen perusteella suurin osa valituista muuttujajoukoista erotteli merkittävästi eri vakavuusryhmiä. Ne vauvat, jotka toipuivat hapenpuutteesta täysin, pystyttiin löytämään EEG-pohjaisella luokittimella tarkemmin kuin pelkän kliinisen arvion avulla.

Avainsanat: EEG, HIE, SVM, Vastasyntynyt

Preface

This study was carried out at BABA research center at the Helsinki Children's Hospital. I would like to thank the whole research team for creating the best possible environment to do clinical research. Special thanks belong to Anton Tokariev and Nathan Stevenson for their valuable support and guidance in technical aspects and to Meri Myöhänen for doing an immense work with the dataset. In particular, I want to thank Sampsa Vanhatalo for offering this possibility and guiding me through the world of neonatal EEG with contagious enthusiasm. I also want to thank my supervisor Mikko Sams for his valuable comments on scientific writing. Finally, I want to thank my family and friends for the support and encouragement they have shown throughout my studies.

Helsinki, 23.1.2017

Julia Jaatela

Contents

Symbols and abbreviations

Symbols

α	Significance level
cov	Covariance
$D(q)$	Q-order singularity dimension
E	Expected value
Fs	Sampling frequency
$\mathcal{H}$	Hilbert transform
$h(q)$	generalized Hurst exponent
$\Im$ or Im	Imaginary part of a complex value
$P(a, b)$	Joint probability
$K(x, y)$	Kernel function
ϕ	Phase
p.v.	Cauchy principal value
$\Re$ or Re	Real part of a complex value
ρ	Pearson's linear correlation coefficient
r_s	Spearman's correlation coefficient
θ	Phase difference
σ	Standard deviation
τ	Time-lag
W	Sum of ranked values

Abbreviations

AEC	Amplitude Envelope Correlation
aEEG	Amplitude-integrated Electroencephalography
AS	Active Sleep
ASI	Activation Synchrony Index
CP	Cerebral Palsy
CPSD	Cross Power Spectral Density
EEG	Electroencephalography
ETDF	Energy-weighted Temporal Dependency Function
FFT	Fast Fourier Transformation
HIE	Hypoxic-Ischemic Encephalopathy
MFDFA	Multi-Fractal Detrended Fluctuation Analysis
MRI	Magnetic Resonance Imaging
NC	Nestedness Coefficient
NICU	Neonatal Intensive Care Unit
PLV	Phase-Locking Value
PPC	Phase-Phase Correlation
PSD	Power Spectral Density
QS	Quiet Sleep
SVM	Support Vector Machine
wPLI	Weighted Phase-Locking Index

1 Introduction

Eight percent of childhood deaths and 24% of newborn deaths worldwide are caused by asphyxia [1, 2, 3]. Asphyxia is a condition where the body does not receive enough oxygen. During birth this condition can be caused by many different reasons such as a prolapsed cord or placental abruption [4]. Although asphyxia has long been the leading cause in neonate deaths in western countries [3], very little is still known about the exact effects and mechanisms of this condition.

Asphyxia often leads to neurological problems that are usually caused by an asphyxia related brain damage called hypoxic-ischemic encephalopathy (HIE). In Finland, about 200 newborns are diagnosed with HIE every year [5]. These babies can later on be diagnosed with various long term neurological disorders like learning disability or cerebral palsy (CP). In the neonatal period the brain has an extremely high level of plasticity, which leads to both opportunities and challenges in the treatment process [6].

Neonates with a mild HIE often recover fully or have only mild neurological problems, while the ones with a severe HIE will usually have severe neurological problems if they survive. The hardest group to evaluate is the group of babies with a moderate HIE. They can either recover fully or have severe neurological problems although the original severity looks similar in a standard evaluation. [4, 7] There are not yet any existing measurements or tests that could always correctly predict the outcome of neonates with HIE.

The goal of this study was to further improve our understanding of the brain activity in neonates that have asphyxia related neurological problems. This is done by using different computational parameters obtained from a four channel EEG measurement and comparing them to the severity of HIE and outcome of the babies. This research also aims to shed light on the possibilities of using EEG in the treatment process of asphyxiated neonates.

The used data was from a continuous EEG that was measured within the first five days of the babies' life. From this data, two different 20-minute epochs were extracted for each baby: one during quiet sleep and one during active sleep. The research questions were the following:

1. Is there a difference between active and quiet sleep epochs?
2. Does the measured epochs reflect the clinical state and history of the baby?
3. What can these epochs tell us about the prognosis of the baby?
4. Can we predict the clinical outcome based only on EEG data?

The first question compares the different phases of neonatal EEG. As with adults, the brain activity of a newborn depends highly on the sleep-wake cycle. While neonates in the intensive care mostly sleep, the comparison in this study is only made between the sleep stages. With this question we try to conclude if the computed EEG features stay the same in different sleep stages and if their predictive value is present only in one of the sleep stages or in both.

With the second question, the goal was to understand if the measured EEG reflects the clinical diagnosis given by a medical doctor. If the calculated EEG variable corresponds to the severity estimate of HIE, it can be considered as a potential tool that could also help the clinician to make the diagnosis. It also confirms that EEG can represent the overall neurological state of the baby.

The third question asks about the relationship between the EEG features and the outcome of the baby. If there is a strong correlation with a certain computational feature and the severity of neurological problems, this feature could be used to evaluate the outcome prognosis. A good estimation of the future is highly valuable not only for the clinician but also for the parents.

With the fourth and final question we try to address to the need of computerized decision making and monitoring. If we could evaluate the severity of brain activity related problems automatically, the correct treatment could be given faster and more precisely. An automatized system would help in constant monitoring and especially in situations when a specialist is not available around the clock.

The subjects of this study were selected to be only babies with HIE, brain damage caused by lack of oxygen in the central nervous system. While the brain activity also depends strongly on the gestational age, only full-term babies were selected for this study. The total number of subjects was 42.

The used computed features were selected based on the literature and with the help of a medical doctor who has experience in evaluating normal and abnormal neonatal EEG visually. There were eight different feature classes in total: amplitude envelope correlation (AEC) [8], phase locking value (PLV) [9], weighted phase lag index (wPLI) [10], nestedness coefficient (NC)[11], activation synchrony index (ASI)[12], multifractal detrended fluctation analysis (MFDFA) [13], power spectral density (PSD) and cross power spectral density (CPSD). The used frequency bands were between 0.25 Hz and 30 Hz.

2 Background

In this section the background of the study field is discussed. The first part offers the basic knowledge about EEG and about the physiological phenomena it measures. Also the use of EEG in a clinical environment is discussed. The second part discusses the characteristic aspects of neonatal brain activity and the techniques that are used to measure it. The third part introduces hypoxic-ischemic encephalopathy which is the common factor across the study subjects.

2.1 Electroencephalography

Nervous system is responsible for maintaining homeostasis. It also enables us to perceive and understand the surrounding world and move through it. The most complex part of the nervous system is the brain, which contains about 86 billion neurons – cells that are able to convert stimuli into electrical impulses called action potentials. Action potentials are able to propagate along the neuron surface due to the ion movement that is mainly caused by sodium and potassium. On average, each neuron connects with 1000 other neurons, which results to thousand trillion (10^{15}) connections called synapses within the brain. [14]

The electrical activity occurring in the brain can be recorded with electroencephalography (EEG). EEG was first developed by Richard Caton in 1875 and has since been widely used in research and in different clinical applications [15].

EEG is measured by placing voltage sensitive electrodes over the scalp. The positions of the electrodes are described by the international 10-20 system that uses landmarks of the skull in order to standardize the electrode locations across subjects and trials [16, 17]. Each position has a name that consists of a number and a letter: F for frontal, C for central, T for temporal and O for occipital. Even numbers refer to the right and odd numbers to the left hemisphere. This positioning system is presented in Figure 6.

The electrodes placed over the scalp measure the potential difference between two distinct points. Potential differences are obtained either between each electrode pair or from each electrode relative to some reference point [14]. The exact number of electrodes and the use of reference point vary depending on the research setup and goals [4].

The measured potential difference is caused by electrical currents generated by the neurons that are located close to the brain surface. This outer layer of the brain is called cerebral cortex. Even if the cell is located in the brain surface, single action potentials are too small to be detected by EEG. Thus EEG measures instead of individual action potentials EEG measures the voltage fluctations of synchronized activity caused by multiple synaptic connections [14]. The mechanisms of these rhythmic oscillations that are seen with EEG are not fully understood but they are hypothesized to be caused by postsynaptic potentials that share a common cortical origin [15, 18]. The postsynaptic potential is considered to last longer than single action potentials and stretch simultaneously over a large cell membrane area. Potential differences formed in the deeper parts of the brain or horizontally oriented

cortical cells are thought to contribute only little to the recorded EEG. [15]

Oscillations are most often studied by dividing the activity into four frequency bands called delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz) and beta (13-30 Hz) [4]. In healthy adults, alpha waves are present in a relaxed but awake state, beta waves in an active state, theta waves in an emotionally stressed state and delta waves during sleep [14]. These different frequency components are presented in Figure 1.

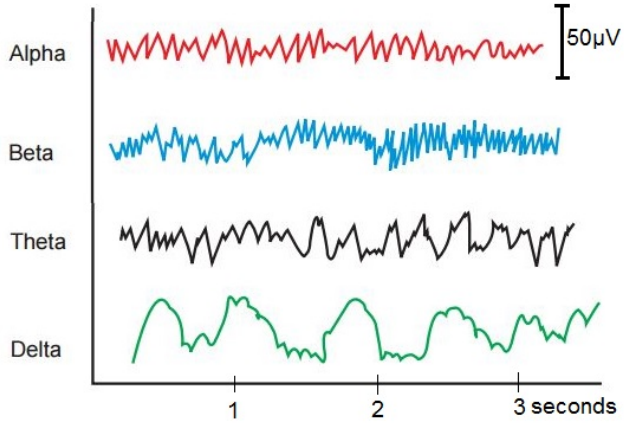


Figure 1: Four main types of EEG oscillatory activity: delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz) and beta (13-30 Hz) waves. [14]

All electrical recordings are vulnerable to artefacts that are caused by the movement of electrodes or disturbances in recording equipment and recording environment. During EEG measurements, all body movements such as blinking affect the obtained signal. Some artefacts are easy to spot because of their unusual appearance or with the help of additional monitoring such as electrocardiogram, which is an electric recording of heart activity. However, some noise is always present and can affect the interpretation of EEG. [15]

Despite the risk of artefacts, EEG is an extremely informative tool when assessing the brain activity. It is also non-invasive and can be easily used together with different equipment such as heart monitoring. EEG allows long term monitoring also in difficult environments like intensive care units.

2.2 EEG of neonatal brain

Development of the nervous system starts during the first weeks of gestation and ends in the young adulthood. In some form the central nervous system changes throughout the lifetime. [14] The highest plasticity of the brain takes place before and after birth. Most rapid transformation happens with the synapses: after the 7th gestational week the number of synapses first increases rapidly and then reaches a peak between first and fourth year of life, depending of the cortex region. The newborn brain has thus more neuronal synapses than an adult brain. [19]

Brain activity plays a critical role in the development of the brain [4]. Sensory inputs based on new experiences create neuronal activity and modify the brain after

birth. Modification happens through neuronal arborization, branching of the axons, through adding and removing synapses, through stabilizing the synaptic connections and through controlled cell death [4, 19]. These modifications can take place between neighbouring neurons or between groups of neurons located further apart. In the newborn brain local connectivity seems to be more dominant than long-distance connections. [20]

However, all neuronal development is not caused by sensory input. It has been shown that immature central nervous system produces internal or spontaneous activity that also modifies the brain [11, 21]. Especially the early brain development is mostly caused by this internal activity and is thus quite free from any environmental influence. This spontaneous activity has typically discontinuous and synchronous pattern: short activity bursts are followed by more silent periods simultaneously in the two hemispheres. Disorganization in these rhythms often lead to abnormalities in brain development. [4, 21]

The overall background activity of neonates can be classified as mainly continuous or discontinuous and usually EEG develops towards a more continuous pattern as the newborns reach full-term age of 37 weeks. The appearance of ongoing and continuous activity needs sufficiently large cortical networks that are activated via connections within the cortex or from thalamus to cortex. Although bursts in discontinuous EEG also include delta, alpha and beta frequencies, they are not physiologically the same as oscillations in mature brains. For example, delta activity is only present in sleeping adults whereas in newborns they are normal in awake state. [4, 15]

Neonatal brain activity does not depend solely on development and maturation but also on the sleep-wake cycle. Full-term babies already have detectable sleep stages that resemble the ones in adults. The sleep-wake cycle can be divided in four classes: waking, active sleep (REM sleep), quiet sleep (non-REM sleep) and unknown states that are considered as transitional phases [15]. One sleep cycle lasts approximately 50 minutes which is faster than in adults [22]. Active sleep (AS) is characterized by a continuous and quiet sleep (QS) by a discontinuous activity. QS usually has a higher voltage and lower frequency than AS. Because the EEG patterns look very similar, active sleep and wakefulness can be difficult to distinguish from each other. [15] Especially quiet sleep is diagnostically sensitive to show abnormalities in the brain activity [23].

When studying neonatal brain activity, multiple methods can be used. Visually EEG can be studied both in short and long term. The dominating background activity, continuity of EEG, localization of activity and synchrony between the hemispheres can be seen quickly from live EEG recording. These recordings are usually done with 9-15 electrodes as seen in Figure 2. Long term monitoring is often presented in a form of amplitude-integrated EEG (aEEG), which is a single or double channel recording that is filtered, rectified and smoothed before presented in a compressed 24-hour period. The aEEG signal presents the maximum and minimum variations of the EEG amplitude as a wide band from which it is easy to see the overall trend of EEG, study the development of sleep-wake cycle and spot periods of inactivity or seizures. [4]

Besides studying EEG visually, there are numerous computational methods that

can be used to characterize not only neonatal EEG but EEG in general. The main components of EEG are frequency, amplitude and phase, and most of the computational feature classes describe these parameters. One can for example calculate the amount of different frequency components of EEG signal, study the correlation of phases in different spatial locations or see if there if a certain waveform occurs simultaneously in every EEG channel.

Figure 2: Newborn EEG is usually recorded from 9-15 electrodes which are the round marks on the cap. Picture by Sampsa Vanhatalo.

2.3 Hypoxic-ischemic encephalopathy

Hypoxic-ischemic encephalopathy (HIE) is a medical condition that is defined as brain damage caused by oxygen loss, deprivation of blood supply or both [7]. The term is used only in cases where the damage takes place right before or after birth. HIE can lead to developmental or cognitive delays, motor impairments or even death [4, 7]. In this section we first discuss the mechanisms of HIE: what is HIE, what are the factors causing it and how does it affect brain activity. Next, the medical aspects such as diagnosis and treatment are presented. Finally, there is a short overview of the different methods that are used to predict the outcome of the neonates with HIE.

2.3.1 Mechanisms of HIE

The most significant risk factor causing HIE is asphyxia, lack of oxygen. Birth asphyxia can originate from low maternal blood pressure, cardiac complications, prolapsed cord, placental abruption, trauma or multiple other reasons. [4] Asphyxia is always a dangerous condition but it has even more severe effects on babies that have not reached full-term age [24]. Birth asphyxia affects the whole body and can cause multiple organ failures [4, 25].

Deprivation of oxygen leads to a widespread cell death. Cells are unable to function correctly without enough oxygen and it causes disturbances in their metabolism. Especially neuronal cells are vulnerable to the lack of oxygen while they use a lot of energy to maintain their depolarized state. [26] Cell death is usually limited to the

time period of oxygen deprivation and it often stops when there is oxygen available again. [4]

The duration and severity of asphyxia mainly determine the range of cell death in the brain. Severe asphyxia damages important parts of the brain such as basal ganglia, thalamus and key parts of sensory and motor cortices. In prolonged moderate asphyxia the most vital areas remain unharmed and the damage might be restricted to watershed area, brain regions that receive blood supply from the most distal branches of arteries. In mild or moderate asphyxia, the babies may recover fully without any permanent changes in the brain. [7, 23]

In addition to short term effects of asphyxia, it also triggers an inflammatory reaction that can continue from days to weeks. The immune system reacts to the cell death by releasing different chemical signals such as protein hormones called cytokines. These chemical signals either stimulate or inhibit normal cell functions. [14, 26] The precise functions of these secondary reactions are not fully known but they are considered to be an important factor in the severity of HIE and in the outcome of the baby [4, 25].

Because of the short and long term effects of asphyxia, brain injuries of HIE develop with a delay. Although clear damage is difficult to detect instantly, asphyxia still affects normal brain development and activity in many ways. One of the clearest signs are epileptic seizures: spikes, sharp waves and combinations of those two. Different abnormal background patterns are also common in neonates with HIE. Continuous low voltage pattern and burst suppression, alternation of very little activity and high-voltage bursts, have been connected to subjects with poor outcome. Abnormalities can be also local: only some parts of the brain show slow or suppressed activity. [4, 23] HIE can also cause disturbances in sleep-wake cycle [27, 28].

2.3.2 Diagnosis and treatment

Diagnosis of HIE is done based on a combination of multiple aspects including alertness, reflexes and brain activity. HIE severity is divided to three classes: mild (class I), moderate (class II) and severe (class III). This classification was first introduced in 1977 by Sarnat and Sarnat who presented a table that uses several visible features to estimate the severity of HIE [29]. This table has been later modified and updated many times but the basic outline of the original table is still widely in use. Table 1 shows an example of the Sarnat classification.

In addition to the visible signs, some measurements can be used to diagnose HIE and to estimate the HIE class. The amount of base present in blood and the pH-value of blood at the time of the birth can indicate if the baby has suffered from asphyxia. As discussed in the previous section, HIE has also been associated with different EEG abnormalities and early EEG monitoring can reveal these patterns. Structural analysis of the brain with different neuroimaging techniques can also be used to aid the diagnosis. [23] The neurological injury is however not a static state so the HIE classes only describe the situation at a certain time point.

After HIE is diagnosed, the main goal in the treatment is to prevent any further damage on the baby. Usually the newborn is medically monitored and treated to

Table 1: Sarnat classification of HIE. Modified from Lagercrantz and colleagues [4].

	Grade I mild	Grade II moderate	Grade III severe
Alertness	Hyperalert	Fatigue	Coma
Posture	Mild flexion	Strong flexion	Decerebration
Stretch reflexes	Overactive	Overactive	Decreased or absent
Muscle tone	Normal or increased	Hypotonia	Flaccid
Seizures	None	Frequent	Uncommon
Pupils	Dialated, reactive	Small, reactive	Variable or fixed
Respiration	Regular	Periodic	Apnea
Duration	<24h	2-14 days	Weeks

maintain a normal blood glucose level and pressure, and to prevent or control any seizures. For example, a drug called Mannitol can be used to reduce swelling and different Barbiturates to suppress seizures. [23]

It has been noticed that reduction of temperature can greatly reduce the amount of cell death in the brain in the secondary phase of HIE. Hypothermia is considered to suppresses the metabolic rate, produce certain neurotransmitters and help to maintain sufficient intracranial pressure. [30, 31]. Because of this observation, hypothermia is now a standard method in the treatment of neonates with HIE. In Neonatal intensive care unit (NICU) of Helsinki University Central Hospital, cooling treatment has been used since 2006.

Cooling is usually used with neonates with moderate or severe asphyxia and Table 2 shows the more detailed inclusion criteria. In hypothermia treatment the body temperature is lowered to 33-34°C for 72 hours after which the temperature is slowly raised back to normal. [5, 32]

Clinical outcomes of the asphyxiated neonates can be divided to different categories based on the severity of the disability that is caused by HIE. Normal outcome is considered when no neural abnormalities are found. Mild abnormality includes mild speech, motor or cognitive delay. Moderate abnormality includes mild dystonic or hemiplegic CP and severe abnormality tetraplegic or dyskinetic CP, severe mental retardation and severe epilepsy. Final outcome group is formed by babies that have died because of HIE.

2.3.3 Evaluating clinical status and outcome

Accurate estimation of the current state of a neonate with HIE is important in order to give the correct treatment. A good estimation of the future outcome is also highly valuable while it helps to modify the therapeutic decisions accordingly and give more knowledge not only for the clinician but also for the parents. Although different clinical parameters such as Apgar scores, heart rate and certain biomarkers correlate with the HIE gradus as well as the outcome, they do not work well individually [33]. The hardest group to evaluate is the HIE class II. They can either recover fully, have a severe disability or end up somewhere between the extreme ends. Thus it is important to have additional information that can help the clinician to make

Table 2: Inclusion criteria for hypothermia treatment in NICU, Helsinki University Central Hospital [5]. Modified from TOBY trial protocol [31].

If criteria 1-3 and some of criteria 4 and 5 are filled, aEEG monitoring and possibly the cooling treatment will be started. All neonates with asphyxia will be monitored with aEEG if it is available to use.

1. Gestational age more than 36 weeks
2. No abnormalities that require surgery in neonatal period
3. Age less than six hours at the time of treatment
4. Significant asphyxia, including one of the following criteria:
   - Apgar point less than 6 at 10 minutes after birth
   - Need for endotracheal or mask ventilation, at 10 minutes after birth
   - Acidosis within 60 minutes of birth, defined as any occurrence of pH $<$7.00
   - Base deficit over 16mmol/L in any blood sample within an hour of birth.
5. Moderate to severe encephalopathy including clinical seizures OR all of the following criteria
   - Altered state of consciousness (decreased or missing reactions to stimuli)
   - Hypotonia
   - Abnormal reflexes including abnormal/missing suck or Moro
6. At least 30 minutes of amplitude integrated EEG recording that shows abnormal background aEEG activity or seizures. There must be one of the following:
   - normal background with some seizure activity
   - moderately abnormal activity
   - suppressed activity
   - continuous seizure activity

conclusions of the clinical state and direction of the baby.

The most common outcome prediction method is the usage of HIE classes. It has been shown that most babies with mild HIE gradus belong to the normal group 1 and some to group 2, whereas about half of the babies with moderate or severe HIE belong to the most severe group, group 3. [29, 34] The duration and changes of the HIE gradus have also been associated with different outcome expectations [35]. However, some clinical parameters used in HIE scoring may be difficult to measure soon after birth and are hard to monitor constantly.

HIE gradus and outcome can also be estimated based on anatomical changes. Ultrasound can be used to image the brain structures when the plates forming the skull are not yet grown together. Magnetic resonance imaging (MRI) is also used to detect abnormalities in the neonatal brain. [4] Besides detecting brain lesions, imaging techniques can show different developmental steps that are connected with outcome estimates. For example, the appearance of the posterior limb of the internal capsule, a structure of the white matter, has been associated with the motor outcome [36].

The problem with structural imaging is that some changes may occur progressively and they can easily be underestimated in the neonatal period. Additionally, these methods are difficult to use during the critical period in NICUs. [4, 23].

One of the most popular technique in estimating the severity of HIE is the aEEG. It is a reliable indicator of the current clinical status and outcome prognosis. For example, when the normal aEEG band varies between 10 and 40 μV, in severe HIE the activity band can be significantly narrower and drop below 10 μV [33, 37, 38]. An example of a normal and burst suppressed aEEG trace are shown in Figure 3. aEEG has also been shown to correlate with metabolic changes in the brain [39]. Easy interpretation and usability have made aEEG a standard monitoring method in NICUs but it is necessary to remember that aEEG analysis is not usually quantitativea and can thus vary between different observers and observing times [28].

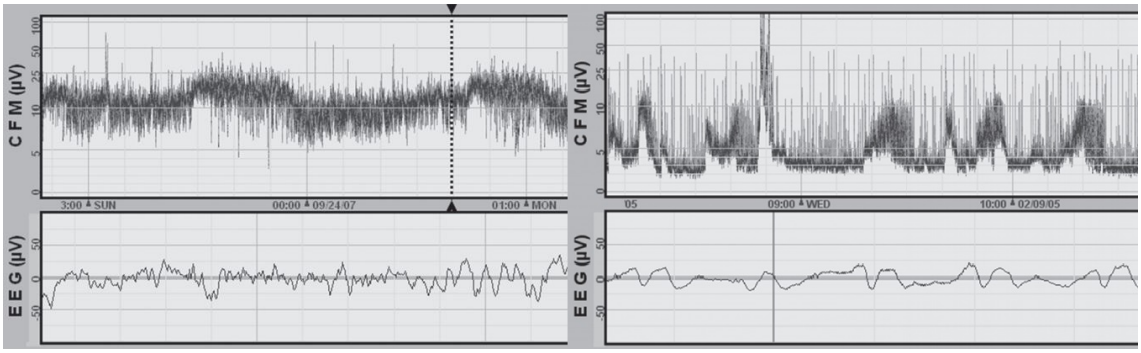


Figure 3: Left figure represents normal aEEG and right figure is an aEEG recording of a paitent with severe HIE with burst suppression and seizures. Upper images show the aEEG trend from a cerebral function monitor, CFM. Signals below show 7-second epochs of raw EEG. [4]

Although aEEG is proved to be a good tool in neonatal intensive care, the most promising method in clinical assessment and prediction is normal EEG. It is able to show smaller changes in the brain activity than aEEG [37]. Traditional way of utilizing EEG measurements has been a visual interpretation of the signal by a specialist. Visual analysis of the amplitude, interhemispheric symmetry, continuity and waveform have been shown to correlate well with the outcomes [37, 38, 40]. Specialist expertise is however not always available and thus there is a need for computational parameters and automatized analysis. Visual patterns also hard to quantify and different studies tend to use varying definitions to determine what is for example considered as low amplitude [28].

Automated systems for distinguishing normal EEG patterns from abnormal ones have been developed. These techniques have concentrated on bursting properties, amplitude variations and frequency content of the signal. For example, increasing amount of slow delta waves [41, 42], duration of burst intervals [43] and low overall amplitude [41, 44] have been associated with poor outcome. These different quantitative methods have shown promising results. Although individually they cannot distinguish all HIE classes, they have good predicting value when used together [41, 45].

3 Research material and methods

This section introduces the used data set and the methodological choices made in this study. The main emphasis is in presenting the selected computational features but also the methods that are used to evaluate and classify the data are discussed.

3.1 Dataset

The dataset used in this study was collected at the Helsinki University Children's Hospital during a four-year period from January 2011 to December 2014. Based on medical reports, newborns who had showed signs of asphyxia were selected for a closer inspection. For the selected 61 newborns an estimation of HIE gradus was then given by a medical doctor.

HIE gradus was estimated for 52 neonates. The status of the remaining nine babies was not possible to estimate or it was diagnosed to be something else than HIE. From the neonates with HIE diagnose, seven were excluded from the study because of missing EEG monitoring data or an unknown outcome. The outcome score was based on the last available evaluation from the routine follow-up clinic. Two neonates were excluded because their gestational age was under 36 weeks. Additionally, one baby was later excluded because of an exceptionally abnormal EEG data. Total number of the subjects was 42.

From the babies used in this study 13 had a mild (class I), 19 had a moderate (class II) and 10 had a severe (class III) HIE gradus. The outcomes at the age of approximately one year were divided to three groups: normal outcome (group 1), mild or moderate abnormalities (group 2), and severe abnormalities or died (group 3). More detailed descriptions of the different abnormalities were presented in Section 2.3.2. 18 of the babies were estimated to belong to group 1, 13 to group 2 and 11 to group 3. Figure 4 shows the respective HIE classes for each outcome group.

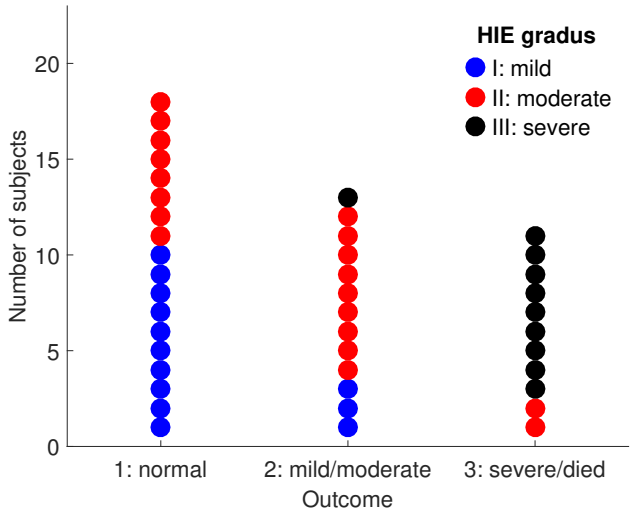


Figure 4: Number of subjects in each outcome group. Colour indicates the estimated HIE class after birth, blue being mild, red moderate and black severe HIE gradus.

EEG was recorded within the first few days after birth, at the age of 70.21 ± 23.87 (mean ± SD) hours. The age distribution is presented in Figure 5 and we can see that the most severe group had the EEG recorded a bit earlier than other subjects. The goal was to get EEG data that is approximately from the age after the hypothermia treatment.

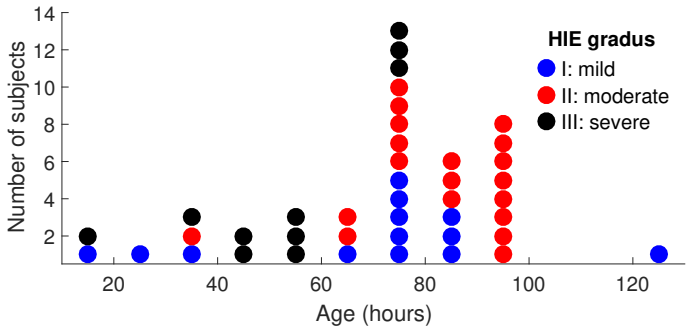


Figure 5: The age of each subject at the time when EEG was recorded. Colour indicates the estimated HIE class after birth, blue being mild, red moderate and black severe HIE gradus.

Data was collected from the two frontal channels, F3 and F4, and from the two parietal channels, P3 and P4, positioned according to the international 10–20 standard. EEG was recorded at a sampling frequency of Fs =250 Hz using NicoletOne EEG amplifier (Cardinal Healthcare/Natus, USA) and 21 channel EEG caps (sintered Ag/AgCl electrodes; Waveguard, ANT-Neuro, Germany). All four channels shared a common reference point at the nasion, which is the point between the eyes, just above the bridge of the nose. The used channels are highlighted in Figure 6.

For this study, 20-minute epochs were extracted from the EEG data. First set of epochs was selected from periods of quiet sleep, which were visually estimated from the EEG signals. An example of quiet sleep recording is presented in Figure 7. Neonatal quiet sleep is characterized by higher voltage and lower frequency compared to active sleep. It also includes bursts of activity separated by more inactive periods.

The second set included epochs from active sleep. Active sleep has lower voltage and higher frequency and it is more continuous than QS as can be seen from Figure 7. For 11 newborns it was impossible to extract AS epochs while there were no clear signs of that sleep stage. From these babies only two came from the HIE class II while others were from the most severe class III. The outcome group was 3 for all others than the HIE class II babies who were estimated to belong in outcome groups 1 and 2. This resulted to a small group sizes in HIE class III and in outcome group 3.

From the 42 subjects 20 were girls and 22 boys, with gestational age of 39.64 ± 1.69 weeks. 26 of the newborns went through hypothermia treatment, during which no EEG could be recorded. Cooling was done for 5 (38%) of grade I babies, 11 (58%) of grade II babies and 10 (100%) of grade III babies. Additionally, information about Apgar points, pH, BE, weight and height at birth, was collected. Information about given medication that can potentially affect the electrical functions of neonatal brain was also included in the research material.

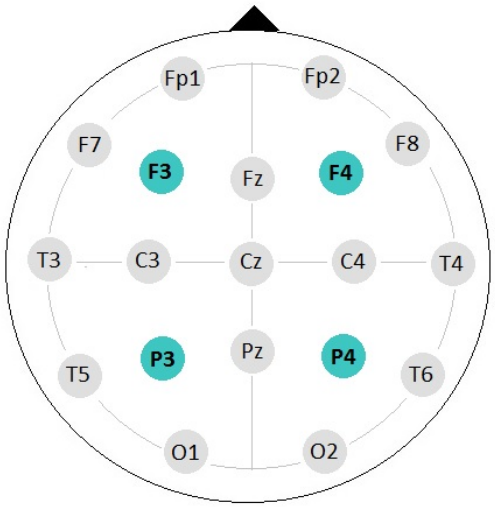


Figure 6: International 10-20 system for EEG recording. Channels that were used in this study are presented in blue. All channels shared a common reference point located at the nasion.

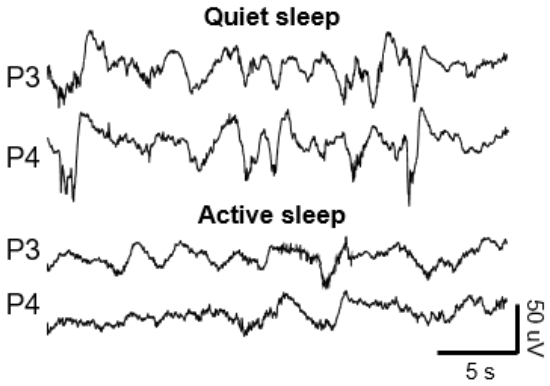


Figure 7: Figure shows an example of neonatal EEG recording from parietal channels during quiet and active sleep stages. In quiet sleep the signal is characterized by higher voltage and lower frequency compared to active sleep. In active sleep the signal is more continuous and the frequency is higher.

3.2 Pre-processing of data

EEG recording is sensitive to external activity and low-pass filtering was thus used to remove artefacts that were caused by movements, sweating and electrical equipment [4]. It is also known that neonatal EEG activity is mostly packed at frequencies below 30 Hz [11, 46]. Because of this, EEG epochs from all subjects were filtered with a low-pass filter with a cut-off frequency of 45 Hz. Because some newborns did not have long-enough periods of quiet or active sleep, 20-minute epoch was collected in several parts that were later concatenated into one continuous epoch.

After the signals were filtered and concatenated, EEG data was down-sampled

from the original 250 Hz to 100 Hz. The selected sampling rate was slightly higher than the Nyquist rate of 90 Hz, which is the minimum sampling rate that avoids aliasing. Nyquist rate is twice the highest frequency component in the sampled EEG, which in this case was 45 Hz after the filtering. The direct current component was also removed from the epochs.

3.3 Computational features

Eight different quantitative feature classes were selected for this study. All of them were considered potentially relevant for this topic, while they give an overall estimation of some aspect of large-scale brain activity and are thought to be stationary. Most of the selected features reflect correlation patters and synchrony. Synchronization is considered to be an important factor in development of brain connections [10, 47]. Communication between different brain regions is essential for complex brain functions. For example, long-range phase synchronization has been connected with attention and local synchronies have been considered to arise from a single sensory modality [48].

Synchrony can be divided to three different correlation categories: amplitude-amplitude correlation, phase-phase correlation and phase-amplitude correlation. Correlation between amplitudes was measured by amplitude envelope correlation (AEC). Phase-phase correlation was estimated with two different measures, phase locking value (PLV) and debiased weighted phase lag index (wPLI). Phase-amplitude correlation was assessed by nestedness coefficient (NC).

Based on clinical observations with neonatal EEG, two computational feature classes were selected. First one, activation synchrony index (ASI), reflects the co-occurrences of activity bursts between the two hemispheres. In a normal neonatal brain, the degree of interhemispheric synchrony is usually high. Second one, multifractal detrended fluctation analysis (MFDFA), estimates the multifractal properties or so called self-similarity of the signals. This measure has been shown to significantly differ between healthy and pathological conditions [49].

Additionally, two feature classes were selected to describe the spectral density of the signal. Spectral density is a more diffuse measure and can reflect well the overall state of the brain. Decreased spectral power has also been showed to correlate with outcome after HIE especially in delta frequency band [50, 51]. Selected features were power spectral density (PSD) and cross power spectral density (CPSD). A summary of selected feature classes is presented in Table 3. Because of multiple frequency bands and channels, each feature class had a set of vectors and the total number of feature vectors in this study was 189.

Neonatal EEG activity takes place at frequencies from near zero to around 30 Hz [11, 46] and this was thus the frequency area of interest. The typical spectrum of neonatal EEG is also presented in Figure 14. The frequency area was split further into four different frequency bands: 0.25–3 Hz, 3–8 Hz, 8–15 Hz and 15–30 Hz. They roughly correspond to delta, theta, alpha and beta frequency ranges, presented in Figure 1.

Table 3: Summary of computational EEG feature classes and their measurement type, used channels and frequency bands. Last column refers to the obtained number of variables for each feature class.

Feature	Type	Channels	Frequency bands (Hz)	N
AEC	Amplitude-amplitude correlation	Between all channels	0.25–3, 3–8, 8–15 and 15–30	24
PLV	Phase-phase correlation	Between all channels	0.25–3, 3–8, 8–15 and 15–30	24
wPLI	Phase-phase correlation	Between all channels	0.25–3, 3–8, 8–15 and 15–30	24
NC	Phase-amplitude correlation	Within each channel	0.2-0.6 against 3–8, 8–15 and 15–30	12
ASI	Interhemispheric synchrony	Interhemispheric	1.5-20	1
MFDFA	Multifractal spectrum	Within each channel	0.25–3, 3–8, 8–15 and 15–30	64
PSD	Spectral density	Within each channel	0.25–3, 3–8, 8–15 and 15–30	16
CPSD	Spectral density	Between all channels	0.25–3, 3–8, 8–15 and 15–30	24

3.3.1 Amplitude envelope correlation (AEC)

Amplitude envelope correlation (AEC) measures the correlation between amplitudes of two signals regardless of the phase [8]. If the signal from one channel has large values simultaneously with a signal from another channel, the AEC value is high. Amplitude–amplitude correlation between signals describes co-modulation of overall neuronal activity levels [52, 53]. These oscillations happen in time scales much longer than those of spike synchronization and phase correlations. The amplitude-amplitude correlations characterize spontaneous resting-state activity and are also closely associated with correlations of slow fluctuations in the blood oxygenation level-dependent signals between brain areas. [53, 54]

At the beginning, all of the four signals were band-pass filtered with one of the chosen frequency ranges: 0.25–3, 3–8, 8–15 or 15–30 Hz. Filtering was done by using a pair of 7th order Butterworth low-pass and high-pass filters. The filters were applied in both forward and backward directions by using a Matlab function `filtfilt` to achieve minimal alteration of the original waveform and a zero lag phase shift [46].

In order to extend signals into the complex plane, Hilbert transform was used. If $f(t)$ represents the original signal, with the help of Hilbert transform we can then produce a two-dimensional representation $S(f,t) = f(t) + j\mathcal{H}(f(t))$ that is located in time–frequency space. This representation is complex-valued: at each

point, it consists of an amplitude value $|S(f,t)|$ and a phase value $\phi(f,t)$. [55] Hilbert transform follows Equation 1, where p.v. refers to Cauchy principal value and τ to time-lag.

$$\mathcal{H}(f(t)) = \frac{1}{\pi}\text{p.v.}\int_{-\infty}^{\infty}\frac{f(\tau)}{t-\tau}d\tau. \tag{1}$$

After filtering and performing Hilbert transform, the signal envelopes were calculated as absolute values. Envelope is an amplitude variation of the signal over time, as presented in Figure 8. Finally the correlation between all signal envelope pairs were calculated using Pearson's linear correlation coefficient

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}, \tag{2}$$

where cov is the covariance of the envelopes X and Y, and σ is the standard deviation of one of the envelopes. Value $\rho_{X,Y} = 0$ means no linear correlation and values $\rho_{X,Y} = \pm 1$ mean a perfect negative or positive correlation. Same procedure was performed for the four frequency bands, which resulted to total number of 24 AEC value vectors.

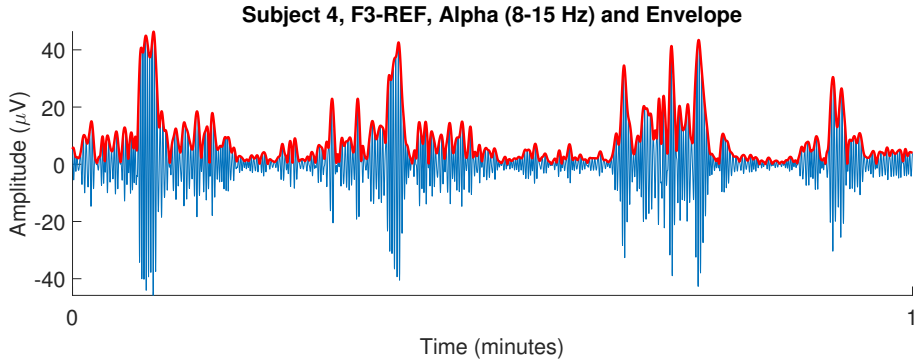


Figure 8: The figure shows a quiet sleep signal (blue) and its envelope (red). Original signal has been band-pass filtered to alpha frequency band 8-15 Hz.

3.3.2 Phase locking value (PLV)

Neuronal activity in large-scale networks shows up in multiple correlation structures. Phase synchrony is the correlation between the phases of the signals and it describes the consistent, non-random phase difference. Phase synchrony can vary from fast spike synchrony to much slower time scales, even hundreds of seconds. [53, 56]. In this research we used two different measures for phase correlation: PLV and wPLI.

PLV estimates the phase relationship between two neuroelectric signals. PLV only uses the phase of the signal and it can thus directly estimate the phase covariance. For example, coherence that is an often used method for estimating the strength of correlation accounts both amplitude and phase simultaneously. [9]

When computing PLV, signals are first filtered in the frequency range of interest with Butterworth filters similarly as described before with AEC. As in AEC, they are also moved to complex plane with Hilbert transformation from Equation 1 [55].

First step is to calculate the phase difference $\theta(t,n) = \phi_1(t,n) - \phi_2(t,n)$ between the two signals. Variables ϕ_1 and ϕ_2 are the instantaneous phases of the two complex time series. Next, the PLV value is calculated as

$$PLV = \frac{1}{N}\left|\sum_{n=1}^{N} exp(i\theta(t,n))\right|, \tag{3}$$

where N is the number of time points in the signal. PLV values range from zero to one, where zero means no correlation and one corresponds to perfect correlation. [9] A graphical illustration of PLV is presented in Figure 9. The same estimation was done for each channel pair and for each of the four frequency bands. This resulted in 24 PLV vectors.

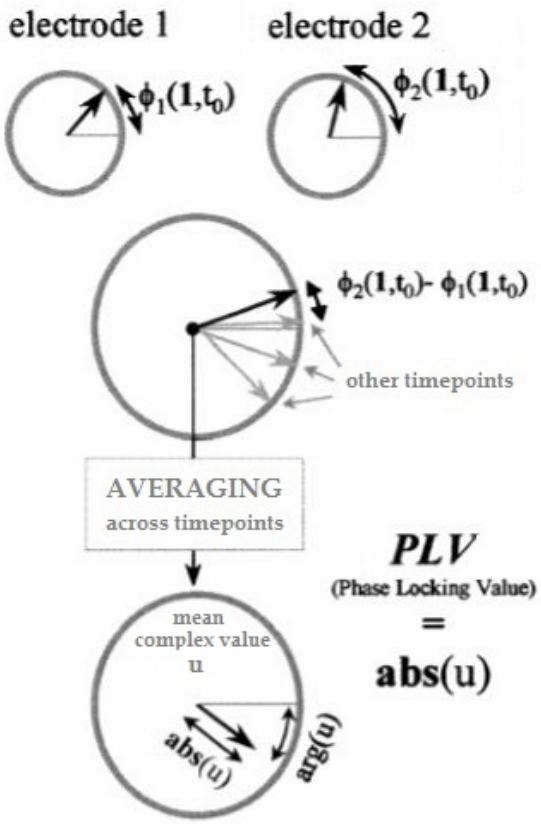


Figure 9: Estimation of PLV. First the instantaneous phase for each electrode is calculated. By averaging over all time points we obtain a complex value u, which amplitude is the phase locking value. Modified from Lachaux and colleagues [9].

PLV can easily be misinterpreted because of background variations. It is also sensitive to volume conduction and noise, because it includes both zero phase lag as

well as nonzero phase lag coupling of the phases. [10, 57]. Because our data contains a common reference point, it can significantly affect the PLV measures.

3.3.3 Weighted phase lag index (wPLI)

wPLI is an improvement of PLV and measures also phase-phase coupling [10]. It is similar to a method called phase lag index (PLI), which was first introduced by Stam et al. [57] with an aim to create an estimate that does not depend on presence of common sources, such as volume conduction or reference electrodes. This is done by excluding phase differences that center around zero or π. PLI is calculated as

$$PLI = |E(\text{sign}(\Im\{X\}))|, \tag{4}$$

where $\Im\{X\}$ is the imaginary part of the cross-spectrum of the signal and E is the expected value operator. If all signals sources would be uncorrelated, the PLI value would be zero. [57] The PLV Equation 3 can be written in same format as $PLV = |E[exp(i\theta)]|$. PLI performs better than PLV in many cases but is still affected by noise. If phase difference value is close to 0 or π, then noise that has somewhat low amplitude, can rotate it across real axis, changing the sign from positive to negative, as presented in Figure 10 [10].

Because of the noise effect, we are more confident on larger values of leading or lagging phase differences. This realization has resulted in wPLI measure [10]. The debiased wPLI weights the phase leads and lags by the magnitude of the imaginary component, as illustrated in Figure 10. Debiased wPLI is computed by first estimating the cross-spectra and its imaginary components as described in Section 3.3.8. After that the average imaginary component of cross-spectra is computed. Finally, the pairwise products of all imaginary components are divided by the pairwise products of magnitudes of all imaginary components. The equation for this process is presented below.

$$wPLI = \frac{\sum_{j=1}^{N} \sum_{k=j+1} W_{j,k} d(X_j, X_k)}{N(N-1)\overline{W}}. \tag{5}$$

In this equation, $W_{j,k} = |\Im\{X_j\}\Im\{X_k\}|$ is the weight, $\overline{W}$ is the weight normalization and $\Im$ is the imaginary part of the signal [10].

Table 4 shows comparison between all phase-phase correlation estimates that were introduced above. In this study, we only used PLV and the debiased form of the wPLI. In conclusion from this table we can say that the PLV in more sensitive to volume conduction effects, but wPLI tends to ignore zero-phase lags correlations even if they are not due to volume conduction. This is because wPLI skips small phase lags while PLV accounts them all. As with PLV, wPLI resulted in 24 value vectors.

Table 4: Summary of performance and behavior of different phase-synchronization indices. Modified from Vinck and colleagues [10].

	PLV	PLI	wPLI
Construction index	Imaginary and real part equally	Consistency of sign of imaginary component	Expected value of imaginary component, normalization by expected value of magnitude
Effect of volume-conduction	Strong increase or decrease	Unaffected	Unaffected
Volume-conducted, uncorrelated noise	Strong increase or decrease	Increase or decrease depending on the distribution relative phase	Decrease if the sign of the imaginary component changes
Change of phase of coherency between sources of interest	Decrease or increase, depending on volume-conduction coefficients	Range statistic is always [0,1], PLI only changes if distribution of sign imaginary component is affected	Range statistic is always [0,1], wPLI only changes if distribution of sign imaginary component is affected
Detecting phase-synchronization	Strong tendency to generate false positives	Even without added noise, PLI may fail to detect phase-synchronization for bimodal/asymmetric relative phase distributions	Steeper relationship with true phase-consistency than PLI even without added noise, always detects non-zero coherence

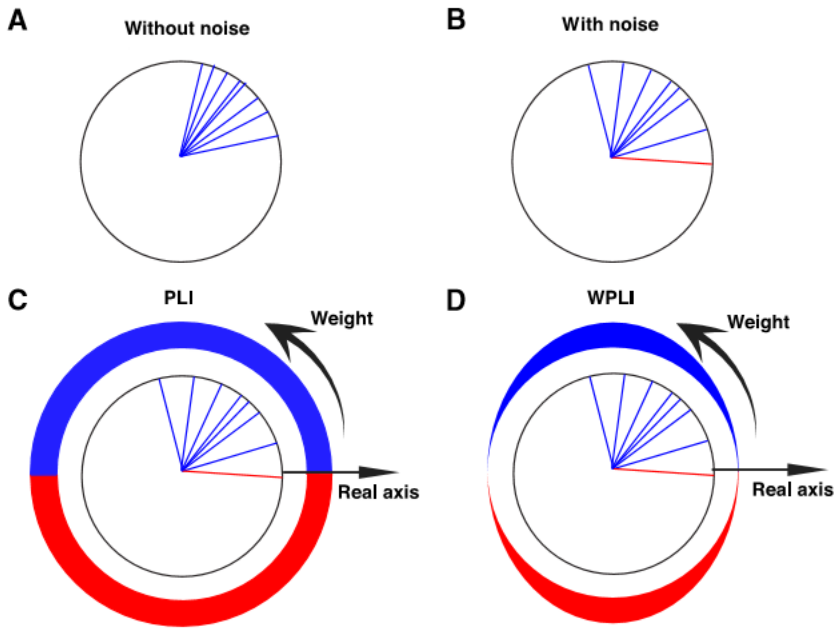


Figure 10: Comparison of PLI and wPLI methods. (A) Phase differences without noise. (B) Phase differences with noise causing the red line to rotate across the real axis. (C) PLI weights all phase differences equally and assigns value of +1 (blue, phase lead) or -1 (red, phase lag) depending on which side of the real axis it lies. (D) wPLI weights phase differences according to the magnitude of the imaginary component. Phase differences around the real axis contribute to a lesser extent than cross spectra around the imaginary axis. Modified from Vinck and colleagues [10].

3.3.4 Nestedness coefficient (NC)

Phase-amplitude coupling is measured by NC. We estimated so called nested oscillations by using phase–amplitude correlation measures. Nested oscillations refer to cross-frequency interactions: amplitude of the fast oscillation is correlated with the phase of the slower oscillation [11]. Nested oscillations are widespread in brain dynamics and may be an essential mechanism for regulating large-scale neuronal activity occurring in several frequency bands [11, 58]. Phase–amplitude correlation may reflect the modulation of the fast oscillations by the slower ones [53]. It can also indicate phase-locking of the slow oscillation to the amplitude peaks of the faster oscillation [46].

NC is proven to be a good method to model cross-frequency interaction during the spontaneous activity bursts in neonatal EEG [11, 59]. It is considered to reflect coordination of spatially overlapping networks with different functionalities [58, 60].

NC is always calculated within one signal and between 2 frequency bands. In this study the lower frequency band was selected to be always 0.2-0.6 Hz, while the faster oscillation was selected to be either 3-8 Hz, 8-15 Hz or 15-30 Hz. With four channels this resulted to a total number of 12 NC value vectors for the used dataset.

The lower component was filtered with a 5th order and the higher component

with a 7th order Butterworth low-pass and high-pass filters. Similarly to AEC, filters were applied in both forward and backward directions. The envelope of the higher component was then extracted and filtered with the same filter that was used for the lower frequency band.

Figure 11 shows an illustration of phase-amplitude coupling. Blue line is the slow oscillation while the red dashed line corresponds to the fast oscillation. Red solid line is the envelope of the fast oscillation. The phase correlation between the two filtered oscillations is calculated with PLV, presented in Equation 3. Before computing PLV, both the filtered envelope and lower oscillation were moved to the complex plane with Hilbert transform in Equation 1.

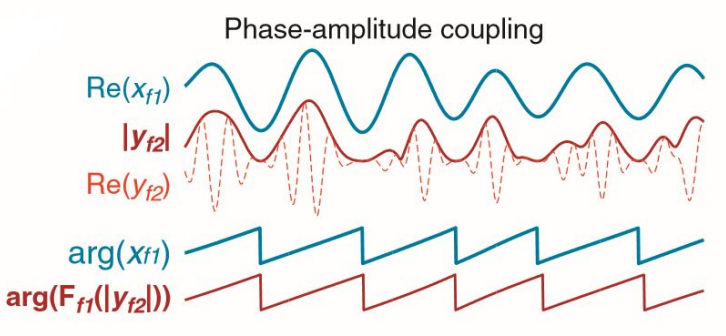


Figure 11: An example of phase-amplitude coupling or nestedness. The phase of the slow and the amplitude of the fast oscillating signal are synchronized. Modified from Palva and Palva [53].

3.3.5 Activation synchrony index (ASI)

ASI measures the interhemispheric synchrony by comparing the co-ocurreces of activity bursts between the two hemispheres, as seen in Figure 12. Interhemispheric synchrony is one of the key components when assessing background activity of the brain. ASI has proved to separate well the normal synchrony from the modest and severe asynchrony [12]. ASI increases within the first few months before birth and can thus also serve as a maturational measure.

In this study we first had to exclude the common reference point. The intrahemispheric signals were extracted from each other: P3 was extracted from F3 and P4 from F4. This resulted to two signals – one from the left and one from the right hemisphere. These signals could then be used to calculate ASI value: the statistical independence between two amplitude envelopes from two hemispheres. Following Koolen and colleagues [61] the two EEG epochs were cut to 2.5-minute parts and ASI was calculated for each part separately. After obtaining ASI values for each part, they were averaged to get one, more stable ASI value.

ASI was calculated by following the work of Räsänen and colleagues [12]. Signals were first down-sampled to 50 Hz from the original 100 Hz. The higher frequencies that are related to the activity bursts were emphasized with a 1st order finite impulse response high-pass filter.

Amplitude envelopes of the filtered signals were obtained by using fast Fourier transformation (FFT). FFT works similarly to Hilbert transform in Equation 1 while it also moves the signal from the original space to frequency domain. FFT was computed by sliding a Hamming window through the signal. The used Hamming window was a bell-shaped curve that had a width of two seconds. It was moved with a step-size of 100 ms through the signal. Amplitude envelopes were obtained by summing those FFT bins that corresponded to the selected frequencies of 1.5-25 Hz.

These envelopes were then quantized to Q different amplitude levels. First, a random subset of the samples was clustered using a standard k-means algorithm and then each sample was assigned to the nearest resulting cluster. As a result, two distinct sequences corresponding to the two input signals A and B were obtained. Quantization was done in order to represent the two signals with sufficiently small number of amplitude levels, that ideally cover the entire scale of amplitudes. The discrete representation also enables the estimation of the joint probabilities of the quantized amplitude values across the two signals. In this study, $Q = 8$ quantization bins were used similarly to Räsänen and colleagues [12], and Koolen and colleagues [61, 62].

Next, an energy-weighted temporal dependency function (ETDF) was calculated between the two quantized envelopes. ETDF measures the statistical dependencies across all possible signal level pairs with different time-lags. These dependencies are also weighted by the relative frequency of each pair. ETDF was calculated with equation

$$ETDF(\tau) = \sum_{a,b} \mathrm{AMP}(a)\mathrm{AMP}(b)\frac{P_\tau(a,b)^2}{P(a)P(b)}, \tag{6}$$

where time-lag τ was set range from -5 to 5 s. $\mathrm{AMP}(a)$ refers to the amplitude corresponding to quantization level a in signal A and respectively $\mathrm{AMP}(b)$ in B. $P_\tau(a,b)$ is the joint probability, which means the probability of observing level a in signal A when second signal B is delayed by τ. $P(a)$ and $P(b)$ are the individual probabilities of the two levels. Resulting ETDF is normalized by extracting the global minimum value from all ETDF values.

Finally the ASI value is calculated by

$$ASI = ETDF_{norm}(\tau = 0)/\Big(\frac{1}{\mathrm{length}(\tau)}\sum_{\tau=-5s}^{5s} ETDF_{norm}(\tau)\Big), \tag{7}$$

where $ETDF_{norm}$ is the normalized ETDF value. Equation 7 gives the ratio between $ETDF_{norm}(\tau = 0)$ and the mean value of $ETDF_{norm}(\tau)$ over the whole time-lag range. This ratio describes the amount of coupling of the signals without a delay when compared to the delayed coupling [62].

Using only two signals (F3-P3 and F4-P4) and one frequency band (2.5-20 Hz) resulted to one ASI value for each subject.

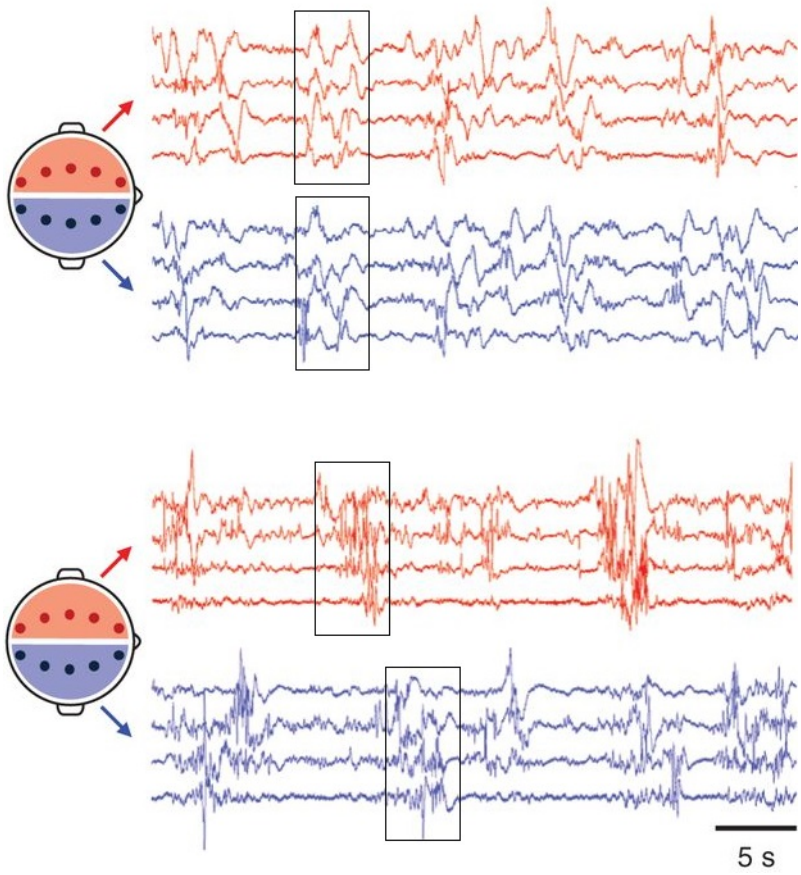


Figure 12: ASI measures interhemispheric synchrony that is seen as a co-occurrence of bursts in both hemispheres (boxes with black outline). The upper example shows high synchrony and high ASI, typical for a normally developing neonate. The example below shows low synchrony and low ASI, typical for an abnormal EEG function. Modified from Videman and colleagues [63].

3.3.6 Multifractal detrended fluctation analysis (MFDFA)

Biomedical signals such as EEG often include fractal structures – some repeating patterns that appear with different intervals and in different scales. These varying fractal structures are found by a measure called MFDFA that estimates the multifractal spectrum of the signal [13]. MFDFA is calculated for each signal separately.

Signals were first filtered to frequency ranges 0.25–3, 3–8, 8–15 or 15–30 Hz, similarly to AEC. After filtering, MFDFA was computed for each of the four signals in five steps that were introduced by Kantelhardt and colleagues [13]. First step was to calculate the sum of the signal in order to obtain signal profile Y:

$$Y(i) = \sum_{k=1}^{i} x_k - \langle x \rangle. \tag{8}$$

In this equation, $\langle x \rangle$ is the mean of the signal and $i \in [1, N]$, N being the number of

data points in the signal.

The second step was to divide the profile Y into N_s segments that do not overlap each other. Segments were defined to have a length of $s = [10, 18, 31, 56, 100]$. The lengths of s were selected by using a logarithmic scale between 10 and 100. N_s was calculated by $N_s = \text{int}(N/s)$ for each value of s.

In the third step, the local trend y_v of each segment was determined. This was done by applying the least-square fit that tries to minimize the difference between each signal point and the fitted model. The model was selected to be a first order polynomial. When the trends were obtained, the variance between the fitting polynomial and the segment was calculated by

$$F^2(s, v) = \frac{1}{s} \sum_{i=1}^{s} (Y[(v-1)s + i] - y_v(i))^2, \tag{9}$$

where $v \in [1, N_s]$ is the segment index and s the window length.

The fourth step was to average over all segments, which resulted to a q^{th} order fluctuation function. Here we used values $q \in [-5, 5]$ with a step size of 0.5, as instructed in paper from Ihlen [64]. Averaging was done with equation

$$F_q(s) = \left\{ \frac{1}{2N_s} \sum_{v=1}^{2N_s} (F^2(s, v))^{q/2} \right\}^{1/q}. \tag{10}$$

The fifth step was to determine the scaling behaviour of each fluctuation function $F_q(s)$. This was done by analysing plots where $F_q(s)$ is plotted against s in a logarithmic scale. The slope of this log-log plot is the Hurst exponent H and function $h(q)$ is called generalized Hurst exponent [13]. With $h(q)$ is it also possible to calculate the q-order singularity dimension:

$$D(q) = \frac{qh(q) - 1}{q - 1}. \tag{11}$$

$D(q)$ plotted against $h(q)$ is the so called multifractal spectrum. From this spectrum, four variables were extracted: *peak hq, width hq, tail Dq* and *heiqht Dq*. First three metrics are more conventional [13, 64] and the last one has been shown to correlate well with abnormal EEG [65]. Selected variables are presented in Figure 13. All four variables were calculated for each channel and four frequency ranges, which resulted to 16 value vectors for each variable and thus to 64 different MFDFA vectors.

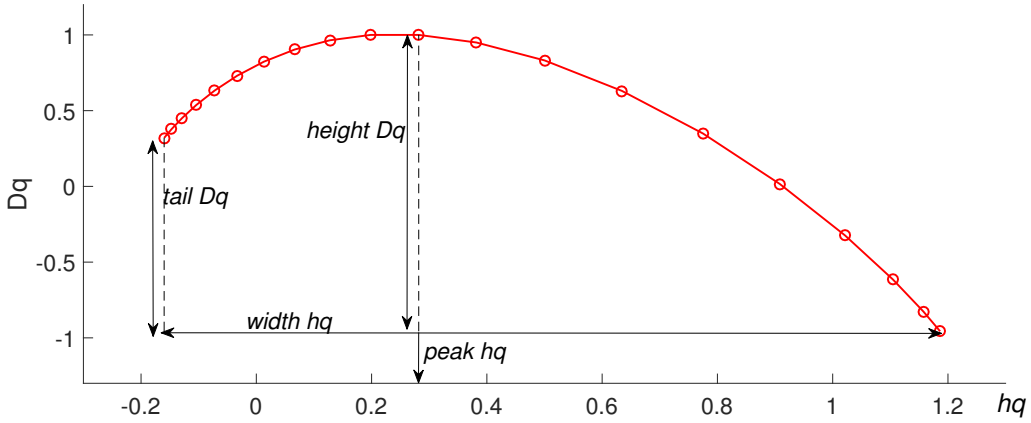


Figure 13: An example of the MFDFA multifractal spectrum, which is a plot of *D(q)* against *h(q)*. The used variables were *peak hq, width hq* and *tail Dq* that all describe the multifractal spectrum.

3.3.7 Power spectral density (PSD)

Spectral density describes the distribution of the signal power into different frequency components. EEG attenuation has been associated with severe HIE [66, 67] and this can be detected with spectral density analysis.

Spectral density was calculated with Welch's power spectral density estimate. Compared to other methods, Welch's method has an advantage of minimizing the estimate variance. [68] Spectral density was computed for each channel individually for each of the four frequency bands. A Matlab function called `pwelch` was used to calculate the Welch's estimate.

Signal was first divided to a maximum of eight sections that are as long as possible and have a 50% overlap. After dividing the signal to overlapping segments, each section was moved to the time-frequency space with a sliding Hamming window. The obtained estimates of the frequency power were then averaged.

This procedure resulted in an single-sided periodogram, a frequency-power presentation that had a bin-size of 1 Hz. In single-sided estimate the power is only accounted for positive frequencies. An example of a periodogram for one subject and for one channel is shown in Figure 14. To estimate PSD for the wanted frequency ranges, the bins within the wanted ranges were averaged.

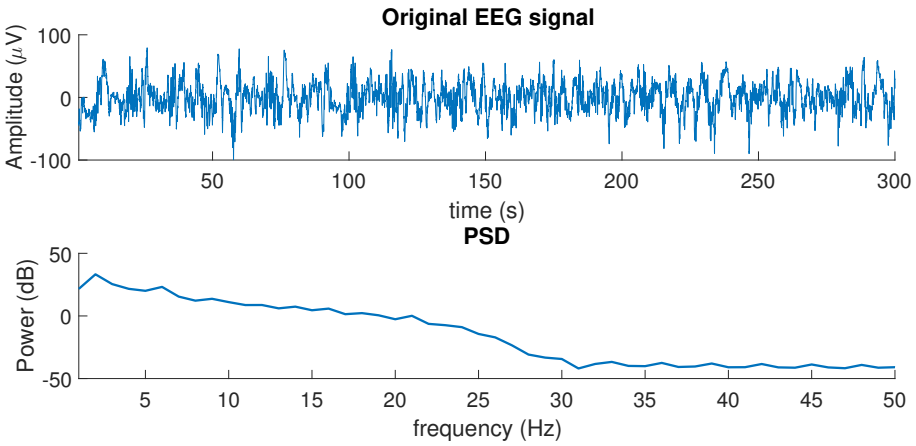


Figure 14: Power spectral density for one subject and for one channel. The power is strongest with small frequencies and then decreases significantly after 30 Hz, which supports the decision to use frequency range 0.25-30 Hz in this study.

3.3.8 Cross power spectral density (CPSD)

While PSD is calculated within each channel individually, cross power spectral density (CPSD) is measured between each channel pair. CPSD can be interpreted as a Fourier transform of a cross-correlation between two signals. Cross-correlation measures the similarity of two signals as a function of time-lag between them so CPSD provides information on the power shared by a given frequency for the two signals. Signals were filtered with the same method as used with AEC for the four frequency band. Cross correlation between signals X and Y was then calculated with

$$R_{X,Y}(\tau) = E[X(t+\tau)Y(t)], \tag{12}$$

where E is the expected value operator and $\tau \in [0, N]$ is the time delay. CPSD for frequency f was then defined by taking the FFT from $R_{X,Y}(\tau)$.

CPSD was calculated using Matlab function `cpsd`. Four frequency bands and four channels resulted in 24 CPSD values for each subject.

3.4 Evaluation and classification methods

This section introduces the tools that are used to analyse the computed EEG features. First part of this section introduces the Wilcoxon rank tests. Wilcoxon rank sum test was used to compare the feature values between the three HIE classes and the three outcome groups. Wilcoxon signed rank test compares paired values and it was used to compare the two sleep stages. The last three parts in this section are connected to the classification of the outcomes. We first discuss the analysis of feature stability, then methods used for feature selection and finally we present the used machine learning method.

3.4.1 Wilcoxon rank tests

To determine which of the computed features vary significantly between the HIE or outcome groups, a statistical test called Wilcoxon rank sum test was used.

Wilcoxon rank sum test is used to determine whether two samples come from the same distribution or not. More specifically, it uses a null hypothesis that the medians of the two samples are the same. It is a nonparametric alternative to the t-test of two samples. By using the ranks instead of actual values, the calculations are simplified and there is no need to make assumptions of the normality of the distribution. Ranking is done by ordering all samples from both groups together and then assigning a rank for each value based on its position: the smallest value is replaced by 1, second smallest by 2 and so on. If two or more observation have the same value, they are replaced by the mean of their supposed rankings. [69, 70]

In Wilcoxon rank sum test, the calculated test statistic parameter is the sum of ranked values of the first group, W. For large enough sample sizes, Wilcoxon rank sum test uses z-distribution to test the null hypothesis, equality of the medians. The z-score is calculated with the expected value $E(W)$ and the variation $V(W)$ of rank sum W with formula

$$z = \frac{W - E(W)}{\sqrt{V(W)}}. \tag{13}$$

After calculating this z-score, it needs to be compared to the z-value table in order to determine the p-value of the null hypothesis. When assuming that null hypothesis is true, the p-value is the probability of observing at least as extreme value as the observed value of the test statistic. The selected significance level was 0.05.

In order to compare the features from quiet and active sleep epochs, Wilcoxon signed rank test was used. The main difference to the Wilcoxon rank sum test is that instead of two separate groups it compares two related samples [70]. When considering features that were calculated from AS and QS epochs, we always have pairs: same feature from the same baby calculated both from QS and from AS.

In the Wilcoxon signed rank test we calculate the difference $d = |x_1 - x_2|$ for each measurement pair. The differences are the ranked similarly to the Wilcoxon rank sum test. The test statistic W is then calculated as the sum of the signed ranks. The

null hypothesis is that the differences between pairs follow a symmetric distribution with a median of zero. Under the null hypothesis, the z is defined with W and its variance $V(W)$ by

$$z = \frac{W}{\sqrt{V(W)}}, V(W) = \frac{N(N+1)(2N+1)}{6}. \tag{14}$$

After calculating the z-value, it is again compared to the z-distribution and a p-value is obtained. [70]

3.4.2 Stability

If the results of this study were to be used in clinical practice, the used EEG features would have to be stable. Stability in this context means that the feature values should not vary too much between different epochs and epoch lengths. If the variation is too great, the feature cannot be used in clinical applications while it is too unreliable.

To test the stability of the used computational features, they were computed individually for epoch lengths of 5, 10 and 20 minutes. After the feature set for each epoch length was computed, the relative variance was studied by

$$D = \frac{\sigma^2}{\mu}, \tag{15}$$

where variance σ^2 was the variance between the 5, 10 and 20 minutes values and mean μ the mean of the same three values for each subject individually. The absolute values of the subject specific relative variances were then averaged to get the feature specific estimation of the relative variance.

Because the epochs were obtained from both quiet and active sleep periods, the relative variances were calculated for all the 189 feature vectors in both of these sleep stages individually. The relative variances from the two sleep stages were also compared to each other in order to estimate possible differences in stability.

3.4.3 Feature selection

Only quiet sleep EEG data was used for the automatic analysis of the outcome. This was done because we could not extract active sleep epochs for most of the newborns from the outcome group 3. Using all the 189 feature vectors from QS to estimate the outcome with a machine learning algorithm would not make sense as there are only 42 subjects. Selecting all the parameters would result to overfitting while there would be too many parameters relative to the number of observations. Thus it was decided to limit the number of used features to three. This decision was made without any quantitative analysis, while the goal of the classification was to perform an exploratory preliminary analysis, not the best possible classification.

The goal of the algorithm was to find the most normal cases (group 1) and the most severe cases (group 3). The performance of the classifier was estimated with sensitivity and specificity. Sensitivity measures the proportion of those babies that are correctly identified as belonging to the first group. Specificity describes the

proportion of babies that are correctly classified to the other remaining group. The exact formulas are given later in Equation 17.

All features that had a relative variance over 0.2 were excluded while they were thought to be unstable. The exact limit was decided based on the inspection of relative variances: about 10% (22) of the features were above this limit while 90% (167) were below it. In order to select the three features, three different approaches were used: forward selection, brute force and Wilcoxon rank sum test's p-value based selection.

In forward selection the features are added one by one to the model. The first parameter was set to be the one with the lowest p-value obtained from Wilcoxon rank sum test using Equation 13. After that each feature was added to the model separately and the one that gave the best prediction sensitivity together with the first parameter was selected. Same procedure was repeated by testing all remaining features together with the two already selected ones. If two or more parameters gave the same sensitivity the one with lower p-value was selected.

Brute force selection refers to testing all possible combinations. In this method all sets of three variables were tested. After excluding the unstable features the total number of features was 167, which resulted to a total number of 762355 sets. The brute force method is time consuming but usually more accurate than the forward selection.

Third approach was to select the three features based solely on the Wilcoxon rank sum test p-values and visual inspection of the features' distribution. Three variables with the smallest p-values may not describe the dataset very well, while they can all be from the same feature. To get a variable set that represents the overall state of the brain, this approach was limited so that there would be only one variable selected from one EEG feature.

3.4.4 Support vector machine (SVM)

SVM, first introduced by Cortes and Vapnik [71], is a machine learning algorithm that is used for supervised classification. It was chosen to be used in this study because it showed good results in initial testing, is easy to modify to different shapes of data and works quite well with small sample sizes. SVM is also intuitive and theoretically well motivated.

SVM attempts to assign binary group labels to data points that are located in some feature space. This is done by finding one or more hyperplanes that separate the two groups from each other. Hyperplanes are selected so that the gap between the two groups is as wide as possible. Data points closest to the hyperplane form so called support vectors. Because hyperplane separates two classes we obtain two support vectors: one for each class. Hyperplane and support vectors are illustrated in Figure 15.

Data points are not always separable by a linear hyperplane. When this is the case, hyperplane can be selected so that it separates many, but not all points: some of the points are classified in the wrong group. Another approach is to use so called Kernel trick. Kernel is a function that maps the data points from the original space

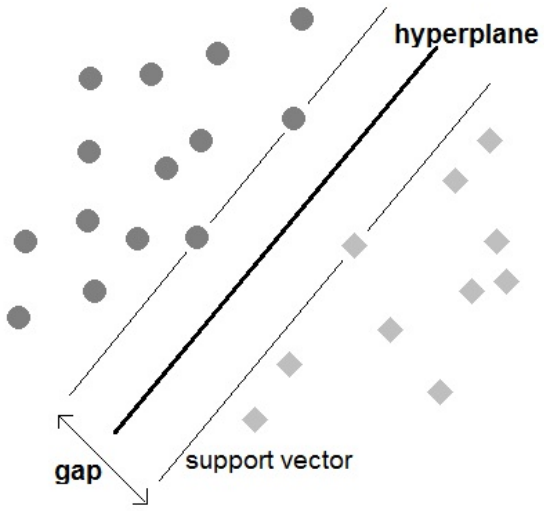


Figure 15: Two dimensional illustration of support vector machine classification. Hyperplane separates the two classes marked with circular and square items with maximum gap. Data points closest to the hyperplane form the support vectors.

to a new, higher-dimensional space. Goal is that the two classes are then linearly separable in the new space. Kernel function can be defined as a dot product in some feature space, following

$$K(x, y) =< \phi(x), \phi(y) > . \tag{16}$$

Separating hyperplane was expected to be linear, which means that no Kernel trick was applied. Because SVM only works with binary classification, it was decided to divide the original question, separating three outcome classes from each other, to two parts: first combining outcome groups 1 and 2, then groups 2 and 3. In first case the sensitivity measured how well could we detect outcome group 1 babies from the other two groups. In the second case the sensitivity then measured how well could we detect outcome group 3 babies from the other two groups.

Before applying the SVM learner, data was divided to train and test sets with a 6-fold cross validation. This fold size was selected while it is computationally less expensive and the bigger fold size is said to be less biased [72]. Because there were 42 subjects, it meant having 7 subjects for testing and 35 for training. Disjoint partition was done so that each subsample had roughly the same class proportions. In forward selection and brute force methods we used also a 5-fold cross validation for the test set. The test set had 35 subjects so it was further divided to trainings set of 30 subjects and testing set of 5 subjects.

Because outcome group 3 was significantly smaller than the combination of groups 1 and 2, we used ADASYN algorithm [73] to create more data points in the subsamples. ADASYN synthesizes new data points to the minority class using a weighted distribution. Using this algorithm, the bias caused by the different group sizes should be avoided. [73]

The SVM classification score for classifying observation x is the signed distance from x to the decision boundary ranging. A positive score for a class indicates that x is predicted to be in that class, a negative score indicates that x belongs to the other class. From the these scores we can then construct a confusion matrix, presented in Figure 5. From the confusion matrix we can then calculate the sensitivity and specificity with the following equations. Sensitivity refers to the true positive rate and specificity to the true negative rate so these two measures give a good overview of the classifier performance.

$$\begin{aligned} sensitivity &= \frac{\text{True Positives}}{\text{True Positives + False Negatives}} \\ specificity &= \frac{\text{True Negatives}}{\text{True Negatives + False Positives}} \end{aligned} \tag{17}$$

Table 5: Confusion matrix is used to estimate classifier performance.

		Predicted group		
		p	**n**	**total**
True group	**p′**	True Positive	False Negative	P′
	n′	False Positive	True Negative	N′
	total	P	N	

4 Results

All the obtained results are presented in this section. The first four parts show the results of the research questions that were introduced in the beginning of this work. The last part of this section discusses the relationship between the features and how they correlate with each other.

4.1 Quiet and active sleep epochs

Brain activity of a newborn is dependent on the sleep-wake cycle. In this study, both active and quiet sleep epochs were extracted for each baby. To determine whether or not the computational EEG features differ in these two sleep stages, we compared each EEG variable obtained from QS and AS epochs.

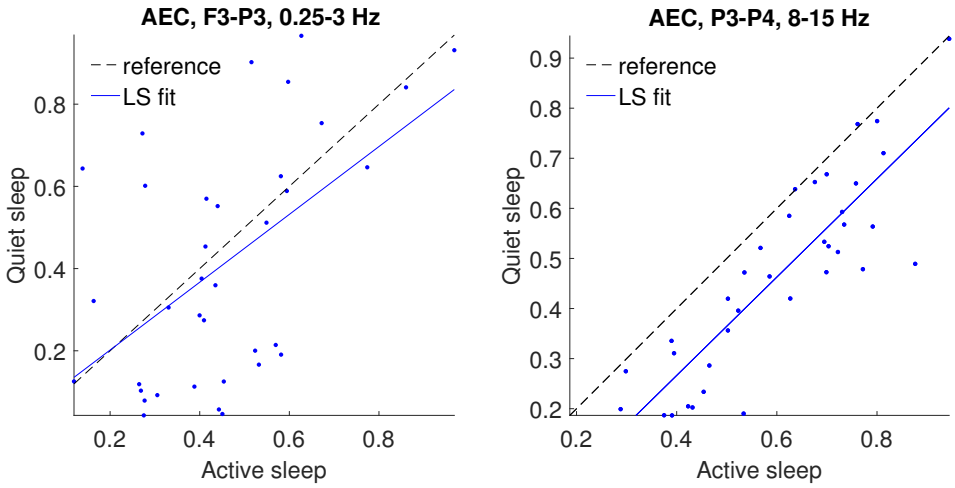
In order to draw conclusions of the similarities or differences between QS and AS, the correlation between QS and AS was calculated using Pearson's linear correlation coefficient described in Equation 2. Because linear correlation can sometimes be misleading if there is a systematic difference, also the Wilcoxon signed rank test was applied. All variables and their statistical test results are presented in Appendix A. Additionally, all variables were plotted and inspected visually. A least square fit was added to each figure to show the direction of the correlation.

In AEC the correlation coefficient was statistically significant with all variables. The coefficients were between 0.48 and 0.86 so the dependence between QS and AS values was clearly positive. The sign rank sum test rejected the null hypothesis with all frequency bands except with 0.25-3 Hz. Rejection meant that the difference between values in AS and QS does not come from a distribution with zero median. This result can be seen visually from Figure 16. From the visual inspection it was clear that the QS AEC values were usually lower than the one from AS, as in Figure 16b.

Also in PLV all the variables showed significant correlation. The Pearson's correlation coefficients were between 0.68 and 0.94 – even closer to one than in AEC. The strong correlation can be observed in Figure 17a. Wilcoxon sign rank test showed only six variables where the null hypothesis could be rejected so most of the variables can be concluded to give very similar results in QS and in AS.

The variables in wPLI showed almost no correlation between quiet and active sleep. Correlation was significant only in two variables and the correlation coefficient was only around 0.4 with these two. Wilcoxon signed rank test was rejected in only two cases and neither QS or AS gave constantly lower or greater values than the other. The only trend seemed to be that if there was a high wPLI value in QS, AS value was then low. This also applied the other way around: if there was a high wPLI value in AS, then the QS value was low. This trend can be seen in Figure 17b.

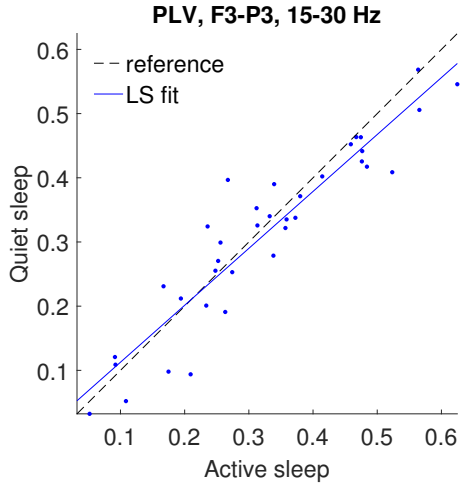
NC had more correlation in the variables from parietal channels. The correlation coefficients were moderate varying between 0.31 and 0.59. The signed rank test showed that the null hypothesis could be rejected in all but two cases. QS values were with the rejected variables significantly lower than the ones in AS. This can be seen also in the Figure 17c.



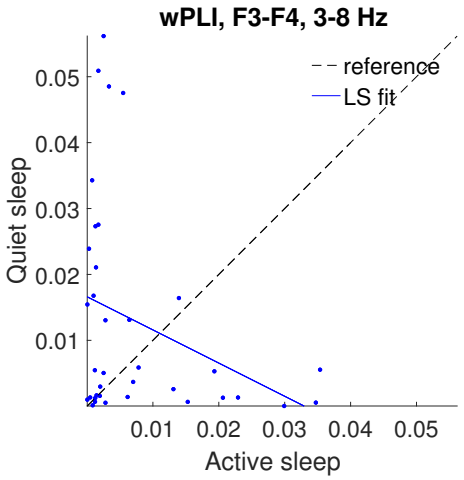
(a) Variable has correlation of 0.5 with p<.001 and the null hypothesis of the Wilcoxon signed rank test holds.

(b) Variable has correlation of 0.86 with p<.001 and the null hypothesis of the Wilcoxon signed rank test is rejected.

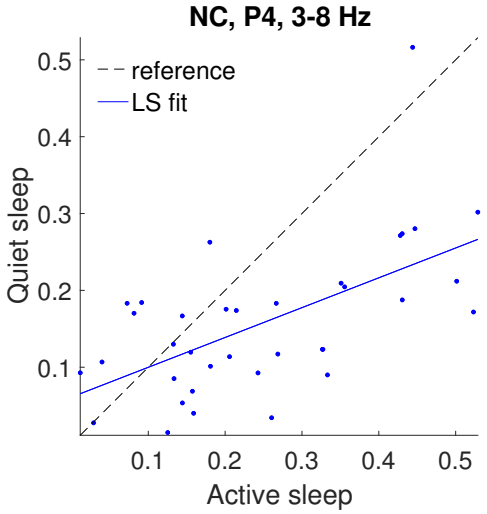
Figure 16: Two AEC variables show the effect of signed rank test. In the left plot the difference between QS and AS values follow a distribution with a zero median. In the right plot the null hypothesis is rejected and we can see that the QS values are lower than AS values.



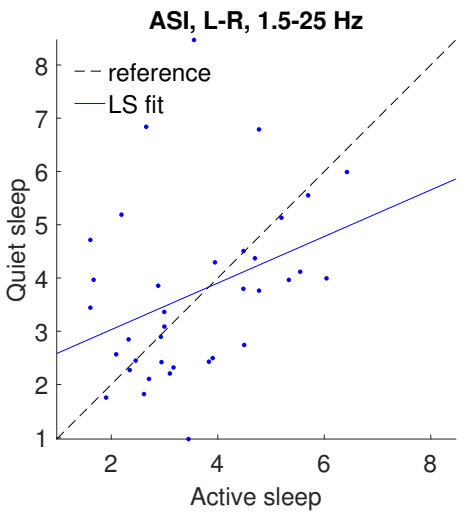
(a) PLV values show strong correlation with coefficient 0.93, p<.001.



(b) Values of wPLI show no significant correlation.



(c) Significant correlation of 0.59, p<.001. QS values are significantly lower than AS values.



(d) ASI values have a correlation of 0.36, p<.05. The signed rank test null hypothesis holds.

Figure 17: Examples of quiet and active sleep correlations.

ASI had a statistically significant correlation with a correlation coefficient of 0.030. The null hypothesis could not be rejected in Wilcoxon signed rank test. The ASI correlation between QS and AS is presented in Figure 17d.

From the four MFDFA features the *peak hq* showed the most significant correlation between QS and AS values. These correlations were all in the higher frequency bands. *Peak hq* also showed many rejected null hypothesis in signed rank test. In all rejected cases the QS values were significantly lower than the AS values, as can be seen in Figure 18a. In the other three features *width hq*, *tail Dq* and *height Dq*, there were almost no significant correlation as seen in Figure 18b. The signed rank test showed few cases from frequency band 3-8 Hz and 8-15 Hz where the QS and AS followed

different distributions. There was no clear direction between QS and AS in *width hq* and *tail Dq* but in *height Dq* the QS values were in all significant cases higher than the AS values.

In PSD the values of QS and AS correlated well and there were only three variables without significant correlation. The correlation coefficients were between 0.39 and 0.79. In the signed rank test all variables from frequency bands 3-8 Hz and 8-15 Hz were rejected. In these variables the QS values were lower than the AS values which is visible in Figure 18c.

All variables in CPSD showed significant correlation between active and quiet sleep. The correlation coefficients were high, varying between 0.57 and 0.98. Again all the variables from frequency bands 3-8 Hz and 8-15 Hz showed p-values under .05 in signed rank test. Figure 18d shows how the QS values are lower than the AS values, similarly to PSD.

The relative variances of the computational EEG features were also compared in QS and AS. Most of the features had a low variance between the different epoch lengths but MFDFA variable *tail Dq*, PSD and CPSD had clearly greater variances. For these features the variance were equally large in both AS and QS. The inspection of relative variances did not provide any evidence of systematic stability difference between the sleep stages.

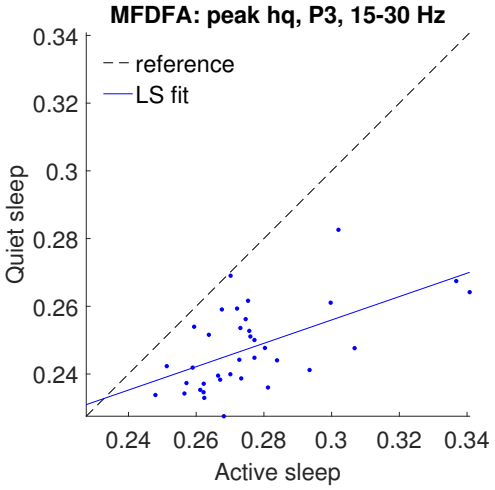
As a conclusion we can say that the simpler methods, such as AEC, PLV and PSD tended to have a good correlation in active and quiet sleep. More complex features like wPLI and MFDFA had often poor correlation. There was also a clear effect of the frequency bands: for example, the lowest band of 0.25-3 Hz often showed less correlation than the other bands. It was also noticeable that many features showed lower constantly lower or higher values in AS than in QS. Because of the big differences between the two sleep stages, it made sense to evaluate the research question two and three individually for both stages.

4.2 Feature correlation with the HIE gradus

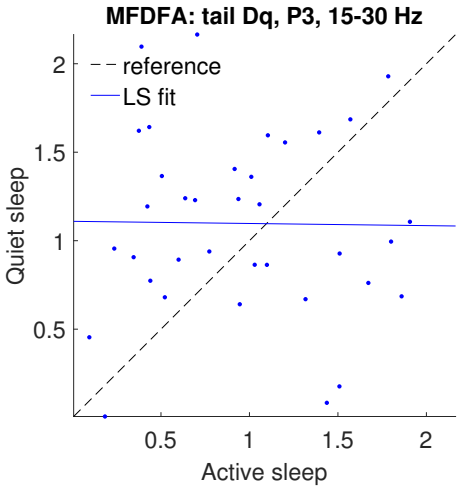
The second research question was about the relationship between the quantitative EEG variables and clinical state of the baby. Clinical condition was considered to be best reflected by the HIE gradus of the baby. There were three separate HIE classes: mild (class I), moderate (class II) and severe (class III). The class for each neonate was estimated by a medical doctor as described previously.

Each variable was treated as an individual measurement. They are summarized in Table 3. All variables were plotted against the HIE classes with box plots and were then examined visually. Wilcoxon rank sum test was used in order to see if the classes differed from each other with a statistical significance. This resulted to an individual p-value between each HIE gradus pair. For quiet sleep, all the statistically significant p-values are listed in Appendix B. For active sleep the results can be found from Appendix C.

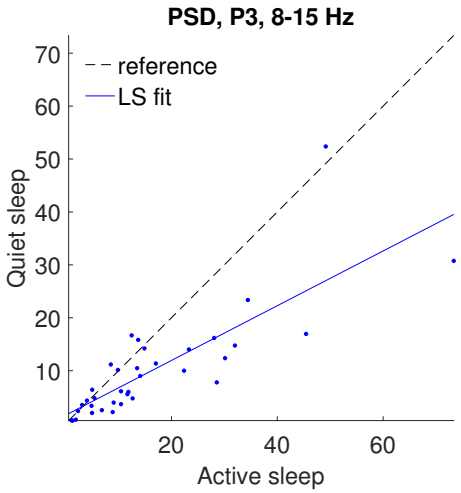
AEC variables showed some significant correlations mostly between HIE classes I and II. In QS the significant differences were mostly in lower frequencies but in AS they were visible in all frequency bands. Additionally, the QS epochs showed one



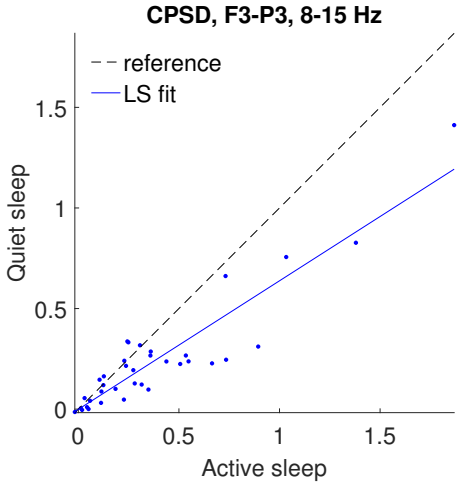
(a) Plot shows a correlation of 0.57, p<.001. QS values are significantly lower when compared to AS.



(b) No significant correlation. The null hypothesis of the Wilcoxon signed rank test holds.



(c) Strong correlation of 0.92, p<.001. QS values are significantly lower than AS values.



(d) Significant correlation of 0.81, p<.001. QS values are significantly lower than AS values.

Figure 18: Examples of quiet and active sleep correlations.

significant result between classes I and III, and two significant difference between classes II and III. All statistically significant differences between AEC distributions were to the same direction: the more severe HIE class, the lower the AEC value. One example is presented in Figure 20a. Although the significant results were quite well distributed between the channel pairs, the parietal pair P3-P4 did not show any results.

PLV showed more significant differences in AS than in QS epochs. All but one of them were between HIE classes I and II, and in the two lower frequency band. The one other significant difference was in QS frequency band 15-30 Hz between classes II and III. All the statistical significances showed similar trend: the more severe HIE, the lower the PLV value. This can be seen in Figure 20b. As with the AEC variables, there were no significant results between channel pair P3-P4. It is noticeable however that this pair showed the biggest PLV values.

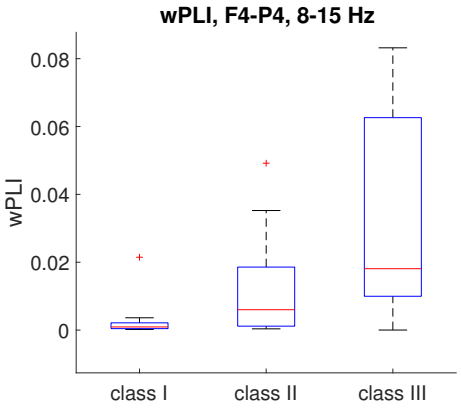
Variables in wPLI showed more significant results in QS than in AS. All frequency bands and HIE class pairs were present among the significant results, also all channel pairs expect the left hemisphere pair F3-P3. Although there were many significant results especially in QS, they were not coherent: some showed that the wPLI value was larger in the more severe HIE classes, some had it the other way around. Based on visual analysis it would seem like the former case was more dominant trend in QS and one example of this kind of wPLI distribution can be seen in Figure 19a. In AS there was no clear trend. All in all, the smaller frequencies had larger wPLI values than higher ones and again the parietal channel pair showed the larger values than the other pairs.

With NC there was only one significant difference that is shown in Figure 20c. It was from AS epoch in frequency band of 3-8 Hz and from channel P4, showing significant difference between classes I and II. It would seem that the more severe HIE class, the more nestedness in active sleep. Lower frequencies showed slightly larger NC values than higher frequencies in both QS and AS.

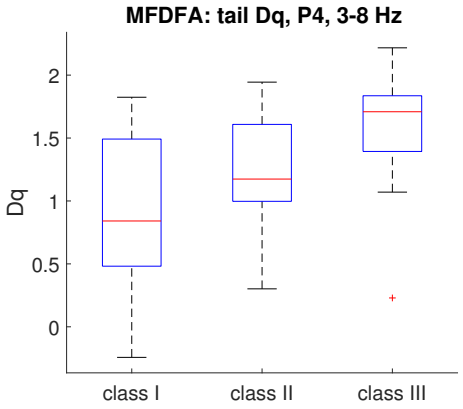
ASI values did not differ significantly between any of the HIE classes. There were also no clear difference between active and quiet sleep variables.

MFDFA had four different measures describing the multifractal spectrum: *peak hq*, *width hq, tail Dq* and *height Dq*. *Peak hq* showed some significant results in frequency band 3-8 Hz between classes I and II and twice between II and III. These variables showed that the *peak hq* was smaller for class II than class I but class III had then larger values than class II. QS showed more significant results than AS and in both cases the *peak hq* values were larger in smaller frequency bands. *Width hq* had significant results also in the higher frequency bands. They all followed the same trend: the more severe HIE class, the larger the *width hq* value. Again QS showed more results than AS.

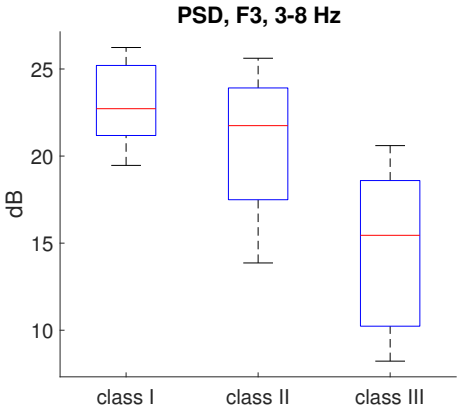
Tail Dq showed differences especially in higher frequencies and in QS. There were significant differences between all classes and all of them followed the same direction: the more severe class, the larger the *tail Dq* value. One example can be seen in Figure 19b. *Height Dq* showed only one significant result in the AS and it was between classes I and II. In QS there were significant differences only between classes I & III and II & III. Only one of these statistically significant differences was



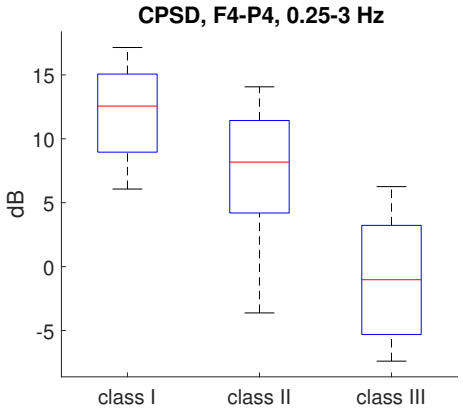
(a) Significant difference between HIE classes I-II (p<.01) and I-III (p<.01).



(b) Significant difference between classes I-III (p<.05) and II-III (p<.05).

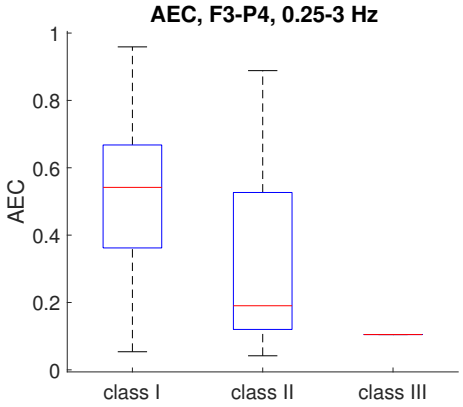


(c) Significant difference between classes I-III (p<.001) and II-III (p<.01).

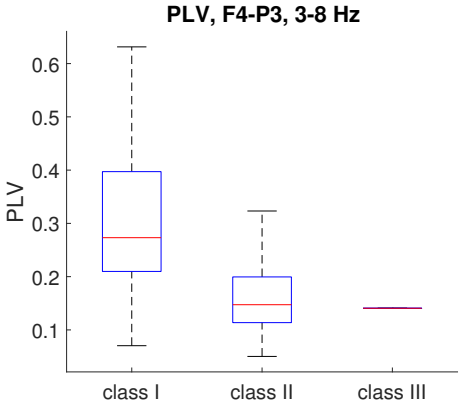


(d) Significant difference between all classes: I-II (p<.05), I-III (p<.001) and II-III (p<.001).

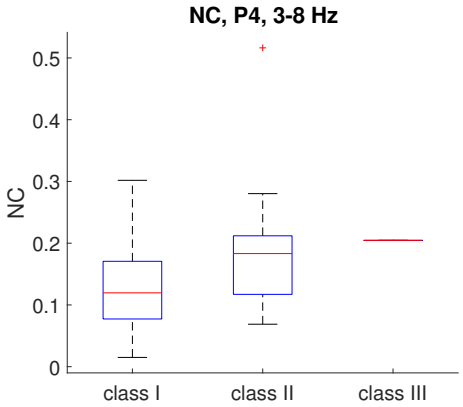
Figure 19: Boxplots present some QS feature values divided to the three HIE classes.



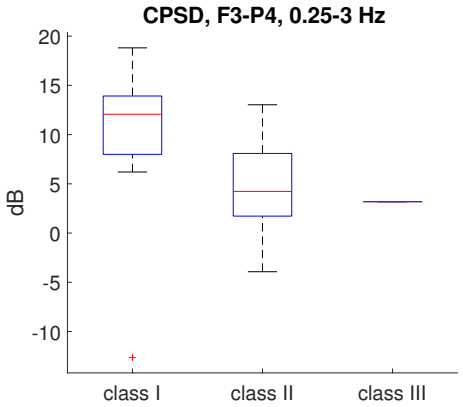
(a) Significant difference between classes I-II (p<.05).



(b) Significant difference between classes I-II (p<.01).



(c) Significant difference between classes I-II (p<.05).



(d) Significant difference between classes I-II (p<.05).

Figure 20: Boxplots present some AS feature values divided to the three HIE classes.

from the lowest frequency range and it showed that the class III was lower than class II. All other *height Dq* variables showed larger values for the more severe classes.

All variables in PSD showed that the more severe is the HIE class, the lower is the magnitude of PSD. It was also noticeable that the the higher the frequency band was, the lower were the PSD values. While in QS there were many significant differences especially between class III and the other two, there were no significant differences in AS. Figure 19c shows one of the significant variables.

The last feature CPSD followed the same trends as PSD: more power in smaller frequencies and the more severe the class, the lower is the magnitude. In QS all variables showed statistically significant differences between HIE classes I and III, most between II and III, and some between I and II. Some even showed a significant difference between all classes as can be seen in Figure 19d. In AS there were only some significant results between classes I and II.

4.3 Feature correlation with the outcome

The third research question aimed to shed light on the relationship between the EEG variables and the clinical outcome of the baby. Outcome was evaluated by a medical doctor on a scale of normal (group 1), mild or moderate (group 2), and severe or died (group 3). Because both HIE class and outcome group describe the severity of the brain injury, it could be expected that the features change similarly between the three groups.

Variables and methods were similar to the previous research question. All variables were plotted against the outcome groups with boxplots and were then examined visually. Statistically significant differences were studied with Wilcoxon ranks sum test, which resulted to an individual p-value between each outcome group pair. The statistically significant results are listed in Appendices B and C next to the HIE class results.

AEC variables did not differ significantly between any outcome groups in AS. In QS there was one significant result between groups 1 and 2 in frequency band of 8-15 Hz. This difference was to the opposite direction than with HIE classes: the more severe case, the higher is the AEC value.

Similarly to AEC, also PLV was unable to separate any outcome group with AS variables. There was one significant variable in QS in frequency range of 0.25-3 Hz. However, here was no linear trend while group 1 and group 2 both had significantly higher PLV values than group 2. When considering HIE classes, the PLV was lower in the more severe cases.

Variables in wPLI showed more significant differences in quiet than in active sleep. As with HIE classes, all channel pairs expect the left hemisphere pair F3-P3 showed significant results. Results were distributed to all frequency ranges and outcome pairs. With HIE classes the differences did not show any trend and that was the case also between the outcome groups in AS. In QS the significant results were all to the same direction: the more severe the outcome, the larger wPLI values. One example is presented in Figure 21a.

NC values had significant differences both between outcome groups 1 and 2, and between 2 and 3. They were present in all frequency ranges and in both active and quiet sleep. In significant cases, group 2 had always higher NC values than group 1. However, group 3 had lower values than group 2 in QS but higher values in AS. At least AS would seem to follow the results obtained from HIE class comparison.

ASI values did not have significant differences between any of the three outcome groups. This was the case also with the HIE classes.

MFDFA showed again more significant results in QS than in AS. In AS there was only one significant difference in *width hq* and one in *height Dq*. In QS there were significant differences in *width hq*, *tail Dq* and *height Dq*. With *width hq* there were differences between all outcome groups but they did not follow the same trend: in AS group 3 had larger values than group 2 but in QS they were lower. Additionally, in QS group 3 had lower values and group 2 higher values than group 1.

In *tail Dq* the significant differences were in the higher frequency ranges between groups 1 and 3 or groups 2 and 3. They all followed the same trend: the more severe

outcome, the larger the *tail Dq* values. Figure 21b shows one *tail Dq* variable with significant differences. This trend goes hand in hand with the results from HIE class comparison. In *height Dq* there was only one significant result in AS where the group 3 had lower values than group 2. In QS there were more significant differences and they were located in higher frequency bands and between groups 1 & 3 and 2 & 3. As with the HIE class comparison, the *height Dq* values were higher with the more severe groups.

Only one PSD variable showed statistically significant difference in AS but almost all variables showed differences in QS. All of the significant differences were between groups 1 and 3 or groups 2 and 3. All the QS variables followed the same trend as with HIE classes – the more severe is the case the lower is the PSD magnitude. The only significant result in AS was however to the opposite direction: group 3 had higher PSD values than group 1. One QS variable is presented in Figure 21c.

CPSD showed also only one significant difference in AS: group 3 had higher CPSD values than group 1. In QS almost all CPSD variables showed statistical difference between groups 1 and 3, and most of them also between 2 and 3. All statistically significant results showed that group 3 had lower CPSD values than the other two groups, which can be also seen Figure 21d. This follows the results with HIE classes except there was no difference between the less severe groups 1 and 2.

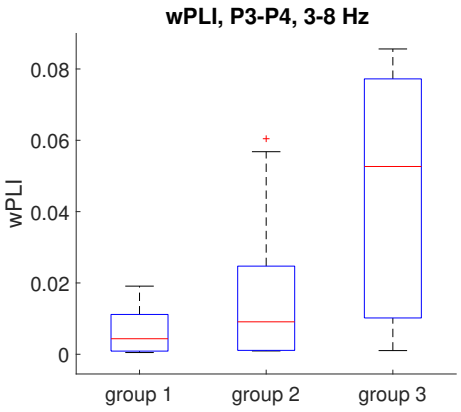
4.4 Classification

The fourth research question was about the computational classification of the subjects. Because of the lower number of subjects in AS, the classification was performed only with the QS variables. Those variables that had a relative variance of over 0.2 were excluded from the classification algorithm and 22 such variables were found. Four of those were from the MFDFA feature *tail Dq*, six from CPSD estimates and the rest from PSD estimates.

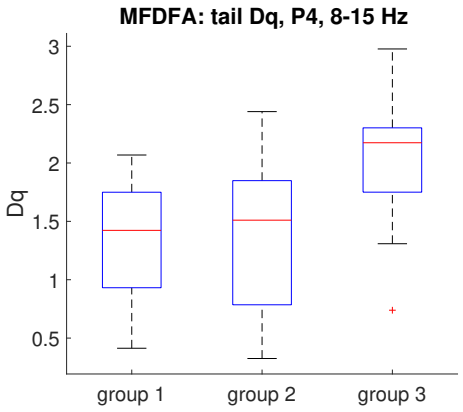
Because the number of newborns in the outcome group 3 was only 11, each fold in the 6-fold partition had only one or two babies from this group. Because of the imbalance between the two classes, we used ADASYN algorithm [73] to synthesize data points to the minority training group. However, this did not improve the classifier performance notably. In some situations, the ADASYN improved the sensitivity but at the same time lowered the specificity. All the SVM algorithms used a linear hyperplane.

Forward selection and p-value based methods were each applied 10 times to get an overview of the classifier performance. The round with the highest sensitivity was considered as the best results and the ranges of sensitivity and specificity were also reported. The results are written in the form: *best result* [*min*, *max*]. Brute force method was applied only once for both of the two classification problem while it takes significantly longer to run through.

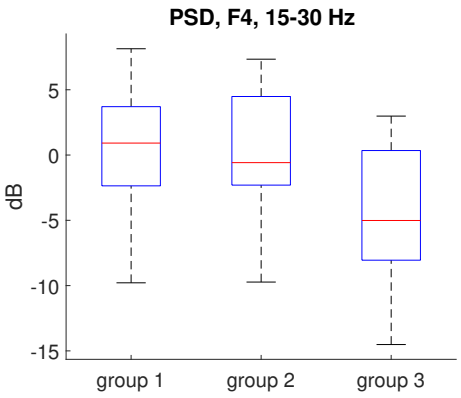
Classification between groups 1 & 2 and group 3 with forward selection achieved a sensitivity of 0.64 [0.27, 0.64] and a specificity of 0.90 [0.77, 0.90] at its best. The confusion matrix obtained using forward selection is presented in Table 6. The selected features and the resulting prediction accuracy varied between each fold,



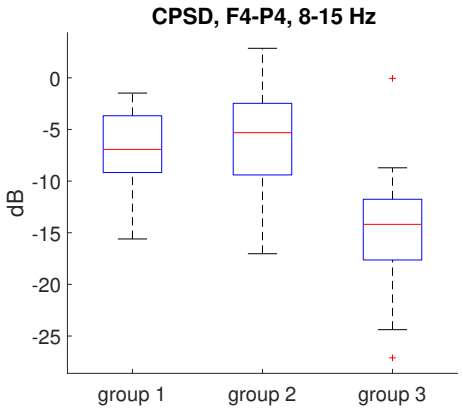
(a) Significant difference between groups 1-3 ($p<.01$) and 2-3 ($p<.05$).



(b) Significant difference between groups 1-3 ($p<.01$) and 2-3 ($p<.05$).



(c) Significant difference between groups 1-3 ($p<.01$) and 2-3 ($p<.01$).



(d) Significant difference between groups 1-3 ($p<.01$) and 2-3 ($p<.01$).

Figure 21: Boxplots present some QS feature values divided to the three outcome groups.

which highlights the fact that the algorithm performance is highly affected by the individuals that are selected to the training and testing sets. The selected features were mainly from the parietal channels of PSD and CPSD but also all of the other features were present, especially wPLI and *height Dq*.

Table 6: Confusion matrix of classifying group 3 with forward selection. The matrix represents the round with the highest sensitivity.

		Predicted group	
		1 & 2	3
Real group	1 & 2	28	3
	3	4	7

The same classification problem with the brute force method gave a sensitivity of 0.73 and a specificity of 0.74. Results are shown in Table 7. The outcome group was identified almost as accurately as with the forward selection method. Again the features varied a lot between the folds and they were mostly located in features *height Dq*, PSD and CPSD.

Table 7: Confusion matrix of classifying group 3 with brute force method.

		Predicted group	
		1 & 2	3
Real group	1 & 2	23	8
	3	3	8

Based on visual analysis of the computed EEG features and p-values from the Wilcoxon rank sum test, three variables were selected for classification of group 3. The selected features were wPLI from channel pair P3-P4 and with frequency band of 3-8 Hz, tail Dq from MFDFA in channel P4 with frequency band of 8-15 Hz, and CPSD between channels F4 and P4 in frequency band of 8-15 Hz. The selected variables are presented in Figure 21.

The resulting confusion matrix is presented in Table 8. The achieved specificity was 0.97 [0.90, 0.97] and sensitivity 0.64 [0.55, 0.64] Although the selected variables were always the same, the results were slightly different between each round because the training groups were different every time.

Classification between group 1 and groups 2 & 3 had a specificity of 0.67 and a sensitivity of 0.67 [0.33, 0.67] with the forward selection. The confusion matrix is presented in Table 9 and we can easily see the low specificity: over half the babies that were classified to group 1 are actually from groups 2 and 3. Again the selected features varied a lot between folds and runs. Interestingly the selected variables were

Table 8: Confusion matrix of classifying group 3 with p-value based method. The matrix represents the round with the highest sensitivity.

		Predicted group	
		1 & 2	3
Real group	1 & 2	30	1
	3	4	7

mostly from wPLI and there were no variables selected from PSD, which was not the case when classifying group 3 with the same method.

Table 9: Confusion matrix of classifying group 1 with forward selection. The matrix represents the round with the highest sensitivity.

		Predicted group	
		1	2 & 3
Real group	1	12	6
	2 & 3	16	8

With brute force method the group 1 was classified with an accuracy of 0.72 and specificity of 0.63 as is shown in Table 10 This means that over two thirds of the babies from group 1 were identified right, which is slightly better than with the forward method. The specificity was however even lower. Similarly to the forward method, the selected variables were mostly from wPLI.

Table 10: Confusion matrix of classifying group 1 with brute force method. The matrix represents the round with the highest sensitivity.

		Predicted group	
		1	2 & 3
Real group	1	13	5
	2 & 3	15	9

For classifying group 1 we again chose three variables based on p-values and visual analysis. One variable was the same as with the group 3 prediction, the CPSD variable that is presented in Figure 21d. The other two were wPLI between channels P3 and P4 in frequency range of 15-30 Hz, and NC in channel F3 with frequency range of 8-15 Hz. These variables resulted to a sensitivity of 0.71 [0.46, 0.71] and to a specificity of 0.63 [0.56, 0.78]. The confusion matrix is presented in Table 11.

Table 11: Confusion matrix of classifying group 1 with p-value based method. The matrix represents the round with the highest sensitivity.

		Predicted group	
		1	2 & 3
Real group	1	14	4
	2 & 3	10	14

When classifying between groups 1&2 and 3, there were two subjects from group 3 that were classified constantly wrong with all three feature selection methods. There were also 11 subjects from groups 1 and 2 that were always classified right. The p-value based method was notably more stable than the forward selection method, in which the misclassified subjects were quite different in each round. In the second classification problem there were only 2 subjects from group 1 that were constantly classified right and no subjects that would have been constantly classified wrong. P-value based method was again more stable than the forward selection but the misclassified subjects varied more from round to round than in the first classification problem.

Although the subjects from group 2 were considered to be the hardest to classify, we could not see any difference when compared to the other two groups. The only group specific difference was seen when classifying between groups 1&2 and 3 with the p-value based method: subjects from group 1 were always classified right.

4.5 Comparison between the features

For this study, a total number of 189 feature vectors were calculated. They came from eight different feature classes, that are described in Table 3. Five of the eight feature classes measured some kind of correlation between the four channels. AEC measures amplitude-amplitude correlation, PLV and wPLI describe the phase-phase correlation, ASI quantifies the interhemispheric synchrony of bursts and CPSD describes the spectral relationship between each channel pair. Three features were computational measures of individual signals. NC describes the phase-amplitude correlation, MFDFA the multifractal spectrum and PSD the spectral form of the signal.

If two features measure the same property, it can be assumed that there would be some noticeable linear correlation. For example, both PLV and wPLI describe the phase-phase correlation of signal pairs, although they have many differences that are summarized in Table 4. Also PSD and CPDS both measure spectral density, even though the former considers only single channels and the latter describes the cross correlation of two channels. We also used four measures to describe the multifractal spectrum calculated in MFDFA so it could be expected that these four variables showed some correlation with each other.

A correlation matrix was calculated from the QS features in order to inspect the

correlation between the features. The resulting matrix is presented in Figure 22, where each row and each column correspond to one of the 189 variables. For example, the first row and column is the wPLI calculated from channel pair F3-F4 in frequency range of 8-15 Hz. The color of each cell tells about the Spearman's correlation coefficient value – red for positive correlation and blue for negative. Spearman's correlation is defined as a Pearson correlation coefficient between the ranked values. If the cell is not coloured, the null hypothesis that there is no relationship between the variables is true with a 95% confidence level. Also the cells that have a low correlation value of $-0.2 < r_s < 0.2$ are white.

From the correlation matrix we can see that the diagonal elements have a perfect positive correlation, while they are the correlations of each variable to its self. It is also noticeable that some of the diagonal boxes that are each describing one feature are more full with coloured cells than other boxes of the matrix. Features wPLI, PLV, AEC, PSD and CPD have many positive correlation coefficients, which means that the different variables within each feature correlate well with each other. For example, the wPLI calculated from channel pair F3-F4 in frequency range of 5-8 Hz has a good correlation with the variable from same position and range of 15-30 Hz. This is the red cell on column 3 and row 4.

In these same features there are some white areas with no linear correlation – for example in PLV it would seem that the frontal pair F3-F4 does not correlate well with the other pairs. We can also see from the white grid in the PSD box that the variables that have a frequency range of 0.25-3 Hz do not have a linear correlation with the other three frequency bands. This is can be due to the vulnerability of the Pearson's correlation coefficient to outliers [74] while there was one subject with substantial PSD values in this frequency band.

NC and MFDFA have less correlation. The NC variables show a clear division between the frontal and parietal channels: variables from channels F3 and F4 correlate well with each other but not with variables from P3 and P4, and the other way around. In the correlation matrix this can be seen from the two separate red areas in the NC box. The MFDFA measures *peak hq, width hq, tail Dq* and *height Dq* have very little correlation within each of them but there is a clear correlation between them. *Width hq, tail Dq* and *height Dq* variables measured from the same multifractal spectrum have a good positive or negative correlation, which can be seen from the diagonal elements in these boxes. *Width hq* has a clear negative correlation with *tail Dq* and *height Dq*, and the last two have a positive correlation with each other. *Peak hq* has very little correlation with the other three measures.

The phase-phase correlation measures wPLI and PLV do not show any clear linear relationship with each other. This can be caused by the fact that even though they measure the same thing, they work quite differently: wPLI skips small phase lags while PLV accounts them all. Surprisingly enough there is a strong correlation between AEC and PLV measures and between AEC and the spectral measures PSD and CPSD even though they measure different aspects of the EEG. A clear relationship is also seen between the two spectral measures.

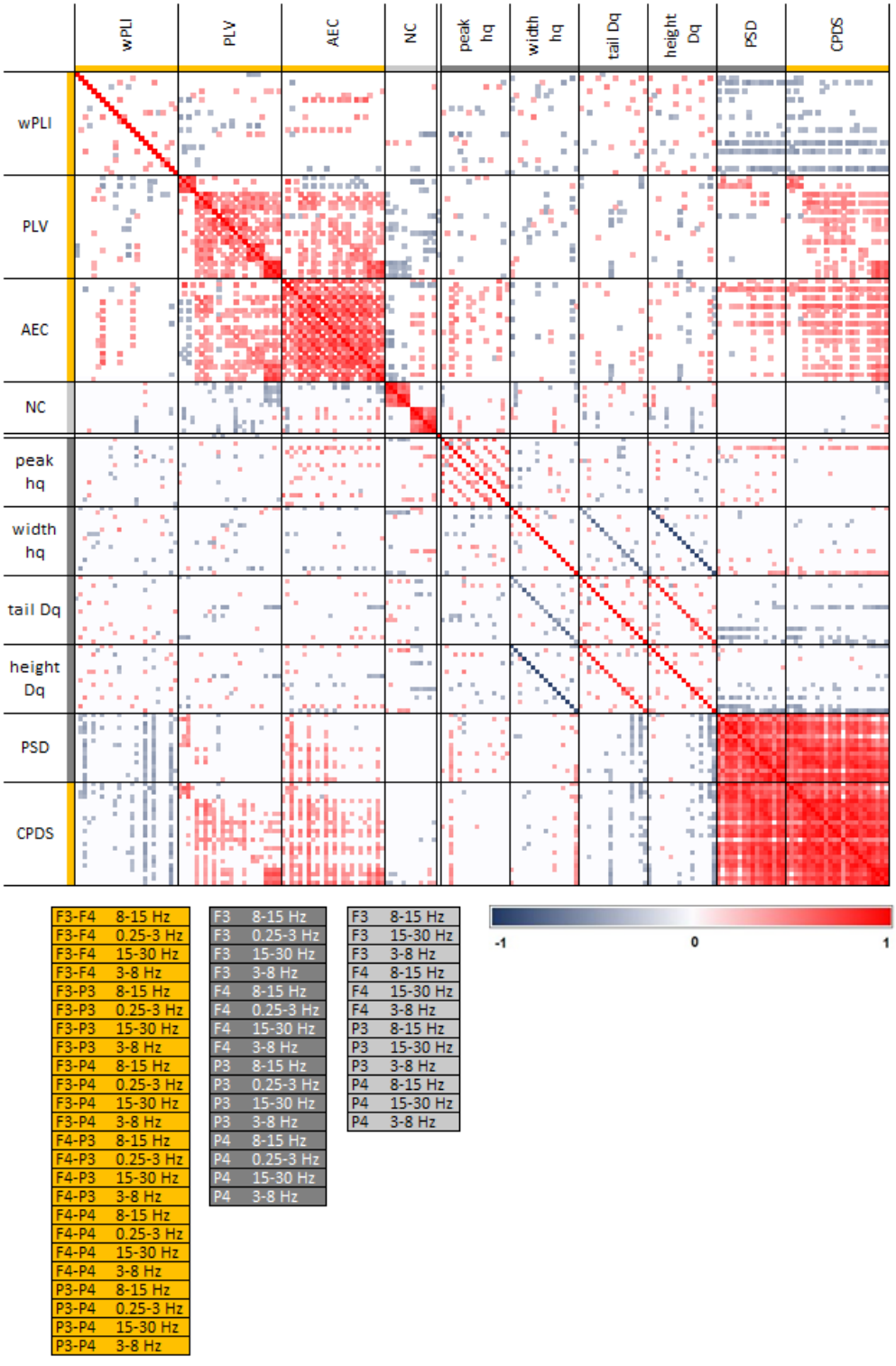


Figure 22: Spearman's correlation matrix of all computational EEG features. Each row and column corresponds to one feature variable. Yellow and grey scales tell the exact features inside every feature class. The brightest red corresponds to 1 (perfect positive correlation) and blue to -1 (perfect negative correlation). White cells either have a p-value over 0.05 or correlation between $-0.2 < r_s < 0.2$. The single row feature in the middle is ASI.

5 Discussion

This study provides further evidence of the usefulness of computational EEG features when assessing neonatal brain activity. The results show that these features can be affected by different sleep stages and different epoch lengths. Some of them showed a good separation of different HIE classes and outcome groups. It was also shown that these features can be used in automated outcome classification with promising results. In this section these results are discussed and evaluated.

5.1 Stability of the features

Most of the features were quite stable between the different epoch lengths. The feature classes that showed most variation between the sleep stages were MFDFA measure *tail Dq* and the two spectral measures. Especially PSD values were substantially different between the three epoch lengths.

Some features were very similar across the two sleep stages. Especially AEC, PLV and CPSD showed strong correlations, while wPLI and some MFDFA measures varied a lot between AS and QS. At some level this also affected the statistically significant differences between HIE classes and outcome groups in AS. The features that had only small variation between the sleep stages also showed more significant results in AS. The comparison between the sleep stages also showed that there were differences especially in the specific frequency bands, not so much on specific spatial locations. This can be explained by the different spectral properties between quiet and active sleep – especially the amount of delta oscillations varies between the sleep stages [4].

Because we were not able to extract AS epochs for many of the most seriously affected babies, the comparison between the two sleep stages was done with a smaller subject number. The correlation between features calculated for quiet and active sleep epochs could be something else than linear or there could be stronger correlations if the subjects were divided to smaller subgroups based on the severity of HIE. Even without these scrutinies included, the comparison between AS and QS supported the claim that QS is diagnostically more sensitive to variations in brain activity [23]. This can be seen from the amount of statistically significant results that are listed in Appendix B and Appendix C.

5.2 Separating HIE classes and outcome groups

There were many features that correlated well with the clinical status. A total number of 69 feature variables showed significant differences at least between one gradus pair in QS which is seen in Appendix B. With a significance level of $\alpha = 0.05$ there should only be by chance about 10 features with significant results. The feature classes that showed statistically significant group differences on this confidence level were AEC, PLV, wPLI, MFDFA, PSD and CPSD. Figure 19 shows some of these significant results.

In AS there were 37 feature variables showing significant group differences with significance level 0.05 which is shown in Appendix C. These variables were mostly from AEC, PLV and CPSD, that were all quite stable between the two sleep stages. Also in the other feature classes the overall trends followed fairly well the ones in quiet sleep. All the statistically significant differences in AS were between HIE classes I and II which is understandable while class III only included one baby.

Less features correlated with the outcome. In QS there were 56 variables that showed significant differences with a confidence level of 0.05 and 36 variables if the level was lowered to the Bonferroni corrected level. In AS there were 9 variables with significance level 0.05 but none with the Bonferroni corrected level. Most of these statistically significant variables in QS also showed significant results when comparing HIE classes. Exceptions were few variables of wPLI, nestedness and MFDFA – they showed significant differences only between the outcome groups. In AS none of the variables that gave significant results between outcome groups showed significant results between HIE classes.

5.3 Classifier performance

The SVM classifiers had a better sensitivity when detecting group 1 instead of group 3. The best sensitivity 0.72 was achieved with the brute force method but it also had the lowest specificity. When detecting group 3, forward selection and p-value based methods achieved a sensitivity of 0.64 at least once. The results varied a lot between each round and especially forward selection seemed to have quite large ranges for both sensitivity and specificity. The p-value based method was quite stable when detecting group 3, but less so when detecting group 1. Brute force method resulted in a slightly higher sensitivity 0.73 but it's specificity was quite low. Because it was only tested once, not much can be said about its stability.

If the EEG based prediction was used in clinical work, it should be at least as accurate as the evaluation of medical doctor. The medical doctor evaluates the condition of the baby with the HIE class and these gradings can be seen in Figure 4. Using only the HIE classes, the prediction of outcome group 3 would have a sensitivity of 0.82 and specificity of 0.98. When searching the normal outcome, outcome group 1, the sensitivity would be 0.56 and specificity is 0.88. The SVM classifier constructed in this study did not reach a performance that would match that of using HIE classes. However, all of the SVM methods gave a better sensitivity than HIE class when predicting group 1. From Figure 4 we can see that the babies HIE class II often fall to the outcome group 1. The comparison of different methods is presented in Table 12.

Because we used information from all subjects when selecting the features in the p-value based method, it is not as data driven as the two other methods. For forward selection and brute force methods it was obvious that the number of subjects was too low while the results varied a lot between each fold and each run in the case of forward selection. When investigating the selected features, we could still see that the features that separated group 3 from the others were different from those that separated group 1. For example, the variables from CPSD appeared to separate

Table 12: Results of predicting the outcome with HIE classes and with different SVM based methods. Forward selection and p-value based methods were applied 10 times and the table presents their minimum and maximum results. The largest value of each row is highlighted.

		HIE class	Forward selection	Brute force	P-value based
Between 1&2 and 3	Sensitivity	**0.82**	[0.27, 0.64]	0.73	[0.55, 0.64]
	Specificity	**0.97**	[0.77, 0.90]	0.74	[0.90, **0.97**]
Between 1 and 2&3	Sensitivity	0.56	[0.44, 0.72]	0.72	[0.56, **0.78**]
	Specificity	**0.88**	[0.33, 0.67]	0.38	[0.46, 0.71]

group 3 well but not group 1, and the wPLI variables the other way around.

The SVM algorithm was used only to do some exploratory preliminary analysis. Because the algorithm was only tested with a set of three features, with linear hyperplane and without any optimization, the results were expected to be moderate. If the classifier parameters were optimized more, it could improve the results.

5.4 Limitations

Several factors can impact the achieved results and should be studied in more detail in the future. The first factor is the selection of the feature classes: there could be more sophisticated measures that would perform better than the ones selected to this work. It could also be possible to combine several different features or simplify the used features for example by calculating the mean over each channel.

One of the main problems with the HIE classes is that the neonatal HIE status is not static: it evolves over time and in this study only one time point was used to select the class for each baby [4, 40]. Because of this, we should always have knowledge of the clinical status at the exact time when EEG is recorded [75]. The outcome prediction based on HIE classes shown in Table 12 could also be more accurate if the fluctuation of the HIE gradus was known. Also the outcome estimation was made from only one time point approximately at age 1 and some neurological problems caused by asphyxia, such as learning deficit, cannot be diagnosed at this age [4, 23]. The true outcomes of the study subjects were thus unknown. The estimates of both HIE gradus and outcome were also made by a single medical doctor, which can cause some error.

Another factor that can affect the results is maturation. The feature correlation to age was inspected visually but more objective and quantitative analysis should take place. The age of the neonates at the time of the EEG recording varied a lot as seen from Figure 5. While the brain activity of a newborn can be quite different after five days [21], the age can have some effect on the results.

Also different medications can have a huge impact on the neonatal brain activity.

Typical medication for critically ill neonates include substances like phenobarbitone, banzodiazepines and morphines, and they all decrease the overall amplitude and increase discontinuity of the EEG signal [23, 28]. This should be taken into consideration while the most severe HIE cases are also those who often receive the most medication.

One additional factor affecting the brain activity is the hypothermia treatment. The rewarming after the treatment changes the EEG signal, especially the discontinuity of it [76]. From the 42 subjects in this study, over 60% (26) were cooled. Hypothermia treatment is more often used to neonates with more severe HIE grading. In this study, all the subjects from HIE class III and only 5 of the 13 babies from class I underwent the hypothermia treatment.

5.5 Future prospects

The results showed that the computational EEG features vary significantly between the different HIE classes and outcome group and should thus be studied more in order to guarantee the best possible treatment for neonates with HIE. Especially the values of wPLI, MFDFA and CPSD seemed to separate well between the severity of HIE and outcome. The classification showed that it is indeed possible to use these features to predict the severity of the neurological problems that follow HIE. Although the predictions only achieved moderate results, they are promising for the future use of EEG in clinical work.

Next step for this study would be to test the most promising features with a larger number of subjects. This would offer a valuable confirmation of the results presented in this study. The features should also be compared to a background activity grading of the EEG epochs made by a specialist. Now we used a measure from the past (HIE gradus) and from the future (outcome) – background grading would offer valuable information about the same time point the features were calculated. The limitations discussed before should also be inspected more carefully. If the classification algorithm was extended to classify multiple groups and optimized more carefully, there is a possibility that it could be implemented in clinical use.

6 Conclusions

The aim of this study was to inspect the potential of computational EEG features when assessing and predicting the neurological state of the asphyxiated neonates. The work was divided to four different research questions, presented in the beginning of this thesis. Eight different feature classes were calculated from the 4-channel EEG recording in frequency bands between 0.25-30 Hz for both quiet and active sleep. This resulted to a total number of 189 features.

The first research question asked, if the computed EEG features were different between quiet and active sleep. Some feature classes, such as AEC, PLV and CPSD, showed fairly good consistency between the two sleep stages. There was quite a lot variation but also many systematic differences.

With the second research question, the goal was to find out if the features reflect the HIE classes of the babies. Most of the feature classes showed significant results and especially AEC, MFDFA, PSD and CPSD separated the classes well. Features computed from quiet sleep epochs reflected the different HIE classes better and only QS features that did not separate the three classes at all were ASI and NC.

The third research question asked about the relationship between the features and the outcomes of the babies. There were slightly less significant results than with the HIE classes and they had a big overlap to the features that were significant in the HIE class separation. However, AEC did not separate outcome groups as well as HIE classes, and NC separated them better than HIE classes. Active sleep showed even less significant results with outcome than with HIE classes.

With the fourth question the aim was to do a preliminary analysis of the usability of these features in automatized classification. The best achieved sensitivity was 0.74 and specificity 0.97 when classifying between groups 1&2 and 3. When classifying between groups 1 and 2&3, a sensitivity of 0.72 and a specificity of 0.67 were achieved. Although the results varied quite a lot between each run and each fold, we were able to show that there is a potential advantage in using these computed EEG features to predict the outcome of the asphyxiated neonate.

This study estimated the clinical usefulness of computational EEG features and we were able to show that there is a lot of potential in them. The results show that the use of quiet sleep epochs could be a better choice when evaluating neonates with HIE. It is important at least to know the sleep stage when selecting the epochs used in the analysis. It was showed that there are many different features, spatial locations and frequencies that can be used when estimating the severity of the injury and especially AEC, wPLI, MFDFA and CPSD should be studied more.

With a larger sample size, more accurate HIE class and outcome group evaluations and a more precisely optimized classifier it could be possible to achieve even better results in the future. It is also important to study the factors that influence the EEG signal: maturation, medication and possible hypothermia treatment. A further development of computational EEG features combined with machine learning methods can become an important tool in the monitoring and diagnosing of neonates after asphyxia.

References

[1] J. Bryce, C. Boschi-Pinto, K. Shibuya, and R. E. Black, "WHO estimates of the causes of death in children," *The Lancet*, vol. 365, no. 9465, pp. 1147–1152, 2005.

[2] L. Liu, S. Oza, D. Hogan, J. Perin, I. Rudan, J. E. Lawn, S. Cousens, C. Mathers, and R. E. Black, "Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: An updated systematic analysis," *The Lancet*, vol. 385, no. 9966, pp. 430–440, 2015.

[3] J. E. Lawn, S. Cousens, J. Zupan, and Lancet Neonatal Survival Steering Team, "4 million neonatal deaths: When? Where? Why?," *The Lancet*, vol. 365, no. 9462, pp. 891–900, 2005.

[4] H. Lagercrantz, M. A. Hanson, L. R. Ment, and D. M. Peebles, *The Newborn Brain: Neuroscience and Clinical Applications*. Cambridge University Press, 2nd ed., 2010.

[5] J. A. Leipälä, M. Metsäranta, R. Marttila, R. Grahn, R. Sund, and A. Malmivaara, "Viilennyksen vaikuttavuus ja turvallisuus vastasyntyneen hypoksis-iskeemisen enkefalopatian hoidossa," *Suomen Lääkärilehti*, vol. 63, no. 49, 2008.

[6] W. Gao, W. Lin, K. Grewen, and J. H. Gilmore, "Functional connectivity of the infant human brain: Plastic and modifiable," *The Neuroscientist*, 2016. doi: 10.1177/1073858416635986.

[7] C. F. Hagmann, J. Rennie, and N. Robertson, *Neonatal Cerebral Investigation*. Cambridge University Press, 2008.

[8] A. Bruns, R. Eckhorn, H. Jokeit, and A. Ebner, "Amplitude envelope correlation detects coupling among incoherent brain signals," *Neuroreport*, vol. 11, no. 7, pp. 1509–1514, 2000.

[9] J.-P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring phase synchrony in brain signals," *Human Brain Mapping*, vol. 8, no. 4, pp. 194–208, 1999.

[10] M. Vinck, R. Oostenveld, M. van Wingerden, F. Battaglia, and C. M. Pennartz, "An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias," *Neuroimage*, vol. 55, no. 4, pp. 1548–1565, 2011.

[11] S. Vanhatalo, J. M. Palva, S. Andersson, C. Rivera, J. Voipio, and K. Kaila, "Slow endogenous activity transients and developmental expression of K+–Cl- co-transporter 2 in the immature human cortex," *European Journal of Neuroscience*, vol. 22, no. 11, pp. 2799–2804, 2005.

[12] O. Räsänen, M. Metsäranta, and S. Vanhatalo, "Development of a novel robust measure for interhemispheric synchrony in the neonatal EEG: Activation synchrony index (ASI)," *Neuroimage*, vol. 69, pp. 256–266, 2013.

[13] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley, "Multifractal detrended fluctuation analysis of nonstationary time series," *Physica A: Statistical Mechanics and its Applications*, vol. 316, no. 1, pp. 87–114, 2002.

[14] G. J. Tortora and B. H. Derrickson, *Principles of Anatomy and Physiology.* John Wiley & Sons, Inc., 12th ed., 2008.

[15] M. J. Aminoff, *Aminoff's Electrodiagnosis in Clinical Neurology.* Saunders, Elsevier, 6th ed., 2012.

[16] H. Jasper, "The 10/20 international electrode system," *EEG and Clinical Neurophysiology*, vol. 10, pp. 371–375, 1958.

[17] V. L. Towle, J. Bolaños, D. Suarez, K. Tan, R. Grzeszczuk, D. N. Levin, R. Cakmur, S. A. Frank, and J.-P. Spire, "The spatial location of EEG electrodes: Locating the best-fitting sphere relative to cortical anatomy," *Electroencephalography and Clinical Neurophysiology*, vol. 86, no. 1, pp. 1–6, 1993.

[18] M. Steriade, P. Gloor, R. R. Llinas, F. L. Da Silva, and M.-M. Mesulam, "Basic mechanisms of cerebral rhythmic activities," *Electroencephalography and Clinical Neurophysiology*, vol. 76, no. 6, pp. 481–508, 1990.

[19] D. H. Sanes, T. A. Reh, and W. A. Harris, *Development of the Nervous System.* Academic Press, Elsevier, 3rd ed., 2011.

[20] A. Qiu, S. Mori, and M. I. Miller, "Diffusion tensor imaging for understanding brain development in early life.," *Annual Review of Psychology*, vol. 66, pp. 853–876, 2015.

[21] S. Vanhatalo and K. Kaila, "Development of neonatal EEG activity: From phenomenology to physiology," in *Seminars in Fetal and Neonatal Medicine* (V. Fellman, L. Hellström-Westas, and I. Roseén, eds.), vol. 11, pp. 471–478, 2006.

[22] M. A. Carskadon and W. C. Dement, "Normal human sleep: An overview," *Principles and Practice of Sleep Medicine*, vol. 4, pp. 13–23, 2005.

[23] H. Pihko, L. Haataja, and H. Rantala, *Lastenneurologia.* Kustannus Oy Duodecim, 2014.

[24] J. A. Low, "Determining the contribution of asphyxia to brain damage in the neonate," *Journal of Obstetrics and Gynaecology Research*, vol. 30, no. 4, pp. 276–286, 2004.

[25] K. Sävman, M. Blennow, K. Gustafson, E. Tarkowski, and H. Hagberg, "Cytokine response in cerebrospinal fluid after birth asphyxia," *Pediatric Research*, vol. 43, no. 6, pp. 746–751, 1998.

[26] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*. New York: Garland Publishing Inc, 5th ed., 2008.

[27] D. Osredkar, M. C. Toet, L. G. van Rooij, A. C. van Huffelen, F. Groenendaal, and L. S. de Vries, "Sleep-wake cycling on amplitude-integrated electroencephalography in term newborns with hypoxic-ischemic encephalopathy," *Pediatrics*, vol. 115, no. 2, pp. 327–332, 2005.

[28] B. Walsh, D. Murray, and G. Boylan, "The use of conventional EEG for the assessment of hypoxic ischaemic encephalopathy in the newborn: A review," *Clinical Neurophysiology*, vol. 122, no. 7, pp. 1284–1294, 2011.

[29] H. B. Sarnat and M. S. Sarnat, "Neonatal encephalopathy following fetal distress.," *Obstetrical & Gynecological Survey*, vol. 32, no. 5, p. 295, 1977.

[30] S. Shankaran, A. R. Laptook, R. A. Ehrenkranz, J. E. Tyson, S. A. McDonald, E. F. Donovan, A. A. Fanaroff, W. K. Poole, L. L. Wright, and R. D. Higgins, "Whole-body hypothermia for neonates with hypoxic–ischemic encephalopathy," *New England Journal of Medicine*, vol. 353, no. 15, pp. 1574–1584, 2005.

[31] D. Azzopardi, P. Brocklehurst, D. Edwards, H. Halliday, M. Levene, and M. Thoresen, "The TOBY study. whole body hypothermia for the treatment of perinatal asphyxial encephalopathy: A randomised controlled trial," *BMC Pediatrics*, vol. 8, p. 17, 2008.

[32] V. Tommiska and M. Metsäranta, "Vastasyntyneiden viilennyshoito," *Duodecim*, vol. 128, no. 12, pp. 1275–1282, 2012.

[33] N. Merchant and D. Azzopardi, "Early predictors of outcome in infants treated with hypothermia for hypoxic–ischaemic encephalopathy," *Developmental Medicine & Child Neurology*, vol. 57, no. S3, pp. 8–16, 2015.

[34] T. W. Pin, B. Eldridge, and M. P. Galea, "A review of developmental outcomes of term infants with post-asphyxia neonatal encephalopathy," *European Journal of Paediatric Neurology*, vol. 13, no. 3, pp. 224–234, 2009.

[35] C. Robertson and N. Finer, "Term infants with hypoxic-ischemic encephalopathy: Outcome at 3.5 years," *Developmental Medicine & Child Neurology*, vol. 27, no. 4, pp. 473–484, 1985.

[36] F. M. Cowan and L. S. de Vries, "The internal capsule in neonatal imaging," in *Seminars in Fetal and Neonatal Medicine*, vol. 10, pp. 461–474, Elsevier, 2005.

[37] M. Toet, L. Hellström-Westas, F. Groenendaal, P. Eken, and L. De Vries, "Amplitude integrated EEG 3 and 6 hours after birth in full term neonates with hypoxic–ischaemic encephalopathy," *Archives of Disease in Childhood-Fetal and Neonatal Edition*, vol. 81, no. 1, pp. F19–F23, 1999.

[38] N. al Naqeeb, A. D. Edwards, F. M. Cowan, and D. Azzopardi, "Assessment of neonatal encephalopathy by amplitude-integrated electroencephalography," *Pediatrics*, vol. 103, no. 6, pp. 1263–1271, 1999.

[39] G. Ancora, S. Soffritti, R. Lodi, C. Tonon, S. Grandi, C. Locatelli, L. Nardi, N. Bisacchi, C. Testa, and G. Tani, "A combined a-EEG and MR spectroscopy study in term newborns with hypoxic–ischemic encephalopathy," *Brain and Development*, vol. 32, no. 10, pp. 835–842, 2010.

[40] V. Matić, P. J. Cherian, K. Jansen, N. Koolen, G. Naulaers, R. M. Swarte, P. Govaert, S. Van Huffel, and M. De Vos, "Improving reliability of monitoring background EEG dynamics in asphyxiated infants," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 5, pp. 973–983, 2016.

[41] N. Stevenson, I. Korotchikova, A. Temko, G. Lightbody, W. Marnane, and G. Boylan, "An automated system for grading EEG abnormality in term neonates with hypoxic-ischaemic encephalopathy," *Annals of Biomedical Engineering*, vol. 41, no. 4, pp. 775–785, 2013.

[42] I. Korotchikova, N. Stevenson, B. Walsh, D. Murray, and G. Boylan, "Quantitative EEG analysis in neonatal hypoxic ischaemic encephalopathy," *Clinical Neurophysiology*, vol. 122, no. 8, pp. 1671–1678, 2011.

[43] K. K. Iyer, J. A. Roberts, M. Metsäranta, S. Finnigan, M. Breakspear, and S. Vanhatalo, "Novel features of early burst suppression predict outcome after birth asphyxia," *Annals of Clinical and Translational Neurology*, vol. 1, no. 3, pp. 209–214, 2014.

[44] M. Hathi, D. Sherman, T. Inder, N. Rothman, M. Natarajan, C. Niesen, L. Korst, T. Pantano, and A. Natarajan, "Quantitative EEG in babies at risk for hypoxic ischemic encephalopathy after perinatal asphyxia," *Journal of Perinatology*, vol. 30, no. 2, pp. 122–126, 2010.

[45] M.-D. Lamblin, E. W. Esquivel, and M. André, "The electroencephalogram of the full-term newborn: Review of normal features and hypoxic-ischemic encephalopathy patterns," *Neurophysiologie Clinique/Clinical Neurophysiology*, vol. 43, no. 5, pp. 267–287, 2013.

[46] A. Tokariev, K. Palmu, A. Lano, M. Metsäranta, and S. Vanhatalo, "Phase synchrony in the early preterm EEG: Development of methods for estimating synchrony in both oscillations and events," *Neuroimage*, vol. 60, no. 2, pp. 1562–1573, 2012.

[47] F. Varela, J.-P. Lachaux, E. Rodriguez, and J. Martinerie, "The brainweb: Phase synchronization and large-scale integration," *Nature Reviews Neuroscience*, vol. 2, no. 4, pp. 229–239, 2001.

[48] F. Varela and E. Thompson, "Neural synchrony and the unity of mind: A neurophenomenological perspective," in *The Unity of Consciousness* (A. Cleereman, ed.), Oxford University Press, 2003.

[49] Y. Zheng, J. Gao, J. C. Sanchez, J. C. Principe, and M. S. Okun, "Multiplicative multifractal modeling and discrimination of human neuronal activity," *Physics Letters A*, vol. 344, no. 2, pp. 253–264, 2005.

[50] A. H. Bell, B. McClure, and E. Hicks, "Power spectral analysis of the EEG of term infants following birth asphyxia," *Developmental Medicine & Child Neurology*, vol. 32, no. 11, pp. 990–998, 1990.

[51] R. Ahmed, A. Temko, W. Marnane, G. Lightbody, and G. Boylan, "Grading hypoxic–ischemic encephalopathy severity in neonatal EEG using GMM supervectors and the support vector machine," *Clinical Neurophysiology*, vol. 127, no. 1, pp. 297–309, 2016.

[52] J. M. Palva and S. Palva, "Roles of multiscale brain activity fluctuations in shaping the variability and dynamics of psychophysical performance," *Progress in Brain Research*, vol. 193, p. 335, 2011.

[53] S. Palva and J. M. Palva, "Discovering oscillatory interaction networks with M/EEG: Challenges and breakthroughs," *Trends in Cognitive Sciences*, vol. 16, no. 4, pp. 219–230, 2012.

[54] M. J. Brookes, M. Woolrich, H. Luckhoo, D. Price, J. R. Hale, M. C. Stephenson, G. R. Barnes, S. M. Smith, and P. G. Morris, "Investigating the electrophysiological basis of resting state networks using magnetoencephalography," *Proceedings of the National Academy of Sciences*, vol. 108, no. 40, pp. 16783–16788, 2011.

[55] B. Boashash, *Time-Frequency Signal Analysis and Processing*. Academic Press, Elsevier, 2nd ed., 2015.

[56] W. Singer, "Neuronal synchrony: A versatile code for the definition of relations?," *Neuron*, vol. 24, no. 1, pp. 49–65, 1999.

[57] C. J. Stam, G. Nolte, and A. Daffertshofer, "Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources," *Human Brain Mapping*, vol. 28, no. 11, pp. 1178–1193, 2007.

[58] A. K. Engel, C. Gerloff, C. C. Hilgetag, and G. Nolte, "Intrinsic coupling modes: Multiscale interactions in ongoing brain activity," *Neuron*, vol. 80, no. 4, pp. 867–886, 2013.

[59] A. Tokariev, M. Videman, J. M. Palva, and S. Vanhatalo, "Functional brain connectivity develops rapidly around term age and changes between vigilance states in the human newborn," *Cerebral Cortex*, vol. 26, no. 12, pp. 4540–4550, 2016.

[60] R. C. Sotero, A. Bortel, S. Naaman, V. M. Mocanu, P. Kropf, M. Villeneuve, and A. Shmuel, "Laminar distribution of phase-amplitude coupling of spontaneous current sources and sinks," *Frontiers in Neuroscience*, vol. 9, p. 454, 2015.

[61] N. Koolen, A. Dereymaeker, K. Jansen, J. Vervisch, V. Matic, M. De Vos, S. Van Huffel, G. Naulaers, and S. Vanhatalo, "Interhemispheric synchrony in the neonatal EEG revisited: Activation synchrony index as a promising classifier.," *Frontiers in Human Neuroscience*, vol. 8, 2014.

[62] N. Koolen, A. Dereymaeker, O. Räsänen, K. Jansen, J. Vervisch, V. Matic, G. Naulaers, M. De Vos, S. Van Huffel, and S. Vanhatalo, "Early development of synchrony in cortical activations in the human," *Neuroscience*, vol. 322, pp. 298–307, 2016.

[63] M. Videman, A. Tokariev, H. Saikkonen, S. Stjerna, H. Heiskala, O. Mantere, and S. Vanhatalo, "Newborn brain function is affected by fetal exposure to maternal serotonin reuptake inhibitors," *Cerebral Cortex*, 2016. doi: 10.1093/cercor/bhw153.

[64] E. A. Ihlen, "Introduction to multifractal detrended fluctuation analysis in Matlab," *Fractal Analyses: Statistical And Methodological Innovations And Best Practices*, p. 97, 2012.

[65] V. Matic, P. J. Cherian, N. Koolen, A. H. Ansari, G. Naulaers, P. Govaert, S. Van Huffel, M. De Vos, and S. Vanhatalo, "Objective differentiation of neonatal EEG background grades using detrended fluctuation analysis," *Frontiers in Human Neuroscience*, vol. 9, no. 189, 2015.

[66] D. Sinclair, M. Campbell, P. Byrne, W. Prasertsom, and C. Robertson, "EEG and long-term outcome of term infants with neonatal hypoxic-ischemic encephalopathy," *Clinical Neurophysiology*, vol. 110, no. 4, pp. 655–659, 1999.

[67] V. Matic, P. J. Cherian, N. Koolen, G. Naulaers, R. M. Swarte, P. Govaert, S. Van Huffel, and M. De Vos, "Holistic approach for automated background EEG assessment in asphyxiated full-term infants," *Journal of Neural Engineering*, vol. 11, no. 6, p. 066007, 2014.

[68] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[69] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

[70] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*. Springer, 2011.

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

[72] J. D. Rodriguez, A. Perez, and J. A. Lozano, "Sensitivity analysis of k-fold cross validation in prediction error estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 569–575, 2010.

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

[74] A. Asuero, A. Sayago, and A. Gonzalez, "The correlation coefficient: An overview," *Critical Reviews in Analytical Chemistry*, vol. 36, no. 1, pp. 41–59, 2006.

[75] N. S. Abend, D. J. Dlugos, and R. R. Clancy, "A review of long-term EEG monitoring in critically ill children with hypoxic-ischemic encephalopathy, congenital heart disease, ECMO, and stroke," *Journal of Clinical Neurophysiology: Official Publication of the American Electroencephalographic Society*, vol. 30, no. 2, p. 134, 2013.

[76] A. Birca, A. Lortie, V. Birca, J.-C. Decarie, A. Veilleux, A. Gallagher, M. Dehaes, G. A. Lodygensky, and L. Carmant, "Rewarming affects EEG background in term newborns with hypoxic–ischemic encephalopathy undergoing therapeutic hypothermia," *Clinical Neurophysiology*, vol. 127, no. 4, pp. 2087–2094, 2016.

Appendices

A Feature comparisons between QS and AS

The tables show the comparison between QS ans AS epochs for each feature. The vectors were compared with the Wilcoxon sign rank test, for which the resulting p-values are listed. If the p-value is lower than 0.05 the null hypothesis is rejected and it can be assumed that the vectors come from two individual distributions. Also the Pearson's correlation coefficients were calculated for each variable. If the p-value is under 0.05, the null hypothesis of no correlation at all, can be rejected. P-values lower than 0.05 are again marked with a grey background.

Feature	Channel	Frequency band (Hz)	Wilcoxon signed rank p-value	Pearson's correlation coefficient	Correlation coefficient pvalue
AEC	F3-F4	8-15	0.000	0.65	0.000
AEC	F3-F4	0.25-3	0.322	0.48	0.003
AEC	F3-F4	15-30	0.000	0.65	0.000
AEC	F3-F4	3-8	0.003	0.74	0.000
AEC	F3-P3	8-15	0.000	0.72	0.000
AEC	F3-P3	0.25-3	0.226	0.54	0.001
AEC	F3-P3	15-30	0.000	0.71	0.000
AEC	F3-P3	3-8	0.000	0.85	0.000
AEC	F3-P4	8-15	0.000	0.68	0.000
AEC	F3-P4	0.25-3	0.470	0.61	0.000
AEC	F3-P4	15-30	0.000	0.68	0.000
AEC	F3-P4	3-8	0.000	0.81	0.000
AEC	F4-P3	8-15	0.000	0.75	0.000
AEC	F4-P3	0.25-3	0.615	0.57	0.000
AEC	F4-P3	15-30	0.000	0.66	0.000
AEC	F4-P3	3-8	0.000	0.80	0.000
AEC	F4-P4	8-15	0.000	0.76	0.000
AEC	F4-P4	0.25-3	0.144	0.65	0.000
AEC	F4-P4	15-30	0.000	0.77	0.000
AEC	F4-P4	3-8	0.000	0.77	0.000
AEC	P3-P4	8-15	0.000	0.86	0.000
AEC	P3-P4	0.25-3	0.346	0.68	0.000
AEC	P3-P4	15-30	0.000	0.84	0.000
AEC	P3-P4	3-8	0.011	0.85	0.000
PLV	F3-F4	8-15	0.102	0.83	0.000
PLV	F3-F4	0.25-3	0.863	0.74	0.000
PLV	F3-F4	15-30	0.008	0.86	0.000
PLV	F3-F4	3-8	0.116	0.68	0.000
PLV	F3-P3	8-15	0.293	0.89	0.000
PLV	F3-P3	0.25-3	0.076	0.85	0.000
PLV	F3-P3	15-30	0.116	0.93	0.000
PLV	F3-P3	3-8	0.167	0.92	0.000
PLV	F3-P4	8-15	0.057	0.76	0.000
PLV	F3-P4	0.25-3	0.551	0.83	0.000
PLV	F3-P4	15-30	0.285	0.88	0.000
PLV	F3-P4	3-8	0.489	0.82	0.000
PLV	F4-P3	8-15	0.004	0.80	0.000
PLV	F4-P3	0.25-3	0.777	0.77	0.000
PLV	F4-P3	15-30	0.177	0.81	0.000
PLV	F4-P3	3-8	0.167	0.86	0.000
PLV	F4-P4	8-15	0.637	0.81	0.000
PLV	F4-P4	0.25-3	0.053	0.81	0.000
PLV	F4-P4	15-30	0.087	0.89	0.000
PLV	F4-P4	3-8	0.038	0.87	0.000
PLV	P3-P4	8-15	0.007	0.91	0.000
PLV	P3-P4	0.25-3	0.001	0.91	0.000
PLV	P3-P4	15-30	0.888	0.94	0.000
PLV	P3-P4	3-8	0.001	0.89	0.000

Feature	Channel	Frequency band (Hz)	Wilcoxon signed rank p-value	Pearson's correlation coefficient	Correlation coefficient pvalue
wPLI	F3-F4	8-15	0.354	-0.08	0.661
wPLI	F3-F4	0.25-3	0.838	-0.07	0.702
wPLI	F3-F4	15-30	0.167	0.46	0.004
wPLI	F3-F4	3-8	0.220	-0.31	0.068
wPLI	F3-P3	8-15	0.271	0.42	0.012
wPLI	F3-P3	0.25-3	0.706	-0.29	0.085
wPLI	F3-P3	15-30	0.044	-0.15	0.368
wPLI	F3-P3	3-8	0.470	-0.03	0.882
wPLI	F3-P4	8-15	0.826	0.24	0.154
wPLI	F3-P4	0.25-3	0.215	-0.11	0.539
wPLI	F3-P4	15-30	0.551	0.23	0.179
wPLI	F3-P4	3-8	0.102	-0.14	0.431
wPLI	F4-P3	8-15	0.014	0.00	0.981
wPLI	F4-P3	0.25-3	0.551	-0.21	0.220
wPLI	F4-P3	15-30	0.362	0.05	0.779
wPLI	F4-P3	3-8	0.167	0.22	0.191
wPLI	F4-P4	8-15	0.423	-0.18	0.303
wPLI	F4-P4	0.25-3	0.850	0.06	0.714
wPLI	F4-P4	15-30	0.660	-0.07	0.676
wPLI	F4-P4	3-8	1.000	0.14	0.430
wPLI	P3-P4	8-15	0.307	-0.17	0.324
wPLI	P3-P4	0.25-3	0.414	0.08	0.638
wPLI	P3-P4	15-30	0.432	-0.07	0.689
wPLI	P3-P4	3-8	0.802	0.22	0.195
NC	F3	8-15	0.001	0.32	0.060
NC	F3	15-30	0.002	0.33	0.048
NC	F3	3-8	0.000	0.31	0.064
NC	F4	8-15	0.001	0.15	0.373
NC	F4	15-30	0.005	0.30	0.074
NC	F4	3-8	0.000	0.36	0.033
NC	P3	8-15	0.037	0.33	0.048
NC	P3	15-30	0.480	0.38	0.021
NC	P3	3-8	0.002	0.54	0.001
NC	P4	8-15	0.041	0.51	0.001
NC	P4	15-30	0.084	0.58	0.000
NC	P4	3-8	0.000	0.59	0.000
ASI	F3P3-F4P4	1.5-25	0.649	0.36	0.030
peak hq	F3	8-15	0.000	0.18	0.298
peak hq	F3	0.25-3	0.251	0.10	0.575
peak hq	F3	15-30	0.000	0.57	0.000
peak hq	F3	3-8	0.182	0.62	0.000
peak hq	F4	8-15	0.000	0.33	0.051
peak hq	F4	0.25-3	0.519	0.12	0.493
peak hq	F4	15-30	0.000	0.80	0.000
peak hq	F4	3-8	0.379	0.62	0.000
peak hq	P3	8-15	0.000	0.43	0.009
peak hq	P3	0.25-3	0.660	0.12	0.490
peak hq	P3	15-30	0.000	0.57	0.000
peak hq	P3	3-8	0.245	0.41	0.012
peak hq	P4	8-15	0.000	0.54	0.001
peak hq	P4	0.25-3	0.016	0.03	0.858
peak hq	P4	15-30	0.000	0.59	0.000
peak hq	P4	3-8	0.582	0.44	0.008

Feature	Channel	Frequency band (Hz)	Wilcoxon signed rank p-value	Pearson's correlation coefficient	Correlation coefficient pvalue
width hq	F3	8-15	0.048	0.05	0.776
width hq	F3	0.25-3	0.900	0.05	0.785
width hq	F3	15-30	0.593	0.06	0.720
width hq	F3	3-8	0.057	-0.14	0.428
width hq	F4	8-15	0.041	-0.23	0.169
width hq	F4	0.25-3	0.460	-0.12	0.477
width hq	F4	15-30	0.251	-0.02	0.913
width hq	F4	3-8	0.330	0.14	0.402
width hq	P3	8-15	0.362	0.01	0.966
width hq	P3	0.25-3	0.826	0.10	0.563
width hq	P3	15-30	0.128	0.09	0.621
width hq	P3	3-8	0.090	0.16	0.354
width hq	P4	8-15	0.405	-0.28	0.097
width hq	P4	0.25-3	0.144	-0.26	0.132
width hq	P4	15-30	0.706	0.03	0.877
width hq	P4	3-8	0.040	-0.08	0.634
tail Dq	F3	8-15	0.000	0.04	0.802
tail Dq	F3	0.25-3	0.489	0.20	0.245
tail Dq	F3	15-30	0.220	0.01	0.949
tail Dq	F3	3-8	0.177	-0.15	0.379
tail Dq	F4	8-15	0.002	-0.29	0.084
tail Dq	F4	0.25-3	0.034	0.23	0.175
tail Dq	F4	15-30	0.220	0.15	0.385
tail Dq	F4	3-8	0.338	0.03	0.842
tail Dq	P3	8-15	0.300	-0.15	0.381
tail Dq	P3	0.25-3	0.850	0.01	0.955
tail Dq	P3	15-30	0.226	-0.01	0.943
tail Dq	P3	3-8	0.002	-0.14	0.427
tail Dq	P4	8-15	0.120	-0.32	0.058
tail Dq	P4	0.25-3	0.838	0.03	0.856
tail Dq	P4	15-30	0.604	-0.03	0.856
tail Dq	P4	3-8	0.018	0.37	0.026
height Dq	F3	8-15	0.000	0.10	0.577
height Dq	F3	0.25-3	0.637	0.10	0.571
height Dq	F3	15-30	0.102	0.13	0.454
height Dq	F3	3-8	0.040	-0.11	0.540
height Dq	F4	8-15	0.001	-0.34	0.040
height Dq	F4	0.25-3	0.441	0.04	0.828
height Dq	F4	15-30	0.084	0.17	0.329
height Dq	F4	3-8	0.099	0.09	0.617
height Dq	P3	8-15	0.113	-0.19	0.267
height Dq	P3	0.25-3	0.900	0.04	0.801
height Dq	P3	15-30	0.011	0.15	0.392
height Dq	P3	3-8	0.004	0.07	0.670
height Dq	P4	8-15	0.071	-0.30	0.078
height Dq	P4	0.25-3	0.059	-0.05	0.788
height Dq	P4	15-30	0.414	0.06	0.734
height Dq	P4	3-8	0.018	0.03	0.856

Feature	Channel	Frequency band (Hz)	Wilcoxon signed rank p-value	Pearson's correlation coefficient	Correlation coefficient pvalue
PSD	F3	8-15	0.000	0.58	0.000
PSD	F3	0.25-3	0.660	0.07	0.679
PSD	F3	15-30	0.660	0.46	0.005
PSD	F3	3-8	0.024	0.55	0.001
PSD	F4	8-15	0.000	0.64	0.000
PSD	F4	0.25-3	0.950	0.09	0.607
PSD	F4	15-30	0.362	0.40	0.016
PSD	F4	3-8	0.004	0.67	0.000
PSD	P3	8-15	0.000	0.81	0.000
PSD	P3	0.25-3	0.706	0.07	0.679
PSD	P3	15-30	0.604	0.53	0.001
PSD	P3	3-8	0.010	0.79	0.000
PSD	P4	8-15	0.000	0.78	0.000
PSD	P4	0.25-3	0.637	0.69	0.000
PSD	P4	15-30	0.480	0.60	0.000
PSD	P4	3-8	0.001	0.76	0.000
CPSD	F3-F4	8-15	0.000	0.86	0.000
CPSD	F3-F4	0.25-3	0.258	0.58	0.000
CPSD	F3-F4	15-30	0.388	0.97	0.000
CPSD	F3-F4	3-8	0.006	0.84	0.000
CPSD	F3-P3	8-15	0.000	0.92	0.000
CPSD	F3-P3	0.25-3	0.239	0.85	0.000
CPSD	F3-P3	15-30	0.396	0.98	0.000
CPSD	F3-P3	3-8	0.000	0.77	0.000
CPSD	F3-P4	8-15	0.001	0.90	0.000
CPSD	F3-P4	0.25-3	0.330	0.88	0.000
CPSD	F3-P4	15-30	0.187	0.98	0.000
CPSD	F3-P4	3-8	0.001	0.81	0.000
CPSD	F4-P3	8-15	0.000	0.90	0.000
CPSD	F4-P3	0.25-3	0.451	0.83	0.000
CPSD	F4-P3	15-30	0.140	0.98	0.000
CPSD	F4-P3	3-8	0.000	0.84	0.000
CPSD	F4-P4	8-15	0.000	0.89	0.000
CPSD	F4-P4	0.25-3	0.338	0.84	0.000
CPSD	F4-P4	15-30	0.338	0.98	0.000
CPSD	F4-P4	3-8	0.000	0.80	0.000
CPSD	P3-P4	8-15	0.000	0.93	0.000
CPSD	P3-P4	0.25-3	0.027	0.88	0.000
CPSD	P3-P4	15-30	0.120	0.98	0.000
CPSD	P3-P4	3-8	0.000	0.88	0.000

B Significant results for QS variables

Statistically significant results from the Wilcoxon rank sum test in QS for each HIE gradus and outcome group pair. We used Bonferroni correction for adjusting the confidence level with the number of individual comparisons – instead of testing if the feature differs between any of the three groups, we tested each pair individually. This means that the used confidence level of $\alpha = 0.05$ should be lowered to $0.05/3 \approx 0.017$. Grey background indicates a p-value $< .05$ while light red indicates $< .017$. The rightmost column shows the relative variances with values > 0.2 highlighted in grey.

Feature	Channel	Frequency band (Hz)	GRADUS			OUTCOME			Relative variance
			I - II	II - III	I - III	1-2	2-3	1-3	
AEC	F3-F4	3-8	0.019	0.663	0.145	0.889	0.772	0.946	0.00
AEC	F3-P3	8-15	0.466	0.945	0.278	0.043	0.105	0.911	0.00
AEC	F3-P3	0.25-3	0.010	0.033	0.642	0.764	0.247	0.252	0.00
AEC	F3-P3	15-30	0.378	0.070	0.028	0.307	0.056	0.200	0.00
AEC	F3-P3	3-8	0.024	0.077	0.642	0.327	1.000	0.406	0.00
AEC	F3-P4	0.25-3	0.029	0.002	0.780	0.920	0.093	0.200	0.01
AEC	F3-P4	3-8	0.038	0.224	0.227	0.889	1.000	0.805	0.01
AEC	F4-P3	3-8	0.038	0.281	0.515	0.238	0.487	0.875	0.00
PLV	F3-F4	15-30	0.788	0.010	0.182	0.589	0.297	0.216	0.00
PLV	F3-P4	0.25-3	0.024	0.063	0.780	0.002	0.037	0.458	0.01
PLV	F3-P4	3-8	0.038	0.224	0.687	0.167	0.093	0.946	0.00
PLV	F4-P3	3-8	0.046	0.207	0.687	0.704	0.385	0.736	0.00
wPLI	F3-F4	0.25-3	0.539	0.113	0.369	0.307	0.049	0.185	0.03
wPLI	F3-F4	15-30	0.318	0.094	0.024	0.704	0.271	0.157	0.00
wPLI	F3-F4	3-8	0.099	0.801	0.402	0.039	0.562	0.669	0.02
wPLI	F3-P4	8-15	0.466	0.191	0.278	0.484	0.049	0.076	0.01
wPLI	F3-P4	15-30	0.135	0.033	0.182	0.952	0.271	0.406	0.01
wPLI	F4-P3	0.25-3	0.234	0.008	0.088	0.180	0.487	0.007	0.03
wPLI	F4-P3	15-30	0.848	0.566	0.687	0.063	0.028	0.381	0.00
wPLI	F4-P3	3-8	0.010	0.801	0.088	0.289	1.000	0.357	0.02
wPLI	F4-P4	8-15	0.008	0.051	0.003	0.222	0.148	0.033	0.01
wPLI	P3-P4	8-15	0.026	0.012	0.515	0.674	0.105	0.144	0.01
wPLI	P3-P4	15-30	0.205	0.029	0.010	0.645	0.072	0.033	0.01
wPLI	P3-P4	3-8	0.539	0.148	0.059	0.238	0.037	0.003	0.02
NC	F3	8-15	0.818	0.077	0.100	0.032	1.000	0.084	0.01
NC	P4	8-15	0.300	0.242	0.828	0.075	0.028	0.637	0.01
NC	P4	15-30	0.420	0.136	0.306	0.089	0.011	0.132	0.01
NC	P4	3-8	0.250	0.302	0.877	0.043	0.082	0.605	0.01
peak hq	F3	3-8	0.032	0.029	0.556	0.347	0.325	0.669	0.00
peak hq	F4	3-8	0.046	0.012	0.251	0.952	0.247	0.200	0.00
peak hq	P3	3-8	0.032	0.261	0.515	0.075	0.325	0.605	0.00
width hq	F3	8-15	0.192	0.041	0.369	0.535	0.020	0.045	0.02
width hq	P4	15-30	0.084	0.011	0.598	0.039	0.020	0.669	0.01
width hq	P4	3-8	0.125	0.020	0.003	0.254	0.354	0.033	0.02
tail Dq	F3	8-15	0.730	0.018	0.077	0.826	0.005	0.021	*0.25*
tail Dq	F3	15-30	0.939	0.029	0.059	0.645	0.105	0.062	0.17
tail Dq	F4	8-15	0.283	0.191	0.114	0.271	0.202	0.004	*0.23*
tail Dq	P3	8-15	0.300	0.176	0.032	0.920	0.118	0.062	0.09
tail Dq	P4	8-15	0.337	0.029	0.051	0.734	0.028	0.004	0.12
tail Dq	P4	15-30	0.645	0.011	0.051	0.589	0.385	0.045	0.15
tail Dq	P4	3-8	0.205	0.033	0.024	0.589	0.118	0.170	0.16
height Dq	F3	8-15	0.490	0.002	0.028	0.535	0.001	0.002	0.06
height Dq	F3	15-30	0.818	0.041	0.129	0.704	0.148	0.271	0.08
height Dq	F4	8-15	0.591	0.176	0.204	0.289	0.132	0.004	0.07
height Dq	F4	0.25-3	0.514	0.029	0.251	0.509	0.224	0.084	0.09
height Dq	F4	15-30	0.099	0.449	0.077	0.389	0.202	0.033	0.07
height Dq	P3	8-15	0.730	0.103	0.145	0.368	0.037	0.234	0.07
height Dq	P4	8-15	0.818	0.010	0.067	0.889	0.015	0.002	0.07
height Dq	P4	15-30	0.220	0.004	0.100	0.509	0.082	0.132	0.05
height Dq	P4	3-8	0.135	0.003	0.002	0.826	0.118	0.023	0.07

PSD	F3	8-15	0.399	0.001	0.000	0.180	0.009	0.009	*0.28*
PSD	F3	0.25-3	0.283	0.018	0.003	0.238	0.064	0.002	*1694.43*
PSD	F3	15-30	0.125	0.010	0.001	0.734	0.072	0.037	0.08
PSD	F3	3-8	0.135	0.002	0.000	0.327	0.006	0.007	*3.52*
PSD	F4	8-15	0.234	0.001	0.000	0.734	0.020	0.018	*0.30*
PSD	F4	15-30	0.116	0.005	0.000	0.826	0.037	0.016	0.08
PSD	F4	3-8	0.084	0.004	0.001	0.889	0.043	0.051	*3.22*
PSD	P3	8-15	0.125	0.006	0.000	0.254	0.005	0.026	*0.26*
PSD	P3	0.25-3	0.050	0.006	0.001	0.562	0.005	0.002	*1604.87*
PSD	P3	15-30	0.035	0.070	0.002	0.459	0.056	0.121	0.07
PSD	P3	3-8	0.167	0.005	0.000	0.435	0.006	0.033	*2.86*
PSD	P4	8-15	0.099	0.000	0.000	0.412	0.001	0.001	*0.20*
PSD	P4	0.25-3	0.156	0.002	0.000	0.238	0.004	0.000	*7698.66*
PSD	P4	15-30	0.078	0.001	0.000	0.764	0.006	0.003	0.07
PSD	P4	3-8	0.078	0.000	0.000	0.734	0.001	0.001	*3.47*
CPSD	F3-F4	8-15	0.514	0.001	0.000	0.484	0.056	0.016	0.00
CPSD	F3-F4	0.25-3	0.107	0.006	0.000	0.704	0.011	0.003	*0.33*
CPSD	F3-F4	15-30	0.145	0.006	0.001	0.857	0.056	0.041	0.00
CPSD	F3-F4	3-8	0.618	0.001	0.000	0.435	0.009	0.013	0.02
CPSD	F3-P3	8-15	0.084	0.002	0.000	0.435	0.005	0.016	0.00
CPSD	F3-P3	0.25-3	0.046	0.016	0.001	0.984	0.049	0.011	*0.28*
CPSD	F3-P3	15-30	0.060	0.094	0.004	0.617	0.056	0.157	0.00
CPSD	F3-P3	3-8	0.179	0.003	0.000	0.412	0.001	0.023	0.01
CPSD	F3-P4	8-15	0.107	0.001	0.000	0.459	0.002	0.013	0.01
CPSD	F3-P4	0.25-3	0.029	0.113	0.008	0.509	0.354	0.062	*0.70*
CPSD	F3-P4	15-30	0.026	0.006	0.000	0.674	0.011	0.033	0.00
CPSD	F3-P4	3-8	0.107	0.003	0.000	0.674	0.001	0.041	0.03
CPSD	F4-P3	8-15	0.167	0.001	0.000	0.238	0.003	0.009	0.00
CPSD	F4-P3	0.25-3	0.055	0.506	0.007	0.704	0.451	0.170	*0.36*
CPSD	F4-P3	15-30	0.035	0.005	0.001	0.562	0.028	0.029	0.00
CPSD	F4-P3	3-8	0.145	0.001	0.000	0.289	0.002	0.007	0.02
CPSD	F4-P4	8-15	0.156	0.000	0.000	0.459	0.003	0.002	0.00
CPSD	F4-P4	0.25-3	0.021	0.002	0.001	0.412	0.072	0.001	*38.82*
CPSD	F4-P4	15-30	0.116	0.006	0.002	0.795	0.064	0.041	0.00
CPSD	F4-P4	3-8	0.084	0.000	0.000	0.645	0.001	0.001	0.02
CPSD	P3-P4	8-15	0.250	0.002	0.000	0.193	0.001	0.004	0.00
CPSD	P3-P4	0.25-3	0.145	0.037	0.003	0.389	0.006	0.023	*0.29*
CPSD	P3-P4	15-30	0.099	0.001	0.001	0.368	0.002	0.002	0.00
CPSD	P3-P4	3-8	0.337	0.006	0.001	0.207	0.003	0.010	0.02

C Significant results for AS variables

Statistically significant results from the Wilcoxon rank sum test for AS. The p-values are presented for each HIE gradus pair and outcome group pair. We used Bonferroni correction for adjusting the confidence level as described in Appendix B. Grey background indicates a p-value that is smaller than 0.05 while light red indicates smaller than 0.017. On the rightmost column presents the relative variances and the ones that are over 0.2 are highlighted in grey.

Feature	Channel	Frequency band (Hz)	GRADUS			OUTCOME			Relative variance
			I - II	II - III	I - III	1-2	2-3	1-3	
AEC	F3-F4	15-30	0.006	0.316	0.714	0.688	0.264	0.413	0.00
AEC	F3-F4	3-8	0.002	1.000	0.429	0.949	0.440	0.488	0.00
AEC	F3-P3	8-15	0.026	0.526	0.571	0.300	0.923	0.850	0.01
AEC	F3-P3	15-30	0.007	0.421	0.571	0.568	0.791	0.659	0.00
AEC	F3-P3	3-8	0.008	0.526	0.857	0.657	0.923	0.659	0.01
AEC	F3-P4	0.25-3	0.039	0.526	0.286	0.244	0.440	1.000	0.02
AEC	F3-P4	15-30	0.004	0.316	0.714	0.434	1.000	0.659	0.01
AEC	F3-P4	3-8	0.017	0.632	0.857	0.626	0.549	0.413	0.01
AEC	F4-P3	8-15	0.036	0.421	0.286	0.079	0.791	0.659	0.01
AEC	F4-P3	15-30	0.010	0.316	0.571	0.568	0.549	0.571	0.01
AEC	F4-P3	3-8	0.007	0.526	0.857	0.816	0.923	0.753	0.01
AEC	F4-P4	15-30	0.036	0.421	0.429	0.320	0.791	0.950	0.01
AEC	F4-P4	3-8	0.009	0.632	1.000	0.816	0.923	0.850	0.01
PLV	F3-F4	3-8	0.036	0.526	0.286	0.949	0.791	0.488	0.00
PLV	F3-P3	3-8	0.012	0.632	0.286	0.719	0.791	0.488	0.00
PLV	F3-P4	0.25-3	0.002	0.632	0.429	0.060	0.791	0.284	0.01
PLV	F3-P4	3-8	0.015	0.211	0.286	0.688	0.659	0.345	0.00
PLV	F4-P3	0.25-3	0.036	0.105	0.286	0.212	0.198	1.000	0.01
PLV	F4-P3	3-8	0.004	1.000	0.429	0.916	0.791	0.950	0.00
PLV	F4-P4	3-8	0.024	0.947	0.571	0.882	0.659	0.488	0.00
wPLI	F3-F4	0.25-3	0.327	0.421	0.143	0.816	0.022	0.051	0.04
wPLI	F3-F4	3-8	0.048	1.000	0.429	0.144	0.659	0.753	0.01
wPLI	F3-P4	15-30	0.389	0.526	0.714	0.751	0.044	0.147	0.00
wPLI	F4-P4	15-30	0.043	0.632	1.000	0.434	0.791	0.950	0.01
wPLI	P3-P4	0.25-3	0.764	0.211	0.429	0.026	0.044	0.488	0.02
NC	F3	8-15	0.133	0.842	0.571	0.112	0.044	0.089	0.01
NC	P3	15-30	0.412	0.211	0.286	0.019	0.549	0.068	0.01
NC	P4	3-8	0.024	0.737	0.286	0.122	0.659	0.659	0.01
peak hq	P3	3-8	0.011	0.737	1.000	0.182	0.264	0.186	0.00
width hq	F3	3-8	0.029	0.421	0.571	0.539	0.659	0.950	0.01
width hq	P4	3-8	0.412	0.526	0.286	0.983	0.022	0.051	0.02
tail Dq	F3	15-30	0.003	0.316	1.000	0.751	0.923	0.659	0.08
tail Dq	P3	15-30	0.032	0.947	0.286	0.197	0.088	0.345	0.19
height Dq	F3	15-30	0.029	0.105	0.429	0.539	0.549	0.850	0.02
height Dq	P3	15-30	0.459	0.842	0.714	0.122	0.044	0.068	0.06
PSD	F4	8-15	0.412	0.526	0.286	0.244	0.132	0.038	*0.52*
CPSD	F3-F4	8-15	0.389	0.842	0.286	0.688	0.088	0.027	0.01
CPSD	F3-P3	0.25-3	0.043	0.105	0.286	0.626	0.132	0.231	*1.76*
CPSD	F3-P3	3-8	0.043	0.737	0.286	0.882	0.549	0.659	0.09
CPSD	F3-P4	0.25-3	0.019	0.316	0.286	0.485	0.791	0.488	*1.20*
CPSD	F3-P4	15-30	0.015	0.632	0.143	0.751	0.549	0.850	0.00
CPSD	F4-P3	8-15	0.036	0.947	0.429	0.719	0.440	0.284	0.01
CPSD	F4-P3	0.25-3	0.017	0.947	0.429	0.539	0.264	0.186	*2.79*
CPSD	F4-P3	15-30	0.032	0.632	0.143	0.657	0.791	0.413	0.00
CPSD	F4-P3	3-8	0.024	0.947	0.143	0.783	0.440	0.488	0.09
CPSD	F4-P4	0.25-3	0.048	0.842	0.429	0.783	1.000	0.950	*1.31*
CPSD	F4-P4	3-8	0.048	0.526	0.286	1.000	0.791	0.753	0.08