THE UNIVERSITY OF CHICAGO

NEUROPHYSIOLOGIC MONITORING AND PROGNOSTIC ALGORITHMS FOR NEONATAL HYPOXIC ISCHEMIC ENCEPHALOPATHY

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

INTERDISCIPLINARY SCIENTIST TRAINING PROGRAM:
NEUROBIOLOGY

BY

SYLVIA UWAILA EDOIGIAWERIE

CHICAGO, ILLINOIS
AUGUST 2023

To my loved ones and my future self who dares to see past her limitations – both real and imagined.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am truly grateful for all the people that have been a part of this journey. First and foremost, I want to thank God for seeing me through this process by grace through faith. Next, I want to thank the members of my committee and my collaborators. It is only right for me to begin with my advisor, Dr. Norm Issa for his patience, his guidance, and for believing in me. His clinical and research expertise on epilepsy has been invaluable and his thorough and empathetic approach to research and work has made completing a challenging PhD a joyous experience. I want to thank Dr. Julia Henry and Dr. Henry David both affectionately referred to as the Henries[2] for their guidance during this process. Julie has always modeled poise, perseverance, and organization, and I am fortunate to have had the privilege to work with her. Hank never fails to make me laugh with his big personality, never ending supply of dad jokes, helpful spirit, and tireless positivity. I also want to thank Dr. Nicholas Hatsopoulos and Dr. Leo Towle for their kindness and compassion and sticking with me through the PhD journey. I started this journey just before the pandemic, and I am glad to have had their support during such a challenging time. Finally, I want to thank my collaborator Dr. Brett Beaulieu-Jones, his guidance has been invaluable for completing the seizure detection work.

I also want to thank those with whom I shared the research space with – namely the pediatrics department and the Voss lab. Thank you to the Voss lab for giving me a space to work and to Bobby and James, who were there when I first came, for always having a wealth of knowledge and positive energy. Next, I want to thank the section chief of the UChicago Pediatrics department, Dr. Douglas Nordli, along with Dr. Carol McMillan, Dr. Chalongchai Phitsanuwong for encouraging me even when it seemed like I was squatting in the peds neuro space because I

was on campus more than my own apartment at one point. Additionally, I want to thank the head EEG tech, Celeste, as well as Chris, Nelson, Brianna, and the entire technical support staff for always being helpful, never getting annoyed when I would ask a thousand times to be let into the EEG reading rooms, and always having snacks and fun keto diet stories to tell. I also want to thank Janice, Jackie, and Sarah for making the early morning and afternoon coffee runs fun. Last but certainly not least in the research space, I want to thank Dr. Nancy Schwartz for her kindness, support, and always encouraging me to bring the best out of myself.

I also want to thank my loved ones and friends for their unwavering support. I want to thank Chimmy, Bree, Ada, and Taylor for their continuous support and for understanding the highs and woes of grad school life. My baby sister for her supportive sass and for inspiring me with her Gen Z wisdom. Finally, I want to thank my parents for reminding me that there is always dignity in hard work, for instilling me with Faith, and for teaching me that both small wins and failures are necessary steppingstones on the path to success.

ABSTRACT

Across all stages of life, infants are amongst the most susceptible to brain injury. This vulnerability is particularly pronounced in a condition called Hypoxic Ischemic Encephalopathy (HIE), where lack of oxygen during the birthing process leads to encephalopathy and neuronal cell death. Clinicians rely on neuromonitoring modalities particularly electroencephalography (EEG) and magnetic resonance imaging (MRI) during the first week of life to assess HIE severity and detect and manage seizures that result from HIE.

This body of work focuses on extracting features from MRI and EEG to prognosticate infant outcome and detect seizures. In the first part of the project that focuses on prognostication, it was found that a combination of cortical insula injury and absence of EEG state change activity could predict abnormal outcome in neonates between three and six months of life. It was also found that periventricular white matter injury correlated significantly with abnormal seizure activity and poor outcomes in term neonates.

In the second part of the project that focuses on seizure detection, it was found that a combination of features from the amplitude-integrated EEG and compressed spectral array could detect seizures with high accuracy using an external dataset of 79 patients with heterogenous disease etiology. Results also revealed that the algorithm performance was comparable in HIE patients within the dataset. A follow-up algorithm validation study using the inhouse cohort of term HIE infants, showed that the algorithm could perform well on an independent HIE cohort and equally as well if only trained using the subset of external data patients that had HIE due to birth asphyxia. In conclusion, this body of work shows that early clinical management for neonatal patients can be improved using a combination of features from EEG and MRI to score outcomes and a combination of quantitative EEG features to better detect seizures.

CHAPTER 1

INTRODUCTION

1.1 Epidemiology and Outcomes of Neonatal HIE

Hypoxic Ischemic Encephalopathy (HIE) represents a significant global disease burden. With an estimated incidence of 1.5 per 1000 births, it accounts for nearly a quarter of infant deaths worldwide [1,2]. In addition to mortality, nearly 60% of HIE cases result in severe disabilities such as cerebral palsy [3,4]. Disability Adjusted Life Years (DALYs) are markers used to measure both premature mortality rate of a condition and the years of life spent in suboptimal health due to that condition[5]. Birth asphyxia is among the top three global causes of DALYs, and in 2010 was estimated to be the cause of 42 million DALYs, which by comparison accounts for twice the estimated disability burden of diabetes [6,7].

1.2 Pathophysiology of Neonatal HIE

HIE in the term infant results from prolonged intrauterine hypoxia [3,4]. Hypoxia specifically refers to impaired cerebral perfusion that leads to a proinflammatory cascade that causes encephalopathy[8]. Perinatal hypoxia has a variety of causes. These include but are not limited to cord prolapse, fetal entrapment, maternal hypotension, and placental abruption[3]. These events lead to loss of cerebral blood flow during the birthing process which precipitates neuronal ischemia. HIE pathophysiology can be divided into four stages: Primary Energy Failure, Secondary Energy Failure, Latent, and Tertiary [3].

Primary energy failure occurs immediately after hypoxic injury. In this stage, lack of oxygen leads to a shift towards anaerobic metabolism. If hypoxia is prolonged, anaerobic

metabolism progresses into excitotoxicity as the Na/K+ pumps fail due to ATP loss [8,9]. ATP loss precipitates a proinflammatory cascade that leads to intracellular sodium accumulation and excessive glutamate release. The result is excitotoxity and neuronal cell death via both apoptosis and necrosis[8,9]. Primary energy failure is followed by the latent phase [9]. During the latent phase, implementation of neuroprotective interventions can help to reduce the magnitude of neuronal cell loss [9].

The latent phase extends into the beginning of the next stage, which is secondary energy failure. Secondary energy failure is characterized by delayed neuronal cell death, cytotoxic edema, excitotoxicity, and microglial activation [8,9]. Secondary energy failure extends from six to 48 hours after injury. The majority of neonatal seizures often occur during this stage[10]. In a subset of patients, tertiary brain injury follows secondary energy failure. Tertiary brain injury includes astrogliosis and reduced neuronal cell count that persists in the weeks to years following the initial hypoxic event [8,9].

The neonatal brain is especially susceptible to seizures during this period due to GABA-mediated excitation [11]. This paradoxical excitatory phenomenon has been attributed to relatively low expression of the potassium chloride transporter, KCC2, that facilitates chloride export from the cell and leads to leads to a buildup of intracellular chloride concentration. The result is that ionotropic GABA receptors allow chloride efflux and cause increased neuronal excitability [11]. In HIE, hypoxic injury leads to energy failure due to diminished ATP, which triggers an inflammatory cascade characterized by failure of ion pumps and calcium-mediated excitotoxicity [3]. In this neuronal environment, seizure activity can proliferate.

1.3 Therapeutic Hypothermia

Therapeutic Hypothermia (TH) is the first line of treatment for term infants with HIE following the hypoxic event. This treatment involves whole-body cooling to lower infants' core temperature to between 33.5 and 34.5 degrees Celsius for 72 hours[12]. The goal is to reduce secondary injury and the resultant proinflammatory cascade. During TH, clinicians depend on EEG and MRI to assess the infant's disease severity and response to treatment.

Studies have investigated the efficacy of therapeutic hypothermia for improving long-term neurodevelopmental outcomes. The largest of such studies is the Total Body Hypothermia (TOBY) Study[13]. This multicenter, prospective, randomized study tested the influence of therapeutic hypothermia on neurodevelopmental outcome in term neonates. They found that term infants with hypoxic ischemic encephalopathy who received 72-hour therapeutic hypothermia at less than six hours of age saw improvements in the outcomes using the Bayle Scales of Infant Development II. TH has also been shown to decrease the risk of poor outcomes such as cortical visual impairment in term neonatates[14].

Most studies to date have focused on therapeutic hypothermia for near-term and term neonates. The minimum recommended gestational age for therapeutic hypothermia is often either 36 weeks and above or 35 weeks and above. Studies that have assessed therapeutic hypothermia in preterm neonates (33-35 weeks gestational age) have found largely mixed results with both a potentially negative effect and no effect[15–17]. This may be due to the differing metabolic demands of the preterm brain and the disruption of oligodendrocyte myelination that is unique to the preterm brain in HIE[15]. Ongoing work is being done particularly by the NICHD Neonatal Research Network to conduct a randomized controlled trial (ID.: CN-01540535) to assessed the efficacy of therapeutic hypothermia in preterm (33-35 weeks gestational age) infants[18].

1.4 Bayley Scales of Infant Development

The severity of HIE outcomes in infants is generally measured using functional outcomes metrics. The Bayley Scales of Infant Development (BSID) at 18 to 24 months is currently one of the most widely used standardized assessment tool for neurodevelopmental outcomes. This metric gauges developmental delay using a series of behavioral tests and patient questionnaires[19]. A child's development index is compared to normative age-matched samples to assess for developmental delay. The Bayley scales were initially developed by Nancy Bayley, a pediatrician, in 1969[19]. Since then, the BSID has been revised three times first in 1993, then 2006, and finally in 2019 to generate the BSID-II, BSID-III, and BSID-IV respectively[19]. The papers cited in this work use the BSID-II and BSID-III. BSID-II measures both cognitive and motor development using a Mental Development Index (MDI) and a Psychomotor Development Index (PDI). Bayley-III was introduced to include subscales to separate cognitive and language skills in the MDI by separately measuring cognitive, receptive language, and expressive language scales[20]. The PDI was also further separated to evaluate fine and gross motor skills independently. The BSID-III is still the most widely used, but a notable concern is that it appears to underestimate rates of developmental delay[20,21].

One of the major downsides to outcome metrics such as the Bayley Scales is the need for follow-up at one to two years of age. Studies have cited follow-up rates among neonates within the first year as often well below 40% [22,23]. Factors such as loss of contact and socioeconomic status have been shown to play a role. High attrition rates are particularly concerning because assessments such as the Bayley scales are not objective but based on normative scores from the population of children that do receive follow-up care. Orton et al 2015 showed that rates of developmental delay were greater among neonates retained in their study compared to those that

were retained as part of standard of care[23]. They argued that high attrition can create a bias in developmental outcome scores that are considered standard by clinicians.

1.5 Electroencephalography for Neuromonitoring and Seizure Detection

Infant brain activity must be monitored in HIE to direct clinical management and follow the evolution of hypoxic injury. The gold standard neuromonitoring tool after hypoxic insult and especially during therapeutic hypothermia is electroencephalography (EEG). EEG displays brain activity using scalp electrodes that measure neuronal postsynaptic potentials[24]. The electrical activity comes primarily from the synchronous postsynaptic potentials of cortical pyramidal neurons[24]. Neonates have higher scalp conductivity than adults[25]. Increased conductivity may be due to both the neonatal skull being thinner and the presence of skull fontanelles – areas where the skull has not fused[26]. The neonatal EEG can be visualized using the standard 10-20 system or a modified version of the 10-20 system that concentrates electrode placement near the centrotemporal and centroparietal regions of the brain[27]. Electrode placement is particularly important in the centrotemporal area because it captures activity from the Rolandic and peri-Rolandic cortices, which are known to be among the most metabolically active regions of the term neonatal brain[28]. These regions have also been shown to be more sensitive for capturing important clinical events such as seizures[29].

The normal term and near-term neonatal EEG contains certain patterns of activity that indicate adequate brain development. These background patterns include EEG synchrony, presence of state changes such as quiet sleep and active sleep, and the presence of specific EEG graphoelements such as a trace alternans pattern and enconches frontales[30]. Term and near-term neonatal EEGs should show synchronous and symmetrical EEG activity, which is characterized

by equivalent timing of burst activity across both hemispheres and equal hemispheric power respectively with a tolerance of 1.5 seconds[30]. Discontinuity should be primarily seen during the trace alternans pattern, which is exemplified by synchronous bursts (50–150μV) with interburst amplitudes greater than 25μV [30]. Trace alternans pattern is seen during quiet sleep. Thus, while discontinuity is not pathologic in neonatal EEGs, term and near-term neonatal EEGs, should have brief interburst intervals. Typically, interburst intervals are around six seconds at full term (37 to 40 weeks) and up to 10 seconds in near term neonates (35 to 37 weeks) [30]. In contrast, during active sleep, higher amplitude (50–300μV) more continuous activity along with faster theta activity can be appreciated[30]. Finally, the encoches frontales pattern is characteristic of term and near-term neonates. Encoches frontales are transient discharges that occur in the frontal electrodes during sleep. These discharges are diphasic, synchronous, and symmetric[30].

In addition to observing the EEG background for evidence of normal brain development, clinicians are also tasked with identifying seizure and epileptiform activity using the EEG. Neonatal seizures are defined as rhythmic discharges with a clear evolution pattern that last 10 seconds or more. The evolution pattern typically involves a change in waveform amplitude, frequency, and/or morphology[30]. In contrast, nonseizure epileptiform activity presents as isolated sharp activity, brief runs of epileptiform activity, typically rhythmic activity with an evolution lasting less than 10 seconds, or periodic discharges, which are rhythmic discharges lasting more than 10 seconds but without a clearly evolving pattern[30].

1.5.1 Quantitative EEG

Despite manual EEG review being the gold standard for both neonatal neuromonitoring during therapeutic hypothermia and neonatal seizure detection, manual review has multiple

shortcomings. First, neonatal seizure detection is resource intensive. 24-hour long term EEG monitoring is the standard for all three days of therapeutic hypothermia, but it generates massive amounts of EEG data. To illustrate, long-term EEG monitoring, which is generally displayed with 20 seconds of EEG per page, generates 12,960 pages of EEG for a single neonate during therapeutic hypothermia. This process is also time-consuming because the data must be manually reviewed for assessment of background and seizure activity, which will be used to guide clinical management and prognosticate possible outcomes. Finally, clinicians with expertise in assessing neonatal seizures and background activity are scarce[31], which means the demand for sensitive and specific review of EEG data greatly exceeds the supply of clinicians equipped with the skills to fulfill it. As a result, many facilities are limited by reliance on remote EEG interpretation from an overtaxed pool of clinicians.[32] This shortage may slow access to care and lead to worsened outcomes. Thus, quantitative EEG algorithms (qEEG) have been designed to meet this demand. qEEG may expedite clinical management and provide relief for families burdened by the uncertainty of their infant's prognosis.

Since the purpose of qEEG algorithms is to simplify and expedite assessment of EEG, qEEG trending modalities are often viewed using a reduced electrode montage, specifically the two and four channel montage. The recommended limited electrode set up are the central and parietal electrodes, specifically C3-P3 and C4-P4[33]. This recommendation exists because the central electrodes are more sensitive to seizures especially in comparison the frontal electrodes, which are the most contaminated by artefacts and where seizures are more often missed [29,34].

1.5.2 Amplitude Integrated EEG

Amplitude integrated EEG (aEEG) is currently the most widely used quantitative algorithm in Neonatal Intensive Care Units (NICUs). aEEG was first designed by Douglas Maynard in 1969 for continuous monitoring of adult patients who had been placed under anesthesia or had suffered cardiac arrest and were receiving intensive[35] . The original monitoring device was referred to as the Cerebral Function Monitor. The aEEG algorithm is constructed by band-pass filtering the raw EEG generally between 2 and 15 Hz to reduce sweating and high frequency muscle and electrical artifact[36]. Then, a full-wave rectifier is implemented to get the amplitude information of the EEG, which is a biphasic signal[37]. The resultant signal is then smoothed or enveloped using a lowpass filter[37]. Then, it is segmented and placed on a semilogarithmic amplitude scale so higher amplitude changes specifically above 10μV (upper margin) appear less exaggerated in comparison to lower amplitude changes (lower margin)[36,37]. Finally, the signal is further time compressed by being displayed at 6cm/h at the bedside[36].

Interpretation of the aEEG is based on simple pattern recognition of amplitude fluctuations. Thus, the aEEG is easily learned, interpreted, and used by the clinical team.[38,39] Since the aEEG time compresses the signal, it allows clinicians to assess EEG background patterns over several hours at a time. Notably, the aEEG background pattern has been shown to have prognostic efficacy in neonates with HIE[40,41]. This utility exists because the aEEG background can be used to identify long period of abnormal voltage activity. Additionally, the aEEG can be used to identify seizures using voltage patterns in the upper and lower margins. To do so, clinicians look for an abrupt rise in the upper and lower margins to detect windows of time suspicious for seizure activity[42].

Despite these strengths, aEEG has poor sensitivity for seizure detection. In fact, multiple studies have cited aEEG as having seizure detection rates of well below 50 percent.[34,43–45]

Additionally, clinicians have cited shorter duration seizures (<30 seconds) as being more difficult to detect[34,38]. Since aEEG is designed to time compress the EEG, the result is loss of accurate information concerning the infant's seizure burden. In addition to the seizure detection problem, lack of consensus between proprietary aEEG algorithms makes aEEG less generalizable across studies.[34,46,47,43] Werther et al found that differences in the moving average filter window primarily account for discrepancies in the performance of proprietary aEEG algorithms for computational tasks.[48] In contrast, for outcome prognostication in HIE, aEEG shows more promise. The main downside is that aEEG's predictive capacity is more variable in neonates that receive cooling therapy.[49–51]

1.5.3 Compressed Spectral Array

Another widely used qEEG trending modality is the Compressed Spectral Array (CSA). The CSA displays frequency trends using spectral analysis[52,53]. Clinicians can view patterns of power from the delta to beta frequency bands (0.5-30Hz). The CSA is often displayed using a color-density spectrogram plot, but it can also be viewed as a waterfall plot. Despite its use in neonatal intensive care units, the CSA has primarily been studied in the adult population for long term monitoring and seizure detection. Notably, Williamson et al 2014 found that the CSA was able to detect 87.9 percent of seizures in a population of 118 critically ill adult patients[54]. The CSA has also been used for seizure detection in children (age range: 0.9-6.8 years) and was found to have higher specificity (95%) but lower sensitivity (64.8%) when reviewed by clinicians without the accompanying EEG. For the neonatal population, work pioneered by Gotman et al 1993 showed that spectral features might be viable tools for automated seizure detection in neonates[55].

1.6 Previous Neonatal Seizure Detection Algorithms

Neonatal seizure detection is particularly important in neonates with HIE because HIE causes the majority of neonatal seizures[56]. Additionally, seizures must be caught and treated quickly to reduce seizure burden and assess clinical management[57]. Thus, there has been a lot of effort aimed at creating algorithms for bedside neonatal seizure detection algorithms particularly from research groups from the Netherlands, Ireland, and Finland[55,58–60]. The primary challenges of currently existing algorithms have been reducing the false positive rate while maintaining adequate sensitivity and multichannel probabilistic detection of the neonatal seizures. The latter is particularly important because most neonatal seizures are focal thus the location of seizures will vary across different patients if a full electrode set is implemented for detection.

Since the focus of this body of work is on currently existing qEEG monitoring efficacy, algorithms that assess aEEG and CSA for neonatal seizure detection will be reviewed. While there have been a host of studies that manually assess the efficacy of these qEEG algorithms in neonates[34,43,61,62], only two algorithms will be addressed that attempted to computationally automate seizure detection using the aEEG and CSA[55,60]. The first was conducted by Lommen et al 2007[60]. They generated a seizure detection algorithm using changes in the baseline aEEG trace. They trained on five neonates and tested on eight. Ten of the infants had birth asphyxia, the other three had unknown diagnosis, hypoglycemia, and familial benign seizure. While they observed a sensitivity of above 90% with one false positive per hour, their methodology presents some caveats. First, they evaluated their algorithm based on seizures labeled by clinicians on the aEEG trace not on the EEG. EEG is the gold standard for neonatal seizure detection. Thus, their assessments primarily reflect how well their algorithm mimicked aEEG seizure patterns. Secondly, their sample size is quite small at only 13 patients total.

The second algorithm was an approach pioneered by Gotman in 1997 using 43 patients. This algorithm focuses on applying features of the CSA for seizure detection. Gotman assessed multiple features from the CSA and compared them across 10-second epochs to generate a computer algorithm using a combination of those features. What was key was Gotman's algorithms compared CSA features epochs to those in prior background epochs around 60s earlier. Gotman's best performing set of features detected 71% of seizures with a false detection rate of 1.7 per hour. With new methodologies such as machine learning that allows for weighted feature assessment, a more sensitive combination of spectral features can be extracted and evaluated for neonatal seizure detection.

1.7 Magnetic Resonance Imaging in Neonatal HIE

Magnetic Resonance Imaging is particularly important in neonates when evaluating adequate brain development and identifying lesions in HIE. MRI uses an external magnetic field gradient and radiofrequency pulses to generate an image of tissue. This image is based on the response time of protons within the tissue to the external electromagnetic field [63,64]. The most common imaging sequences used for HIE are the T1-Weighted, T2-Weighted, and Diffusion-Weighted Imaging (DWI)[65]. T1-weighted imaging is based on the longitudinal relaxation time (T1) or the time for protons excited by the external radiofrequency to go back to realigning with the external magnetic field. Tissues with shorter T1 such as white and gray matter appear brighter on T1-Weighted imaging in comparison to CSF which appears dark[64,65]. In contrast, T2-Weighted imaging is based on the transverse relaxation time (T2), and tissue areas with a longer T2 such as CSF will appear brighter than those with a shorter T2 like white and gray matter [64,65]. DWI-MRI takes a slightly different approach by measuring the rates of diffusion of protons within water

using a metric called the apparent diffusion coefficient (ADC). Thus, the ADC is particularly important for assessing ischemic neuronal tissue that would be susceptible to cytotoxic edema and restricted diffusion to due energy failure[66].

Several areas of the brain have been identified as particularly susceptible in HIE. These locations are the basal ganglia, thalamus, internal capsule, periventricular white matter, and cortex [67]. There are two common injury patterns of HIE in term neonates: the basal ganglia-thalamus pattern and the watershed pattern. The basal ganglia-thalamus pattern is generally associated with a severe, acute hypoxic event[67,68]. This pattern typically involves the bilateral ventrolateral thalami, posterior putamina, perirolandic cortex, and posterior limb of the internal capsule (PLIC). Lesions can be seen primarily as hyperintensities on T1-imaging with abnormal T2 hypointensity in the PLIC. DWI-MRI within the first week of birth typically shows restricted diffusion in these areas[69]. In contrast, the watershed injury pattern typically occurs after partial and prolonged hypoxia. If hypoxia is prolonged, cerebral autoregulation leads to redistribution of cerebral blood flow to areas that are most metabolically active in the term neonatal brain. This redistribution leads to injuries in the vascular watershed zones supplied by the middle cerebral artery[67]. These areas include the periventricular white matter and in more severe events subcortical white matter.

While the focus of this study is on term and near-term neonates who receive cooling (>35 weeks gestational age), preterm infants with HIE have characteristic injury patterns that are worth noting. Ostensibly, preterm infants suffer less basal ganglia injury following hypoxic injury and much more perirolanadic and white matter injury compared to term infants[65,67,69]. Basal ganglia myelination occurs at 33 to 35 weeks gestational age[70,71]. Myelination increases the metabolic demand, which may cause the basal ganglia to be more susceptible to hypoxic injury. Additionally, some studies have hinted that preterm neonates may be less susceptible to cortical glutamate

mediated excitotoxicity due to a lower density of Ca2+ permeable AMPA glutamate receptors in the preterm cortex[72,73].

1.8 Thesis Aims: Addressing Critical Gaps in Knowledge

The goal of this thesis is to generate algorithms that can be used for neuromonitoring and clinical management of term HIE patients particularly during therapeutic hypothermia. This objective led us to consider two of the most important clinical objectives following HIE diagnosis during cooling therapy: 1. Prognosticating neonatal outcomes and 2. Detecting neonatal seizures. This study aims to address the first objective by identifying sensitive electrographic and imaging biomarkers of outcome during cooling therapy. The two most important modalities currently used for evaluating infant's brain development during HIE are EEG and MRI. Clinicians synthesize important clinical information from both EEG and MRI to guide clinical management and counsel families. The first objective was to survey the features that have been shown to be useful for prognostication in EEG and MRI and generate a multimodal prognostic algorithm using important EEG and MRI features and an in-house cohort of term and near-term neonates with HIE.

The second objective was to address the ongoing challenge of neonatal seizures detection by addressing the shortcomings of currently existing qEEG monitoring algorithms based on aEEG and CSA. The approach is to generate seizure detection algorithms by extracting features that mimic the manual visual assessment done by clinicians. This study then couples these features with supervised machine learning algorithms for training and testing to generate sensitive and specific detection algorithms. Currently, no study addresses both aEEG's algorithm variability and poor seizure resolution. Studies that use aEEG often just implement the standard aEEG for manual review[74,75]. Other studies bypass aEEG completely and focus on applying signal analysis and

statistical metrics for either seizure detection or outcome prognostication[58,76–80]. Since both aEEG and the CSA are still part of standard care, they remain useful for their ease of implementation and familiarity to clinicians. Addressing their shortcomings may help improve their efficacy for automated detection. Additionally, while convolutional neural networks and other forms of unsupervised machine learning are increasingly being implemented for neonatal seizure detection, they present notable downsides[76,81,82]. The most salient downsides of this approach are the need to decode the black box of deep learning to make it more accessible for clinicians. Demystifying deep learning algorithms will require explainable artificial intelligence approaches, which are arguably still in their nascent stages[83]. Another is the need for very large datasets to prevent overfitting[84]. For these reasons, this thesis focuses on beginning with supervised machine learning and engineering features based on clinician's assessment's current qEEG that are part of standard care.

CHAPTER 2

EARLY EEG AND MRI FEATURES FOR HIE PROGNOSTICATION: A SYSTEMATIC REVIEW

2.1 Abstract

Objective

To review features in early EEG and MRI features associated with long-term neurological outcomes in infants after perinatal HIE.

Methods

Articles were extracted from PubMed, CINAHL, and SCOPUS. 21 studies were included that assessed EEG and/or MRI patterns in neonates who received therapeutic hypothermia and were followed to determine their outcomes.

Results

Abnormal EEG background and burst suppression severity were significantly associated with poor outcomes. Higher MRI injury scores in the basal ganglia and thalamus were also correlated with poor outcomes. Finally, studies also revealed restricted diffusion and greater lesion size in the subcortical gray matter correlated with poor outcome.

Conclusions

EEG background patterns, MRI scoring, subcortical lesion burden, and MRI diffusivity are sensitive metrics for predicting outcome. Both early EEG and MRI features may serve as high

fidelity biomarkers for secondary energy failure and for counseling families of neonates at high risk for devastating neurologic outcome.

2.2 Introduction

Neonatal Hypoxic Ischemic Encephalopathy (HIE) results from prolonged intrauterine hypoxia and represents a significant global disease burden [3,4]. With an estimated incidence of 1.5 per 1000 births, it accounts for nearly a quarter of infant deaths worldwide[1,2]. In addition to mortality, nearly 60% of HIE cases result in severe disability such as cerebral palsy [3,4]. Birth asphyxia is among the top three global causes of Disability Adjusted Life Years (DALYs), and in 2010 caused 42 million DALYs - twice the estimated disability burden of diabetes [6,7].

HIE pathophysiology can be divided into four stages. In the first stage, primary energy failure, anaerobic metabolism predominates in the brain due to lack of oxygen, and if prolonged, it progresses to excitotoxicity and neuronal death through failure of Na^+/K^+ pumps [8,9]. After primary energy failure, there is a latent phase, a variable-duration period during which therapeutic hypothermia can reduce the magnitude of neuronal loss[9]. During the latent phase, there is a transient return of normal cerebral perfusion and partial recovery of neuronal damage[85]. Afterwards, secondary energy failure then ensues, and is characterized by delayed neuronal cell death, cytotoxic edema, excitotoxicity, and microglial activation, occurring six to 48 hours after injury[8,9]. Decreased cellular metabolism and widespread neuronal cell death during secondary energy failure may manifest on the EEG as burst suppression and/or reduced amplitudes[86]. The majority of neonatal seizures also start during this stage[10]. In a subset of patients, tertiary brain injury follows secondary energy failure, and results in further reduction of neuronal cell counts and astrogliosis [8,9].

HIE outcome prognostication is important for guiding clinical interventions and counseling families. Here, twenty-one articles are reviewed that evaluate prognostic features from Magnetic Resonance Imaging (MRI) within the first few weeks of life and EEG during and/or shortly after therapeutic hypothermia. This article discusses the various features that have been proposed for prognostication with the goal of identifying those found most useful for clinical management of infants with HIE.

2.3 Methods

2.3.1 Search Strategy

Articles were collected from three databases: PubMed, CINAHL, and SCOPUS using the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) principles (Figure 2.1)[87]. This search was formulated using the Population, Intervention, Comparison, and Outcome Framework (PICO) [88]. A reproducible search string is shown in Appendix 1. Titles were screened by authors S.E. and H.D. using the data extraction and screening tool Covidence [89].

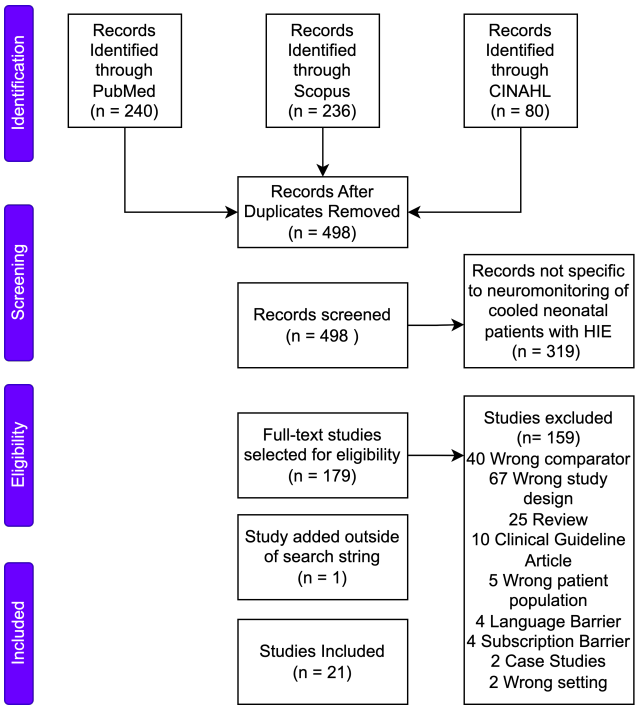

Figure 2.1: PRISMA flow diagram depicting article selection criteria.

2.3.2 Inclusion and Exclusion Criteria

Articles were included if they reported neurodevelopmental outcome at six months of life or later for patients who received therapeutic hypothermia and were monitored with EEG and/or MRI during the neonatal period. Studies were excluded if they did not report established metrics for functional outcome such as Bayley Scales of Infant Development (BSID), Griffiths Mental Development Scales, Gross Motor Function Classification System, the development of cerebral palsy, or post neonatal epilepsy rate [90–92]. Studies with only nonfunctional outcomes, such as length of hospital stay or post-neonatal MRI results, were excluded. One article, Trivedi et al 2017, was identified from another article, Rusli et al 2019, which was captured by the search criteria [93,94].

2.4 Results

From 498 studies initially screened, 20 met inclusion criteria (Figure 2.1). Seventeen studies were retrospective, two were prospective, and one was a multicenter randomized control trial (Tables 2.1-2.3). One prospective study, Trivedi et al 2017, was identified from other included articles, independent of the search criteria [93]. The included articles are presented in three themes: 1. electrographic (EEG) features associated with HIE outcomes, 2. MRI features associated with HIE outcomes, 3. Combination of MRI and electrographic features associated with HIE outcome. Most articles (n=15) assessed outcomes using either the BSID-II or BSID-III scores. The BSID measures cognitive development with composite scores below 70 in the BSID-II and below 80 in the BSID-III signifying risk of poor neurodevelopmental outcome [95].

Table 2.1: Summary of articles assessing EEG features (n=8)

Author, Year	Setting (Country)	Patients included	Outcome Criteria	Features Assessed	Important Modality Features associated with outcome (assessment metric)	Comments
Csekő 2013 [40]	NICU Semmelweis University (Hungary)	70 term Neonates GA (>37 weeks)	Abnormal Outcome = BSID-II (MDI < 70) or death	aEEG background SWC on aEEG	Abnormal aEEG background had PPV of 0.92 of 60 hours for abnormal outcome PPV of no SWC was 0.73 for abnormal outcome	Retrospective Study Independent and blinded reviewers
Dereymaeker 2019 [96]	NICU of UZ Leuven (Belgium)	19 neonates (GA 36-41 weeks)	Abnormal outcome = Death, CP, or BSID-II (<70) at 24 months	Dynamic IBI scored for severity from 1 to 4	Dynamic IBI severity score from 18-24 hours of life (AUC = 0.93)	Retrospective Study
Fitzgerald 2018 [97]	NICU of Children's Hospital of Philadelphia (United States)	93 neonates (mean GA 38.7 weeks)	Abnormal Outcome = language and motor delays at 24 months	Epileptic Seizure Exposure in EEG EEG background	High epileptic seizure exposure predicts abnormal language development (p=0.04) Moderate/severely abnormal EEG background predicts motor delay (p=0.01)	Retrospective Study Also showed seizure exposure could predict abnormal MRIs (p=0.02)

Table 2.1: Summary of articles assessing EEG features (n=8)

Kharoshankaya 2016 [10]	Cork University Maternity Hospital (Ireland)	47 neonates (median GA = 40.7 weeks)	Abnormal Outcome measured via BSID-III, GMDS, and CP at 24-28 months	EEG Seizure Burden (TSB and MSB)	TSB>40 minutes and MSB>13 minutes per (p=0.001 and p=0.003 respectively)	Retrospective Study
Koskela 2021 [98]	University College London Hospitals (United Kingdom)	41 neonates (GA 36.3 to 41.6 weeks)	Abnormal Outcome scored using BSID-III	EEG cortical Burst Power (8-30Hz) recovery at post-natal day 3 & above	High burst power associated with worsened outcomes across channels C3, C4, T4, ($p<0.05$) independently of MRI	Retrospective Study Controlled for MRI injury severity to show independence
Leroy-Terquem 2017 [99]	Necker-Enfants Malades Hospital (France)	40 neonates (GA≥36 weeks)	Abnormal outcome = WHO disability score	EEG Asynchrony within first 7 days of life	EEG asynchrony had 97% sensitivity and 80% specificity for predicting major disability	Retrospective Study Choice of 1.5 seconds as threshold for burst separation could be further justified
Sewell 2018 [41]	Children's National Health Systems (United States)	80 neonates (GA≥35 weeks)	Abnormal outcome scored via BSID-II and BSID-III	aEEG background aEEG latency factors TTDC, TTN, TTC	aEEG BG pattern ($p<0.005$) aEEG latency factors predicted outcomes (sensitivity= 0.944 and specificity=0.852)	Prospective Study Also looked at short term outcome metrics using MRI severity as a metric
Takenouchi 2011 [100]	New York-Presbyterian Hospital (United States)	29 neonates (GA≥36 Weeks)	Outcome based on BSID-II MDI & ambulation without or support	Acquisition of SWC during the first 144 hours of life	Failure to acquire SWC by 120 hours in severe HIE neonates is associated with poor outcomes ($p=0.02$)	Retrospective Study Also used ROC to predict 120 hours is most sensitive for outcome prediction but with a low AUC of 0.53

Table 2.2: Summary of articles assessing MRI features (n = 9); all studies were retrospective.

Author, Year	**Setting (Country)**	**Patients included in Study**	**Outcome Criteria**	**Modalities Assessed**	**Features Assessed**	**Important Modality Features associated with outcome (assessment metric)**	**Comments**
Chang 2020 [101]	Seoul St. Mary's Hospital (Korea)	107 neonates (GA≥35 weeks)	Abnormal Outcome = BSID-III scores < 85 or death	MRI (DWI)	Lesions scored via NICHD scoring system Lesion size (pixels)	MRI lesion size <100 and >500 ($p<0.05$) MRI lesion counts < 2 & between 14-40 ($p<0.05$) NICHD stages 0 to 2A ($p<0.05$)	Retrospective Study

Table 2.2: Summary of articles assessing MRI features (n = 9); all studies were retrospective.

Chintalapati 2021 [102]	St. Louis Children's Hospital (United States)	50 neonates (GA≥35 weeks)	Abnormal outcome = dystonia or spasticity	MRI (DWI)	Striatal and Thalamic ADC	Striatal ADC <1.014 × 10^{-3} mm^2/s (100% specificity and 70% sensitivity) Thalamic ADC <0.973 × 10–3 mm2/s (100% specificity and 80% sensitivity)	Retrospective Study Somewhat restrictive region of interest by only focusing on striatal and thalamic regions
Hayakawa 2018 [103]	Red Cross Kyoto Daiichi Hospital (Japan)	17 neonates (mean GA = 38.3 weeks)	Abnormal Outcome =major disability via GMFCS for CP & standard neurologic exam	MRI (DWI)	DWI pseudonormalization in MRI from week 2 of life	Week 2 pseudonormalization negativity (100% sensitivity, 100% specificity)	Retrospective Study Small sample size
Jung 2015 [104]	Northwestern University: (United States)	73 neonates (GA≥35)	Abnormal Outcome = Post Neonatal Epilepsy within 1 year or death	MRI (DWI)	MRI injury location	9 of 13 neonates with brainstem injuries died Post neonatal epilepsy associated BG/T & brainstem injury (p<0.003))	Retrospective Study 3 late preterm infants Neonates with only subcortical white matter injury had good outcomes
Lakatos 2019 [105]	Semmelweis University (Hungary)	108 (GA≥36 Weeks)	Abnormal Outcome = BSID-II score < 85	MRI (DWI), MRS	MRI Injury Pattern ICH on MRI Lac/NAA ratio	Presence of HIE via both MRI and MRS significantly associated with poor outcomes (p = 0.0246)	Retrospective Study Concomitant ICH had no significant effect on outcome
Mastrangelo 2019 [106]	Sapienza University of Rome (Italy)	29 neonates (GA≥34 weeks)	Abnormal Outcome = GMDS-III global quotient <85	MRI (DWI)	MRI scored via Bednarek severity scores	PPV of MRI global score at 12 & 24 months = 36.36% and 50% respectively, NPV of MRI at 12 months = 93.75%	Retrospective Study Included single neonate with GA below 34 weeks.
Trivedi 2017 [93]	St. Louis Children's Hospital (United States)	57 neonates (GA ≥ 35 Weeks)	Bayley-III score at 18–24 month	MRI (T1W, T2W, DWI)	Injury severity score across subcortical structures, white matter, cortex, cerebellum, & brainstem	Increased MRI injury grade was significantly associated with poor cognitive and motor outcomes (p<0.001, p<0.012)	Prospective Study

Table 2.2: Summary of articles assessing MRI features (n = 9); all studies were retrospective.

Rusli 2019 [94]	NICU of Universiti Kebangsaan Malaysia Medical Center (India)	19 neonates (GA ≥36 weeks)	Abnormal Outcome = death or CP by 2 years based on clinician notes	MRI (T1W, T2W, DWI)	MRI scored using Trivedi et al 2017 scoring system	Trivedi scoring system was not significantly associated with poor outcome (p=0.350)	Retrospective Study Used different outcome metric to Trivedi et al (Trivedi used BSID-III)
Takenouchi 2010 [107]	NICU at New York-Presbyterian Hospital (United States)	34 neonates (GA ≥36 Weeks)	Abnormal outcome BSID-III MDI score at 18 months	MRI (DWI)	ADC of splenium, corpus callosum	Restricted diffusion in splenium is significantly associated with poor outcomes (p=0.002)	Retrospective Study

Table 2.3: Summary of articles assessing both EEG and MRI features (n=4)

Author, Year	**Setting (Country)**	**Patients included in Study**	**Outcome Criteria**	**Modalities Assessed**	**Features Assessed**	**Important Modality Features associated with outcome (assessment metric)**	**Comments**
Basti, 2020 [108]	NICU San Salvatore Hospital in L'Aquila (Italy)	30 Term Neonates	Abnormal Outcome = death, spastic quadriplegia & BSID-III	aEEG MRI (T1, T2, DWI)	Seizure Burden aEEG background (CNV, DNV, BS, LV, FT) MRI Injury Pattern (PLIC, multifocal, BG/T, WML)	Significant features (p<0.05) = High seizure burden, abnormal aEEG over 48 hours, & abnormal MRI pattern	Prospective Study MRI taken within 4 weeks of life (median = 17 days) Used aEEG alone to identify seizures
Lin 2021 [109]	Seoul St. Mary's Hospital (South Korea)	97 Neonates (GA≥35 Weeks)	Abnormal Outcome = BSID-II score < 85	MRI, aEEG	Clinical Seizures (evidence by use of AEDs) aEEG background NICHD MRI Pattern	Abnormal aEEG associated with poor outcome (p<0.05) BG/T and PLIC lesion groups in 2A and 2b associated with abnormal outcomes (p<0.001)	Retrospective Study MRI taken (≤10 days of life) Also showed seizure severity were associated with BG/T lesions
Peeples 2021 [110]	Children's Hospital and Medical Center (United States)	486 neonates (mean GA = 38.8 weeks)	Abnormal Outcome based on BSID-III & GMFCS score	EEG, aEEG, MRI (DWI)	MRI HIE severity (either cortical or deep gray injury), EEG Background aEEG background	Combo of either severe grade HIE or abnormal aEEG/cEEG at 24 hours is associated with poor outcomes (p<0.001)	Retrospective Study

Table 2.3: Summary of articles assessing both EEG and MRI features (n=4)

Weeke 2016 [111]	Wilhelmina Children's Hospital (Netherlands)	26 neonates (GA mean 40.4 weeks)	Abnormal Outcome via BSID-III, CP, epilepsy, hearing or vision loss, death	EEG, MRI (T1, T2, DWI)	EEG TSB and EEG MRI Pattern via Barkovich Score	EEG Background at 36 HOL, TSB, and MRI all associated with outcomes ($p=0.009$, $p = 0.036$, $p<0.001$ respectively)	Multicenter Randomized Trial No multivariate analysis

Abbreviations: BG/T (Basal Ganglia/Thalamus), aEEG (amplitude-integrated EEG), TH (Therapeutic hypothermia therapy), CNV (Continuous Normal Voltage), Discontinuous Normal Voltage (DNV), BS (Burst Suppression), LV (Low Voltage), Flat Trace (FT), ADC (Apparent Diffusion Coefficient), Mental Developmental Index (MDI), IBI (Interburst Interval), Gross Motor Function Classification System (GMFCS), CP (Cerebral Palsy), WHO (World Health Organization), HOL (Hour of Life), TSB (Total Seizure Burden), MSB (Maximum Seizure Burden), MRS (Magnetic Resonance Spectroscopy), AEDs (Anti-Epileptic Drugs), GFMDS (Griffiths Mental Development Scales), PICU (Pediatric Intensive Cre Unit), T1W (T1-Weighted), T2W (T2-Weighted), TTDC (time to discontinuous), TTN (time to normalization), TTC (time to cycling), RCT (Randomized Control Trial), WML (White Matter Lesion), NICHD (National Institute of Child Health and Human Development)

2.4.1 EEG Features Associated with HIE Outcomes

EEG Qualitative Assessments: Two studies, Weeke et al 2016 and Fitzgerald et al 2018, evaluated EEG background qualitatively based on visual assessment during hypothermia [97,111]. They graded EEG background using ordinal scales. The lowest grade was given to EEGs with mild discontinuity (interburst interval duration ≤ 10 seconds) for gestational age while the highest grade was given to very abnormal EEG backgrounds with severe discontinuities (interburst interval duration ≥ 30 seconds) and attenuated voltage (<25μV). Both articles found that severely abnormal EEG background within the first 24 [111] and 36 [97] hours of life as associated with abnormal outcomes.

Burst and interburst patterns: Koskela et al 2021 assessed EEG bursts quantitatively [98]. They computed the burst power between 8-30 Hz of EEGs in the hours directly following therapeutic hypothermia (≥ postnatal day 3). They found that elevated bilateral central, occipital,

and right temporal burst power (across channels C3, C4, O1, O2, and T4) was inversely correlated with BSID-III language and motor scores with correlation coefficients (R values) ranging from -0.49 to -0.31. High burst power directly after hypothermia was therefore associated with worse BISID-III scores [98].

Dereymaker et al 2019 used an automated assessment of interburst intervals (IBIs) to grade EEGs during therapeutic hypothermia in 19 neonates [96]. These bursts, by definition, are abnormal and not the pattern seen in tracé alternant [112]. They used a metric called dynamic IBIs (dIBIs), which measures both IBI duration and amplitude, to score the severity of EEG abnormality from one to five. EEGs with high-amplitude and short-duration IBIs were assigned a low severity score while EEGs with low-amplitude and long-duration IBIs were assigned a high severity score. They found that at 19-24 hours of life, median IBIs duration < 10 seconds and IBI amplitudes ≥15μV were associated with favorable outcome ($p<0.001$). EEGs with moderately high IBI amplitudes and brief IBI durations were less pathologic and predicted better outcomes.

Amplitude-integrated EEG (aEEG): Five articles evaluated aEEG background during at least one hour of therapeutic hypothermia as a tool for outcome prognostication [40,41,108–110]. All five articles used the BSID-II or BSID-III at 18-24 months as part of their criteria for outcome prognostication. Four studies found that any of the most abnormal aEEG patterns, including burst-suppression (BS), low voltage (LV), or a flat trace (FT), was associated with poor outcome ($p<0.05$) [40,108–110]. In addition, Csekő et al reported that an abnormal aEEG at 60 hours of life had a positive predictive value of 0.92 for poor outcome [40].

Interhemispheric Dynamics: Leroy-Terquem et al 2017 evaluated EEG asynchrony as a predictor of outcomes using the World Health Organization (WHO) disability score in 40 term neonates [99]. Asynchrony was defined as a discontinuous background with periods of abnormal

burst activity with burst onsets separated by at least 1.5 seconds over both hemispheres [99]. EEG discontinuities identified were noted to be pathologic and independent of the normal tracé alternant pattern observed in the quiet sleep of term neonates [112]. Asynchrony was assessed within the first 48 hours of life and again from another day within the first week of life. Asynchrony during the first seven days of life identified neonates who would develop major disabilities with a sensitivity of 80% and a specificity of 97%.

Sleep-Wake Cycles: Takenouchi et al 2011 evaluated EEGs following therapeutic hypothermia from 72 to 144 hours of life for evidence of sleep-wake cycles (SWCs) in 29 neonates [100]. Neurocognitive outcomes were based on the BSID-II Mental Development Index (MDI). Takenouchi et al classified a neonate as having SWCs if their EEG contained at least two state changes across a six-hour EEG epoch. These state changes indicated a transition into quiet sleep or into wakefulness. Failure to acquire SWCs within the first 120 hours of life had a sensitivity of 90% and a specificity of 60% for poor outcome ($p=0.02$).

Seizures: Five studies assessed seizures as predictors of functional outcome [10,97,108,109,111]. Seizure burden was quantified in four of these articles [10,108,109,111]. Basti et al 2020 found that increasing seizure burden, as assessed on aEEG across 30 neonates, was significantly associated with poor outcome ($p=0.0004$) [108]. Fitzgerald et al 2018 used a different method for calculating seizure burden, epileptic seizure exposure, which is defined as the total number of seizures in the EEG during both hypothermia therapy and rewarming [97]. The total duration of EEG varied depending on the clinical needs of the child, so seizure rates were not reported. They found that high epileptic seizure exposure (≥4 seizures during cooling and rewarming) was associated with motor delay ($p<0.01$) and having ≥3 seizures during cooling and rewarming was associated with language delay ($p=0.01$) using the BSID-III. In conclusion, these studies are consistent in showing

that a high seizure burden (>30 minutes per hour or ≥3 seizures total) during therapeutic hypothermia is associated with worse outcomes.

2.4.2 MRI Injury Patterns Associated with HIE Outcomes

Qualitative MRI injury scoring: Nine articles evaluated MRI features alone for predicting outcomes [93,94,101–107] (Table 2.2). Five of the studies assessed MRI using published injury severity scoring systems [93,94,101,106,109]. Two studies applied the NICHD scoring system, which ranks lesion severity across the basal ganglia, thalamus, internal capsule, watershed, and cerebral hemispheres [113]. Injury scores span six categories: 0, 1A, 1B, 2A, 2B, and 3; zero signifying a normal MRI and three signifying hemispheric devastation. Chang et al 2020 found that the NICHD scoring system had an Area Under Curve (AUC) of 0.756 for predicting poor outcomes at 18-24 months. Prognostication using the scoring system was marginally better than simply using lesion size or lesion count, which had AUCs of 0.718 and 0.705 respectively [101]. Lin et al found the specific pattern of injury was also predictive of poor outcomes: MRI lesions scored as 2A and 2B, which involve basal ganglia/thalamic and posterior limb of internal capsule (PLIC) locations, were associated with poor outcomes ($p<0.001$) while those without were not [109].

In contrast, Mastrangelo et al used the Bednarek Severity Scoring System – to characterize the MRI [114]. Like the NICHD score, the Bednarek score measures HIE injury severity, but unlike the NICHD, it sums the individual injury scores across the basal ganglia, brainstem, white matter, cortex, and cerebellum into a single global injury severity score [114]. Mastrangelo et al suggest that a global MRI injury score of 55 can be used as a cutoff to separate good from poor outcome groups. Bednarek scores below 55 (range 48-55) were associated with better neuromotor outcomes scores

at 12 months while Bednarek scores above 55 (range 56-186) were associated with worse neuromotor scores (p=0.02) [106].

One study found that a single MRI-based scoring system might not capture all poor neurodevelopmental outcomes. Rusli et al 2019 used the Trivedi MRI scoring system and found no significant association between MRI injury score and outcomes of cerebral palsy or death by two years of age (p=0.350) [94]. Trivedi et al (2017) developed the scoring system that sums MRI injury severity across the brainstem, cortex, white matter, and five subcortical locations: the globus pallidus, putamen, caudate nucleus, thalamus, and PLIC to generate a composite MRI injury score [93]. Trivedi et al assessed their metric for association with outcomes using the BSID-III across 57 neonates with HIE. To evaluate their scoring metric, they dichotomized outcomes by labeling neonates with BSID-III scores < 85 as poor outcomes and BSID-III scores > 85 as good outcomes. Trivedi et al found that their MRI injury score was significantly associated with poor cognitive ($p<0.001$) and motor outcomes ($p<0.012$). In contrast, Rusli et al only evaluated the scoring system using a cohort of 19 neonates and their functional outcome metric was the development of cerebral palsy by two years of age. These discrepancies may explain the poor performance of the Trivedi scoring system in the study by Rusli et al.

Two additional studies evaluated MRI injury patterns independently of a specific MRI scoring system to prognosticate HIE outcomes [104,105]. Jung et al 2015 assessed the MRI for injury in specific locations, including the cortex, subcortical white matter, basal ganglia, and thalamus. They found that the development of post-neonatal epilepsy was associated with subcortical injuries involving the basal ganglia, thalamus, and brainstem. Lakatos 2019 considered both MRI and Magnetic Resonance Spectroscopy (MRS) findings as potential predictors of outcome [105]. For MRI, they considered three patterns of injury: basal ganglia-thalamus, watershed pattern, and total

brain injury as well as the presence of concomitant intracerebral hemorrhage (ICH). For MRS, they considered a high lactate/N-acetylaspartate (Lac/NAA) ratio on MRS as indicative of injury. Outcome was assessed using BSID-II score at 18 to 26 months. On multivariate regression, they found that infants with these MRI or MRS patterns of HIE had higher odds of poor outcome (Odds Ratio = 6.23; CI95% = [1.26; 30.69], p = 0.025) than those who were HIE negative on both MRS and MRI. Interestingly, a concomitant intracerebral hemorrhage was not significantly associated with worse outcomes.

Quantitative scoring for lesion burden: Quantitative MRI measures were also assessed as predictors of outcome. In a study of 107 term neonates using diffusion-weighted MR images, Chang et al 2020 determined the lesion size and the number of lesions in the NICHD injury score locations [101]. They found that DWI-MRI lesion sizes >500 pixels and lesion count between 14 and 40 were both independently associated with poor outcome. Chintalapati et al 2021 and Takenouchi et al 2010 assessed diffusion restriction using the apparent diffusion coefficient (ADC) [102,107]. Chintalapati et al assessed the ADC in the striatum and thalamus and found that an average striatal ADC less than 1.014×10^{-3} mm^2/s across free-drawn regions of interest in the left and right striatum had 100% specificity and 70% sensitivity for the development of dystonic cerebral palsy. In addition, an average thalamic ADC of less than 0.973×10^{-3} mm^2/s across free-drawn regions of interests in both the left and right thalamus had 100% specificity and 80% sensitivity for the development of dystonic cerebral palsy. Finally, using a cohort of 34 neonates, Takenouchi et al 2010 compared infants who had restricted diffusion changes in the splenium of the corpus callosum to those without changes. They found that those with restricted diffusion in the splenium had higher rates of poor neurocognitive outcomes (p=0.002). Restricted diffusion in the splenium

had a positive predictive value of 90% for predicting poor motor outcomes and a negative predictive value of 71% [107].

2.5 Discussion

This systematic review was conducted to identify features for predicting neurodevelopmental outcome in term and near-term neonates with HIE who received therapeutic hypothermia. This review focuses on articles that applied EEG and MRI since they are two of the most frequently implemented modalities for assessing neonatal brain structure and function in clinical practice. Functional outcomes were assessed in the reviewed studies primarily using standardized metrics such as the BSID at 18 to 24 months or by evaluating for the presence of conditions such as cerebral palsy or post-neonatal epilepsy. Overall, multiple features and combinations of features of EEG and MRI were found to be associated with or predictive of neurodevelopmental outcomes.

Four EEG features observed during therapeutic hypothermia were associated with poor functional outcome: an abnormal EEG background pattern, interhemispheric asynchrony, lack of sleep-wake cycle recovery, and increased seizure burden. In the period after therapeutic hypothermia, burst and interburst characteristics were useful predictors. Because these EEG features manifest during and shortly after the therapeutic hypothermia window, they may be predictive of outcome and even serve as biomarkers of secondary energy failure.

MRI patterns were a bit more complex to assess, in part because the most widely used MRI scoring systems – the NICHD, Barkovich, Rutherford systems – were designed prior to therapeutic hypothermia becoming the standard of care [115–117]. Thus, these scoring systems were created by observing injury patterns in heterogeneous cohorts that contained both neonates that underwent therapeutic hypothermia and others who did not. Cooling may cause discrepancies particularly in

scoring systems that evaluate MRI diffusivity, due to pseudonormalization [114,118]. Pseudonormalization occurs when the MRI diffusivity returns to baseline after dipping below baseline due to tissue injury [114]. Thus, injuries are no longer apparent on the MRI. Non-cooled neonates undergo MRI pseudonormalization around 6 to 8 days, whereas cooling pushes pseudonormalization out to 10 to 11 days [114,118]. Additionally, MRI injury severity scores in the basal ganglia and watershed region were found to be significantly lower in cooled neonates than noncooled neonates [119]. To constrain potential heterogeneity in results, the articles selected in this review only included HIE cohorts that received therapeutic hypothermia. These articles showed that injury within deep subcortical structures, particularly the basal ganglia and thalamus, were consistently associated with poor outcomes. This pattern may exist because the basal ganglia and thalamus are among the most metabolically active brain regions in term neonates [120]. This finding is corroborated by regional hyperperfusion in the basal ganglia and thalamus, which was captured using MRI-arterial spin labeling [120]. Thus, these structures seem particularly susceptible to changes in brain perfusion, with the extent of their damage correlating with the severity of hypoxic injury and ultimately the severity of long-term neurological outcomes.

There is also the question of how EEG compares to MRI for outcome prediction, since EEG may be more accessible than MRI in certain conditions and settings, for example, in infants not stable enough to be transported for an MRI or in hospital units without ready access to an MRI scanner [121]. Studies have shown that severe EEG background abnormalities, attenuated EEG power, and electrographic seizure burden correlate with MRI injury severity [111,122,123]. Despite these correlations, it is unclear if EEG patterns can be used to predict specific MRI injury patterns. Clarifying these questions will allow clinicians to ascertain the extent to which EEG can serve as a biomarker for both MRI injury location and neurodevelopmental outcome.

The ultimate goal of identifying the most accurate and useful prognostic features is to provide clinicians and families with sound data on which to base care decisions, and to aid clinicians in counseling families about their newborn's probable neurodevelopmental outcome [124]. Families must make critical decisions in the acute neonatal period that influence the continuation of life-sustaining therapies and overall goals of care for the neonate. These decisions often include whether to place tracheostomy and gastrostomy tubes, whether to escalate to invasive ventilation, and whether to initiate extracorporeal membrane oxygenation (ECMO). These decisions depend, in part, on the etiology, severity, and prognosticated outcome of hypoxic ischemic injury. Higher fidelity prognostication provided by sensitive and specific biomarkers may allow clinicians to feel more confident in the counseling they provide to families and allow families to feel more confident in their clinical decisions in the acute setting.

This review also identified limitations that may serve as opportunities to better characterize prognosticators for neonatal HIE in future studies. First, many of the current studies have small sample sizes; for example, seven of the twenty articles reviewed had fewer than 30 subjects (Tables 2.1-2.3). This limits the number of variables that can be assessed in a particular case series and raises the risk that a few outliers can bias the results. Larger studies, or even reanalysis of several previous datasets with a predefined protocol, may help confirm which markers are the best, independent predictors of outcome.

Additionally, while MRI has been proven important for prognostication, its use must be standardized in clinical practice to allow consistent interpretation. For example, MRI pseudonormalization after week one of life causes injuries to appear less severe [103]. One study, Basti et al 2020, assessed MRIs at various time windows (range 5-30 days). MRIs performed after one week of life may have been susceptible to the effect of diffusion pseudonormalization, which

would underestimate injury severity. This approach may have caused MRI results to vary [108]. MRI-based prognostic features should therefore be defined for a specific imaging time window for clinical practice. In addition, most MRI scoring systems focus on scoring the extent of cortical injury as opposed to assessing particular areas of cortex that may have differential effect on prognosis [93,111,115]. There is a need for validated scoring systems that allow prognostication of specific types of disability based on injury to particular cortical areas.

One more subtle limitation of the studies reviewed is the focus on long-term outcome data at 18 to 24 months. While this follow-up window likely gives an accurate description of the ultimate disability a child will face, it does not capture the full demographic spectrum of those at risk. Neonates with less access to clinical care may be more likely to have HIE and are more susceptible to follow-up loss [22]. Swearingen et al 2020 showed follow-up loss of 62% across a cohort of 237 neonatal patients with rates varying significantly across different demographic strata such as median income and race.

2.6 Conclusions

The objective of this review is to amalgamate clinical features taken from both EEG and MRI that are significantly associated with long-term neurodevelopmental outcomes in neonates undergoing therapeutic hypothermia for HIE. 72-hour therapeutic hypothermia coupled with continuous EEG monitoring and followed by MRI within seven days of hypoxic injury has become the standard of care. High fidelity, prognostic EEG and MRI features may help guide clinical management of time-sensitive conditions and escalation of care. Ultimately, the various features identified will need to be combined into a multimodal model for sensitive and specific outcome prediction that can be easily used by clinicians to benefit patients and family members at the NICU bedside.

CHAPTER 3

MRI AND EEG FEATURES FOR OUTCOME PROGNOSTICATION IN NEONATES WITH HIE

3.1 Abstract

Objective

To determine early EEG and MRI features that are significantly associated with graded, functional outcome in infants with hypoxic ischemic encephalopathy (HIE).

Methods

EEG characteristics, MRI injury patterns, and outcomes were retrospectively assessed in 23 full-term neonates (mean gestational age 38 weeks) with HIE. EEG features were extracted from recordings made at between 24 and 26 hours of life during therapeutic hypothermia. MRIs within the first 1.5 weeks of life were scored for injury. Functional outcomes were graded from 0 (best) to 5 (worst) and were assessed at between three and six months of life. The relationships between EEG or MRI features and outcomes were evaluated using ordinal logistic regression analysis.

Results

Of the features assessed, the absence of EEG state changes at 24 hours of life and insular injury on MRI were the most strongly associated with poor outcomes at 3-6 months post-partum. When combined in a multivariate model, regression analysis suggested both features were independently predictive of outcomes.

Conclusion

A lack of EEG state changes during therapeutic hypothermia and insula injury within the first 1.5 weeks of life are predictive of poor outcomes.

3.2 Introduction

Hypoxic Ischemic encephalopathy (HIE) is a major cause of global morbidity and mortality in neonates. With an estimated incidence of 1.5 per 1000 births, HIE accounts for nearly a quarter of infant deaths worldwide [1,2] and is the cause of approximately 60% of neonatal seizure cases [125]. HIE is also associated with long-term morbidity such as cerebral palsy [4]. Since the severity of long-term outcomes can range from mild to extreme, accurate algorithms are needed for early prognostication of outcome severity.

Both magnetic resonance imaging (MRI) and continuous electroencephalography (EEG) monitoring are routinely used to evaluate patients with HIE. Many studies have considered whether individual findings on the EEG or MRI are associated with outcome in neonates with HIE [97,101]. In studies examining EEG, the background pattern appears to carry sensitive prognostic information particularly at 24 hours of life [97,126]. In infants that receive therapeutic hypothermia, this prognostic period may extend to 48 hours of life [100]. In addition, injury on MRI may be seen clearly within the first 6 to 8 days in normothermic infants and within the first ten days in hypothermia-treated infants [118]. After this time period, injury on MRI may appear less severe due to diffusion pseudonormalization [114]. Despite this process, lesions that remain apparent during the pseudonormalization period may be particularly severe and thus retain significant prognostic accuracy [103].

While some features predictive of poor outcome have been identified, this study addresses two understudied issues. First, previous studies consider predictors of long-term outcomes, defined at 2-3 years of age. However, the population at highest risk for HIE has a high follow-up loss rate, so long-term outcome studies might miss the highest-risk population. Second, few studies have looked to combine both EEG and MRI into multivariate algorithms for outcome prognostication

in neonatal HIE [97,127], and none have identified combinations that correlate with outcomes within six months. In this study, imaging and electrographic features are used to construct a prognostic model for early outcomes in neonates who underwent therapeutic hypothermia for HIE.

3.3 Methods

3.3.1 Patient Inclusion criteria

A retrospective chart review was conducted to identify neonatal patients with HIE who received therapeutic hypothermia between March 2010 and March 2021. The study was approved by the University of Chicago Institutional Review Board. Patients were included if they underwent 72-hours of therapeutic hypothermia, had MRI within 11 days of birth, had one hour of continuous EEG data between hours 24-26 and 48-50 of life, and had follow-up assessments between three and six months of life (Figure 3.1). Sarnat stage 1 HIE patients that were included met primary physiologic criteria for hypothermia therapy based on the presence of either prolonged resuscitation, a sentinel event like cord prolapse before delivery, 10-minute Apgar score ≤ 5, severe acidosis (pH ≤ 7.1), or abnormal base excess (≤ - 10 mEq/L) [128]. Neonates who underwent the hypothermia protocol but who had a normal MRI were excluded to eliminate non-HIE causes of encephalopathy like chromosomal or metabolic abnormalities.

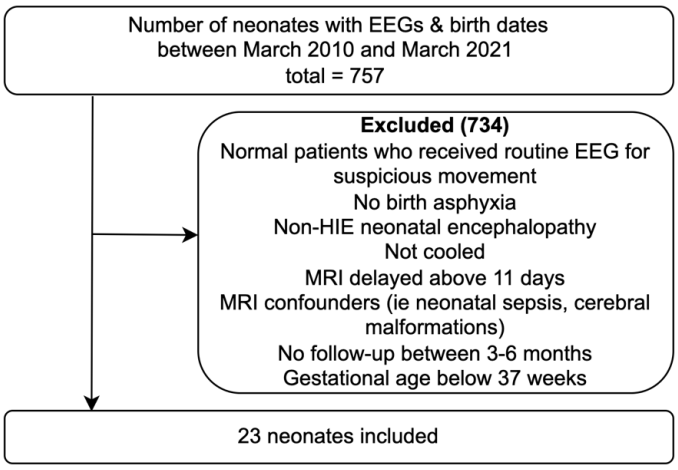

Figure 3.1. Patient inclusion and exclusion criteria flow chart.

3.3.2 EEG Assessment

Clinical EEGs were collected in the neonatal intensive care unit of Comer Children's Hospital using Natus or Nicolet One EEG systems with electrodes Fp1, Fp2, T3, T4, C3, C4, O1, O2, and Cz and were viewed using the neonatal bipolar and/or common average reference montages. One-hour epochs recorded between 24 and 26 hours of life and another between 48 and 50 hours of life were analyzed. Each EEG was read at 15mm/sec at a 7μV/mm sensitivity. The low and high frequency filters were set to 1 and 30 Hz respectively. EEG features were identified by a pediatric neurologist specializing in epilepsy (J.H.) and blinded to patient outcomes. EEG findings were corroborated using the EEG report in the patient's chart interpreted at the time of care. Each feature was scored as present or absent (0 or 1 respectively).

3.3.3 MRI Injury Scoring

Structural MRIs were acquired using a General Electric 3T scanner and included diffusion-weighted images (DWI), T2-FLAIR sequences, and T1 sequences. MRIs were scored using an

adaptation of the Rutherford Scoring System using DWI sequences and were confirmed using apparent diffusion coefficient (ADC) sequences (Figure 3.2) [129]. Scoring was done by a pediatric neurologist (H.D.) blinded to patient outcomes. Injury locations were corroborated by injury locations listed in each patient's clinical MRI report; clinical MRI interpretation was performed by neuroradiologists at the University of Chicago Medical Center. All MRIs were acquired within the first 11 days of life (mean = 5.64 days, median = 5 days). This age range is in accordance with findings that cooling therapy prolongs pseudonormalization in hypothermia-treated infants to 11 – 12 days compared to 6-8 days in normothermic infants [114].

For this study, the Rutherford MRI scoring system [129] was adapted to ask whether injury in individual brain areas correlated with outcomes. The original Rutherford scoring system sums injury severity scores across four brain areas: the basal ganglia/thalamus, posterior limb of internal capsule (PLIC), periventricular white matter, and the cortex. The current study considers injury to each Rutherford area as a separate variable as follows: injury to the basal ganglia/thalamus and periventricular white matter are scored from zero to three, with zero indicating no injury, one indicating mild focal injury, two indicating multifocal injury, and three indicating widespread severe injury. The PLIC is scored from zero to two with zero indicating no injury, one indicating equivocal or reduced signal intensity, and two indicating loss of signal intensity. Cortical injury is scored as a binary variable (injury or no injury) in three locations: the Central Sulcus, Interhemispheric Fissure, and Insula. Thus, six features were extracted from the MRI; each was treated as an ordinal variable.

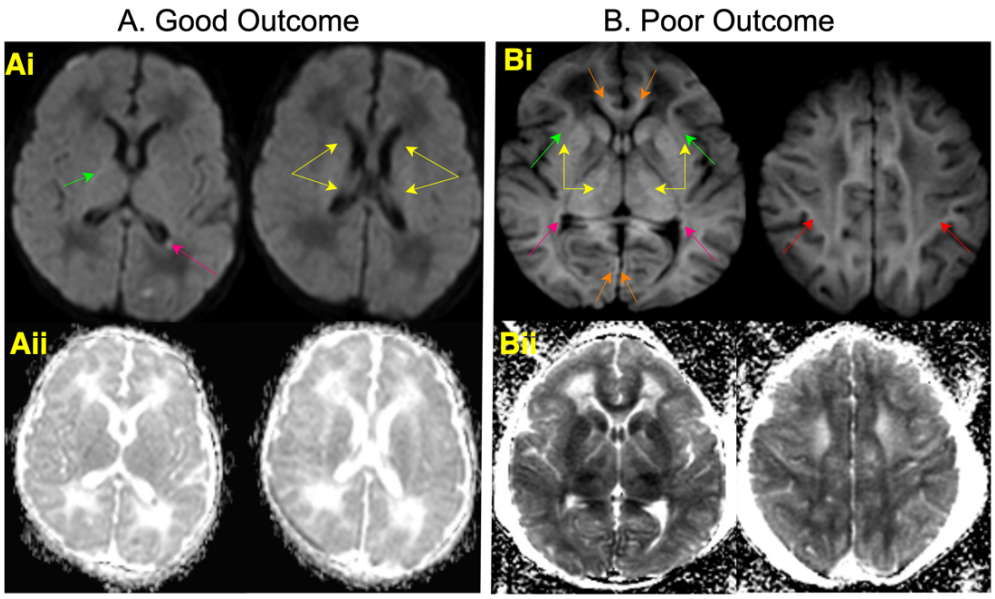

Figure 3.2. Example MRI from patients with a good outcome (A) or poor outcome (B). A. DWI-MRI of good outcome patient (same patient as Figure 3.2A). Ai. MRI injury locations are labeled in the posterior limb of internal capsule (green arrow), periventricular white matter (pink arrow), and basal ganglia and thalamus (yellow arrows). Aii. Injury pattern confirmed by ADC. Lesions appear dark in ADC and bright in DWI. B. DWI-MRI from a patient with poor outcome (same patient as Figure 3.2B). Bi. MRI injury locations in the interhemispheric fissure (orange arrows), insula (green arrows), basal ganglia and thalamus (yellow arrows), periventricular white matter (pink arrows), and central sulcus (red arrows). Bii. Injury pattern confirmed by ADC.

3.3.4 Outcome Assessment

Functional outcomes were assessed at three to six months of age. The presence or absence of five potentially aberrant neurodevelopmental markers was determined by chart review. These markers are the presence of abnormal tone (axial or appendicular hyper- or hypotonia) on neurological exam, spasticity (specifically hypertonicity increased by movement), abnormal visual tracking behavior, failure of hearing screen, and use of antiepileptic treatments such as phenobarbital, levetiracetam, or phenytoin during the follow-up period. The outcome score ranged from zero to five based on the number of markers a patient displayed.

3.3.5 Statistical Methods

Statistical analysis was performed using Stata 17 (StataCorp LLC, College Station, TX). Each feature was initially assessed using univariate ordered logistic regression. The two most significant EEG and/or MRI features were then selected for multivariate assessment. McFadden's pseudo-R^2 was used to assess the goodness-of-fit for logistic regression models [130], with values between 0.2-0.4 representing an excellent fit [130].

3.4 Results

3.4.1 Patient Demographics and Clinical Features

23 full-term neonates were included in this study. 52% (n=12) were male (Table 3.1). The Sarnat score, 10-minute Apgar score, cord gases (arterial, pCO_2 and base deficit) were also collected. None of these measurements varied significantly with functional outcomes (Table 3.1).

Clinical Features (All Numeric)	**Odds Ratio (95% Confidence Interval)**	**McFadden's R^2**	**P-Value**	**Comments**
Gestational Age (in weeks)	1.05 (0.46, 2.39)	0.0002	0.909	min = 37, max = 41.14, mean = 39.46, median = 39.86
Sarnat Score	1.22 (0.30, 5.04)	0.0012	0.780	n=7 (stage 1), n=13 (stage 2), n=3 (stage 3)
Apgar	0.78 (0.54, 1.12)	0.028	0.184	10-minute Apgar
Cord Gas (Arterial)	10.43 (0.069, 1565)	0.013	0.359	Taken within one hour after resuscitation
Cord Gas (pCO2)	0.97 (0.95, 1.0)	0.062	0.059	Taken within one hour after resuscitation
Cord Gas (Base Deficit)	1.09 (0.95, 1.24)	0.023	0.234	Taken within one hour after resuscitation

Table 3.1. Clinical features for each neonate in the study (n=23).

3.4.2 Performance of EEG Features

Seven EEG features were assessed per patient at both 24 and 48 hours (Table 3.2): Low EEG voltage (<25μV), flat trace (<2μV), presence of state changes, hemispheric asynchrony, interburst intervals >10s, presence of seizures, and hemispheric asymmetry. Two features were significantly associated with outcomes in univariate logistic regression analysis (Figure 3.3, Table 3.2): presence of EEG state change at 24 hours of life was associated with good outcomes (p=0.013) and interburst intervals longer than 10s at 24 hours of life was associated with poor outcomes (p=0.018). Figure 3.3 shows an example EEG associated with a good outcome (Figure 3.3A) and an example EEG associated with poor outcome (Figure 3.3B). Figure 3.4Ai and 3.4Aii show the distributions of outcomes for the two significant EEG predictors.

Presence of EEG state change was ascertained by either evidence of EEG reactivity, the appearance of normal graphoelements such as enconches frontales, the appearance of transitions between wakefulness, active, or quiet sleep based on EEG amplitude and frequency changes such as those seen in transitions between activité moyenne and tracé alternant [131]. Infants who had prolonged interburst intervals exhibited intervals of attenuated EEG (<25μV) that lasted longer than 10-seconds following bursts, which exceeds the maximum normal interburst interval of six-seconds for term neonates [131].

Univariate logistic regression analysis was also done to evaluate a possible confounder which is whether these predictors were associated with phenobarbital administration during therapeutic hypothermia; only interburst interval durations ≥10s at 24 hours of life was significantly correlated with phenobarbital administration during the first three days of life (p=0.031). Phenobarbital administration alone during hypothermia was not associated with outcomes (p=0.339).

EEG Predictors (All Binary)	**Odds Ratio (95% Confidence Interval)**	**McFadden's R^2**	**P-Value**	**Comment**
State Change Absence at 24 Hours	**8.97 (1.58, 51.1)**	**0.109**	**0.013**	**Ie transitions into or out of wakefulness**
State Change Absence at 48 hours	2.50 (0.52, 12.0)	0.0209	0.251	See previous row
Interburst Intervals ≥ 10s at 24 Hours	**7.80 (1.42, 42.76)**	**0.0973**	**0.018**	
Interburst Intervals ≥ 10s at 48 Hours	5.00 (0.97, 25.79)	0.061	0.055	
Low Voltage at 24 Hours	5.00 (0.96, 25.79)	0.0613	0.055	
Low Voltage at 48 Hours	2.31 (0.51,10.5)	0.0185	0.280	
Asynchrony at 24 Hours	3.97 (0.78, 20.18)	0.0448	0.097	
Asynchrony at 48 Hours	3.47 (0.66, 18.14)	0.0349	0.141	
Presence of Seizure(s) at 24 Hours	4.75 (0.70, 32.40)	0.0408	0.112	
Presence of Seizure(s) at 48 Hours	2.20 (0.10, 50.31)	0.0038	0.621	
Asymmetry at 24 Hours	2.08 (0.40, 10.81)	0.0118	0.385	
Asymmetry at 48 Hours	2.35 (0.40, 13.9)	0.0138	0.347	

Table 3.2. Univariate assessment of EEG predictors of outcome across all included patients (n=23). All features are binary variables indicating presence or absence. Bolded features are significant ($p<0.05$).

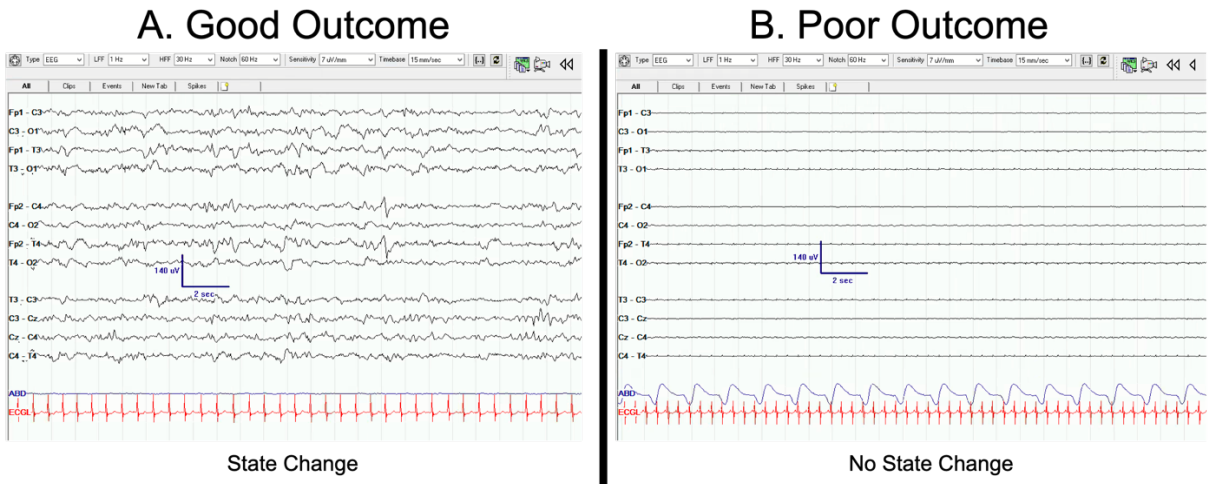

Figure 3.3. Example EEGs. A. EEG of patient with good outcome. EEG shows continuous normal voltage >25μV and state change via presence of normal graphoelements such as encoches frontales. B. EEG of patient with poor outcome shows flat EEG with severely attenuated voltage and no state changes.

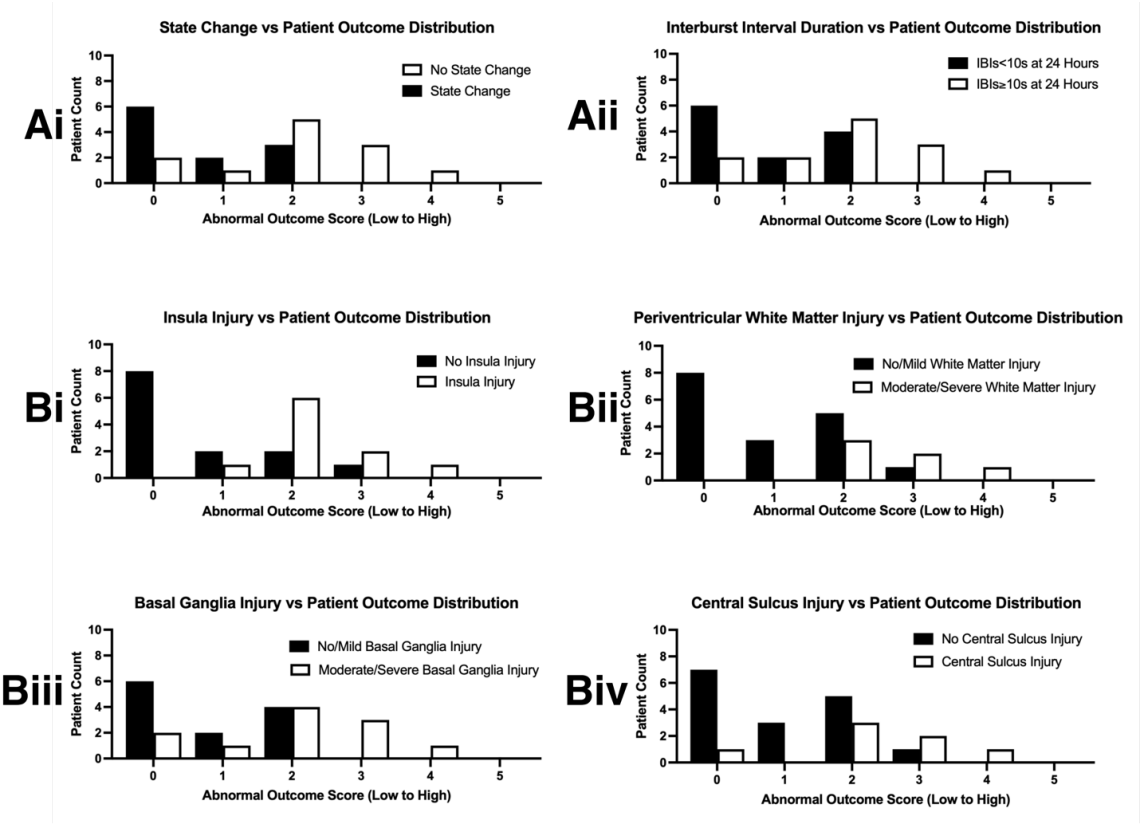

Figure 3.4 Patient outcome score distribution with each of the six significant univariate predictors. A. Shows the two EEG predictors. B. Shows the four MRI predictors.

3.4.3 Performance of MRI Features

The brain injury pattern was characterized using diffusion-weighted images on MRI scans and confirmed using ADC images. The presence and extent of injury was classified using the Rutherford scoring system, which included scores for deep gray matter locations (basal ganglia/thalamus), the PLIC, periventricular white matter and three neocortical regions – the central sulcus, the interhemispheric fissure, and the insula [129]. Based on univariate logistic regression analysis, injury in any of four brain areas was significantly associated with poor outcomes. These areas are the insula (p=0.004), periventricular white matter (p=0.015), basal ganglia (0.017), and central sulcus (p=0.022) (Table 3.3). The distributions of outcomes are shown as a function of injury for each of these four MRI locations in 3.4B.

MRI Predictors (All Ordinal)	**Odds Ratio (95% Confidence Interval)**	**McFadden's R^2**	**P-Value**
Cortex: Insula Score (0 to 1)	**21.20 (2.86, 157.05)**	**0.181**	**0.003**
Periventricular White Matter Score (0 to 3)	**4.18 (1.32, 13.17)**	**0.114**	**0.015**
Basal Ganglia Score (0 to 3)	**2.95 (1.22, 7.15)**	**0.100**	**0.017**
Cortex: Central Sulcus Score (0 to 1)	**9.91 (1.39, 70.74)**	**0.095**	**0.022**
Cortex: Interhemispheric Fissure Score (0 to 1)	3.97 (0.78, 20.18)	0.045	0.097
Posterior Limb of Internal Capsule Score (0 to 3)	0.45 (0.16, 1.31)	0.034	0.143

Table 3.3. Univariate assessment of MRI predictors of outcome using Rutherford scoring. All predictors were treated as ordinal variables. Features found to be significant via univariate logistic regression ($p<0.05$) are bolded.

3.4.4 Multivariate Assessment

With respect to outcome, insula injury was the most significant MRI predictor and the presence of state changes on EEG was the most significant EEG feature. When both predictors were combined in multivariate and multimodal regression model, both were independent predictors of poor outcome (Table 3.4; insula injury, odds ratio = 17.99, p=0.006; EEG state change absence, odds ratio = 7.13, p=0.038). A McFadden's pseudo-R^2 of 0.25 indicates a good model fit. When patient outcome distribution is assessed using both metrics, patients with neither insula injury nor absence of state change tended to have more favorable outcomes (mode outcome score 0), while patients with both insula injury and absence of EEG state change had uniformly poorer outcomes (Figure 3.5A). Patients with either insula injury or absence of EEG state change, but not both, had outcomes that were more dispersed across the range (Figure 3.5A).

When all significant features were considered, the best overall performing combination was insula injury and basal ganglia injury (Table 3.4; insula injury, odds ratio=22.94, p=0.004; basal ganglia injury, odds ratio=2.92, p=0.026). The pseudo-R^2 value here was 0.27 indicating a potentially better model fit. Basal ganglia severity and insula injury presence also appears to be able to separate patients into three distributions with patients displaying both insula injury and moderate/severe basal ganglia injury experiencing uniformly bad outcomes Figure 4B. In contrast, most infants without insula injury and no/mild bad ganglia injury had no markers of abnormal outcome.

Predictors Used	Odds Ratio (95% Confidence Interval)	McFadden's R^2	P-Value
Cortical Insula Injury	17.99 (2.28, 142.1)	0.25	0.006
State Change at 24 Hours	7.13 (1.11, 45.6)		0.038
Cortical Insula Injury	22.94 (2.78, 189.6)	0.27	0.004
Basal Ganglia Injury	2.92 (1.14, 7.48)		0.026

Table 3.4. The two significant and independent multivariate algorithms. Row 1 – Multivariate assessment results using the top EEG and MRI predictors: insula injury and absence of EEG state change. Row 2 – Multivariate assessment results using the other significant multivariate pair: basal ganglia injury score and insula injury score.

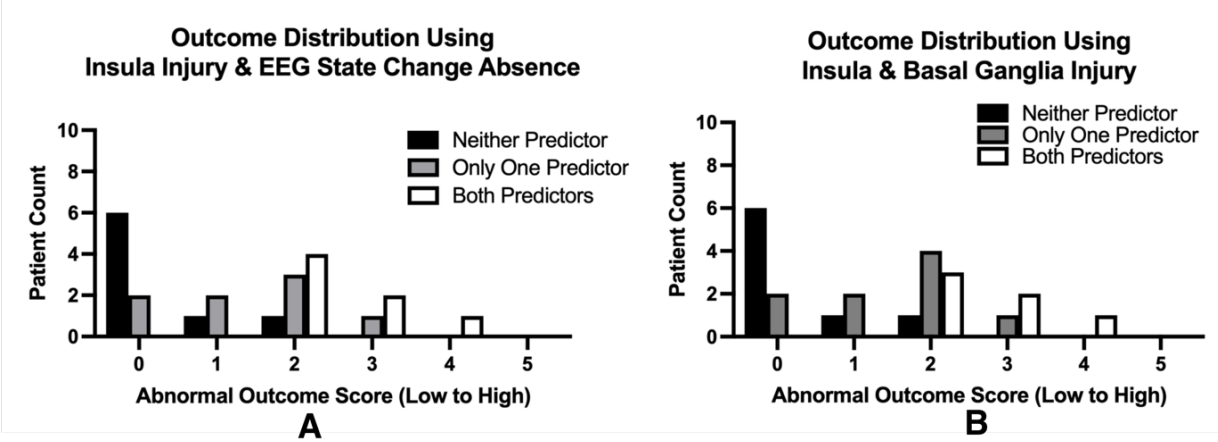

Figure 3.5. Patient outcome distribution based on the two multivariate predictor models. A. Distribution using insula injury and EEG state change. B. Distribution using insula injury presence and basal ganglia injury severity (none/mild or moderate/severe).

3.5 Discussion

In a cohort of 23 term neonatal patients with HIE, injury to the insula identified by DWI-MRI and lack of state change on EEG at 24 hours of life were independent multimodal predictors of poor outcome at three to six months of age.

3.5.1 Outcome and Feature Parameter Modifications

Parameterization of the outcome variable and several of the independent variables differed from previous studies. Most previous studies [101,107] consider outcome as binary, either good or bad, based on a cutoff from developmental scores such as the Bayley Scale of Infant Development-II (BSID-II) Mental Development Index. This study instead used an ordinal outcome score taken as the number of abnormalities in five domains: the higher the score, the worse the outcome. The gradation of outcomes in a multivariate model allows for higher fidelity prognostication compared to an outcome model with a single predictor, and thus may be more clinically useful (Figure 3.5).

In addition, this study adapted the Rutherford MRI injury scoring system. This system places an emphasis on lesion severity within deep structures like the basal ganglia and thalamus, and it combines lesions from varied cortical areas into a single variable, which would obscure whether damage to a particular cortical region predicts poor outcomes[129]. In the current study, injury in three cortical territories was assessed separately, and injury in each area was tested as an independent variable in univariate analyses. This yielded the finding that injury to the insula was most predictive of poor outcomes.

3.5.2 Use of an Early Follow-Up Assessment

The decision to use an early follow up date within the three-to-six-month window was to create a useful prognostic tool that could address the issue of long term follow up loss and reflect the window for standard neurological follow-up tests for abnormal movement. Standard clinical follow-up for neonates with HIE involves an early follow-up between three to six months and a long-term follow-up between one and two years of life [132]. However, infants from families with limited socioeconomic resources, a population with a high incidence of HIE, are also more likely to be lost to follow-up by 18 months [22,133]. Studies have found follow-up rates among neonates within the first year as well below 40% [22,23]. Thus, a major challenge of outcome metrics such as the Bayley Scales of Infant Development (BSID) is the need for long term follow. High attrition rates are particularly concerning because assessments such as the BSID are based on normative scores from the population of children that do receive follow-up care.

This study aims to bridge that gap by assessing outcome metrics on a graded scale using conditions that can be readily observed in early follow-up and that have also been shown to persist even in long term follow-up studies. These include the assessment for post-neonatal epilepsy via use of antiepileptic medication, early markers of cerebral palsy via appendicular and axial abnormal tone markers, and early markers of hearing and vision problems [134–137].

3.5.3 Feature Relevance to HIE Pathophysiology

In the pathophysiology of HIE, secondary energy failure is thought to occur 6 to 48 hours after the initial hypoxic injury, and is characterized by an ATP deficit, excitotoxicity, cytotoxic edema, and neuronal cell death [138]. An EEG lacking normal graphoelements or reactivity may be an electrographic marker of severity of secondary energy failure. Indeed, sleep wake cycles, a form

of EEG state change, are present in neonates with more favorable outcome [100]. This study shows that both absence of EEG state changes after the first 24 hours and prolonged interburst intervals at 24 hours were important markers for poor outcome in neonates. These markers could be an effect of the underlying pathology of HIE and/or an effect of another confounder like phenobarbital treatment [139]. However, while phenobarbital administration was associated with prolonged interburst interval duration, it was not associated with the presence or absence of state changes.

3.5.4 MRI Injury Predictors

Prior studies suggest that injury to the basal ganglia/thalamus is the best predictor of poor outcome [109,115]. This dataset, however, suggests that injury to the cortical insula may better predict poor outcome than injury to the neighboring basal ganglia. Both the insula and basal ganglia receive blood supply from branches of the middle cerebral artery (MCA), so injury to both regions is likely if there is significant compromise of the MCA. Unsurprisingly, outcomes are poor when both regions are injured Figure 3.5B. Three patients had insular injury without basal ganglia involvement; all had an outcome score of two which signifies moderately poor outcome. Despite this, the small number of patients with pure insular injury makes it difficult to determine if insular injury alone leads to poor outcomes. Finally, while periventricular white matter injury is the characteristic injury pattern in premature infants, this dataset suggests that periventricular white matter injury might also play a role in the outcome of term infants.

Even when univariate logistic regression is applied to only the subset of infants that received MRIs on day five of life and below (n=14) to address the potential for post day five MRI pseudonormalization as a confounder, injury in the insula (p=0.017), periventricular white matter (p=0.019), and basal ganglia (p=0.024) are still significantly associated with outcome. Since

outcomes have not been previously associated with injury to specific cortical areas, future studies may better delineate the role of injury to the cortical insula as a biomarker for HIE outcomes.

3.5.5 Study Limitations

The primary study limitation is the small sample size. With 23 neonates, only two predictors could be tested in a multivariate model, despite the identification of three significant predictors of outcome on univariate analysis. A second limitation is the qualitative nature of the EEG and MRI scoring systems deployed. While these assessments can be performed by trained clinicians, they may be more difficult to standardize across clinicians, which can lead to poor interrater reliability. Additionally, when scoring the MRI regions, cortical regions were scored according to presence or absence of injury but not by injury severity, and no distinction was made between unilateral or bilateral lesions. Ultimately, a larger, prospective trial would allow additional parameters to be tested in a multivariate analysis.

Follow-up at three to six months may be considered a limitation since this follow-up occurs before the milestones of walking or talking emerge. Despite this possibility, earlier follow-up allows this study to include a higher risk population that is more likely to be lost to follow-up by 12 months.

3.6 Conclusions

In a cohort of 23 neonates with HIE from a single center, the combination of absence of EEG state change and insular injury on MRI was an indicator of poor short-term outcomes. When neither feature was present, outcomes were typically good.

CHAPTER 4

NEONATAL SEIZURE DETECTION USING AEEG AND CSA ENSEMBLE ALGORITHM

4.1 Abstract

Objective

To build a clinically translatable neonatal seizure detection algorithm using EEG trending tools (aEEG and CSA) that are currently applied at the bedside for manual neuromonitoring with the goal of reducing the morbidity associated with uncontrolled neonatal seizures.

Methods

Using a public dataset of annotated neonatal EEGs, features of the aEEG and CSA were extracted from bipolar channels (C3-P3 and C4-P4). These features were then used to train and test three machine learning classifiers, Random Forest, Support Vector Machines, and Artificial Neural Networks.

Results

The trained RF, SVM and ANN classifiers had an AUC of 0.80, 0.71, and 0.80 and an average accuracy of 0.91, 0.90, and 0.92 respectively. Median accuracy scores were higher among non-HIE patients (median = 1) than HIE patients (median = 0.92, 0.93, 0.93).

Conclusion

An aEEG-CSA algorithm is feasible for neonatal seizure detection by extracting clinical features and coupling these features with a supervised ML classifier.

4.2 Introduction

Among all age groups, neonates have the highest risk of developing seizures [140,141]. Thus, neonatal seizure detection is a major challenge. Additionally, most newborns have subclinical seizures, so clinicians cannot depend on clear visual clinical cues to detect neonatal seizures [142]. For this reason, they must rely on long term electroencephalography (EEG) monitoring for seizure identification. But 24-hour EEG monitoring, presents the added challenge of generating massive volumes of EEG data that must be reviewed by clinicians with specialized expertise. Unfortunately, these experts are in short supply [143].

Despite these challenges, it is imperative that seizures be detected in a timely manner. Neonates with seizures have seizure burdens proportional to the length of time allowed to pass between their first seizure and the administration of antiepileptic medication [57]. High seizure burden increases the risk of adverse outcomes [10,56], so prompt and accurate identification of seizures using is critical. To expedite EEG review, quantitative EEG (qEEG) trending analysis is used to time compress and transform EEG signals, so seizures can be manually identified using pattern recognition.

Amplitude-Integrated EEG (aEEG) and the Compressed Spectral Array (CSA) are two of the most used EEG trending analysis tools in Intensive Care Units (ICU). aEEG is the standard long term monitoring trending tool used in the neonatal ICU [38,42]. The aEEG transforms and time compresses the EEG to enable the clinical team to detect seizures by identifying abrupt rises in the aEEG upper and lower margins [44,144]. Despite its popularity, aEEG has notable drawbacks, which include the need for manual review, low sensitivity for seizure detection, variability across

different proprietary aEEG algorithms, and the inability to assess seizure burden [34,145]. Thus, aEEG is often only used as a supplement to the EEG time series.

The CSA, another qEEG trending tool, displays changes in the EEG power from the delta to beta band (0.5-30Hz). Previous studies have shown spectral analysis can be useful for seizure detection in adults[54,55,146,147], but has not been validated in neonates. The most easily identifiable seizure pattern on the CSA is referred to as the flame pattern, which is a sudden increase in power across a range of frequencies that contrasts sharply with the background activity [148]. In practice, the CSA like the aEEG is used alongside the EEG timeseries to expedite review.

In this study, features of the aEEG and CSA are extracted to train and test supervised machine learning (ML) classifiers for neonatal seizure detection. The aim of this work is to serve as a proof of concept that a combination of aEEG and CSA can be used for detecting neonatal seizures and determining seizure burden.

4.3 Methods

4.3.1 EEG Preprocessing

EEG data was collected from a publicly available database of neonatal EEG recordings from Helsinki University [149]. This dataset contains annotated EEG recordings sampled at 256 Hz from 79 neonates, 39 of whom had at least one seizure. EEG recordings were annotated separately for seizures by three clinicians. Annotations on seizure location were not included. Each recording was evaluated using the consensus ('and') of all three reviewer labels for training and testing as it provides the most stringent seizure classification.

EEGs were bandpass filtered from 0.5 to 40 Hz using a 6-pole Butterworth filter and MATLAB function filtfilt for zero-phase distortion. Bipolar referenced electrode pairs C3-P3 and C4-P4, typically used in calculating qEEG trends in neonates, were used. The parietal and central electrodes have the greatest sensitivity for detecting neonatal seizures[29,33,34]. An example raw EEG of patient four from the inhouse UChicago cohort is illustrated in Figure 4.1A. EEG channel C4-P4 with consensus annotation is plotted in Figure 4.1A. All data analysis was completed using MATLAB 2022b and GraphPad Prism.

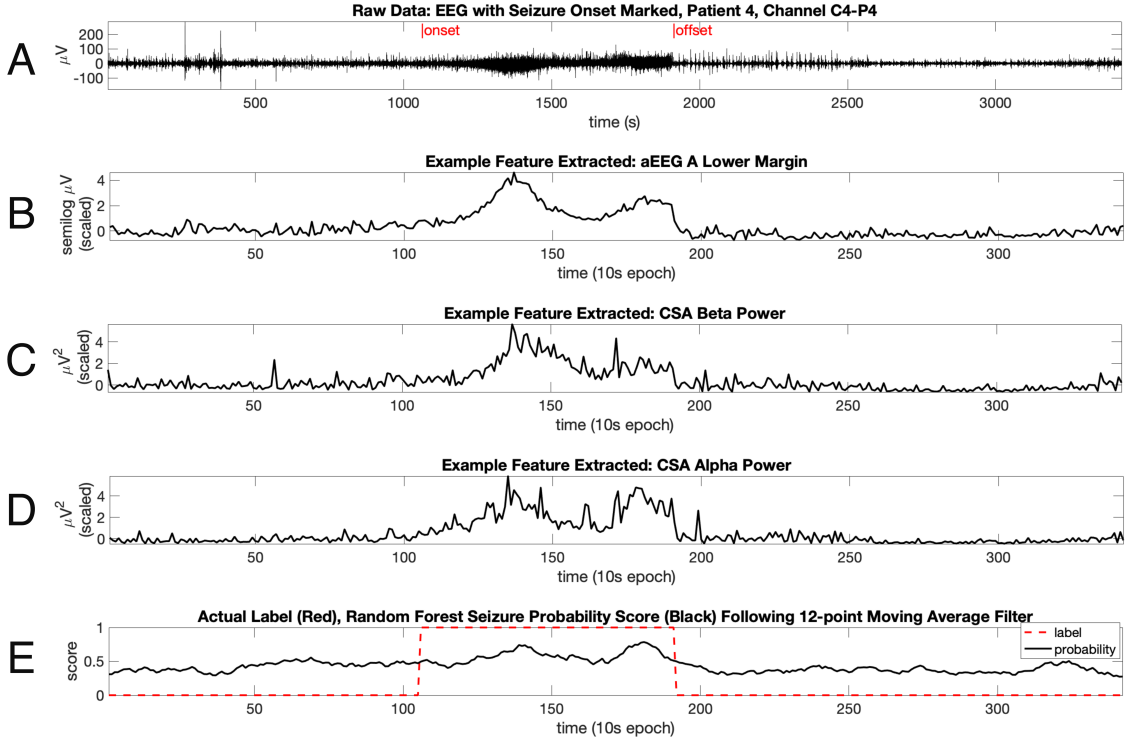

Figure 4.1 Example features extracted from the annotated EEG of patient four along with seizure probability scores. A. EEG time series of C4-P4 with seizure onset and offset marked using consensus label. B. Lower margin of aEEG A feature shows pattern of a rise in amplitude trend during seizure epochs. D. CSA alpha power feature shows increased alpha power during seizure windows. E. Random Forest classifier probability scores for each 10-sec window in black with consensus labels for each window in red dashed lines.

4.3.2 Feature Extraction

aEEG and CSA Feature Extraction

Eleven features were extracted separately from the left and right centroparietal electrodes: six aEEG features and five CSA features. 10-second nonoverlapping EEG windows were chosen for feature extraction since 10-seconds is the minimum required duration for a neonatal seizure. Each 10-second epoch was represented by a vector of features from the aEEG and CSA.

The procedure of Zhang et al 2013 was adapted to compute the aEEG margins and aEEG envelope [150]. Briefly, the EEG was filtered using an asymmetric bandpass filter (2-15Hz), absolute value rectified, and smoothed with a moving average filter (Fig. 2B). The EEG was then segmented into 10-second windows and the amplitude distribution in each window was calculated to extract the upper (93rd % sample) and lower (9th % sample) amplitude values. The amplitudes were converted into a semilogarithmic scale, with amplitudes larger than 10 microvolts replaced by the base-10 logarithm and those below 10 microvolts kept their value.

Two changes were made from the published protocol of Ding and Zhang 2013. One was increasing the temporal resolution of the aEEG. While Ding and Zhang 2013 computed the aEEG in 5-minute windows to mimic the resolution of the clinical aEEG[37], a 10-second window was chosen since it is more optimal for capturing neonatal seizures and allows for finer seizure resolution. The second change was to generate two aEEGs – aEEG A and aEEG B – that differ by the duration of the moving average window (Table 4.1). Werther et al found that the smoothing filter accounted for differences in the performance among proprietary aEEG algorithms for various computational tasks[48]. Thus, six aEEG features were extracted for every 10-second EEG window:

the upper margin, lower margin and median envelope using the aEEG A algorithm and aEEG B algorithm.

aEEG Version	*Reference*	*Asymmetric Bandpass Filter (2-15Hz)*	*Enveloping Filter*
aEEG A	Werther et al 2018	Linear Phase FIR filter (Parks-McClellan) Filter 2	Rectangular Moving Average Filter (3-second window)
aEEG B	Werther et al 2018	Linear Phase FIR filter (Parks-McClellan) Filter 5	Rectangular Moving Average Filter (0.5-second window)

Table 4.1: Documented parameters for generating two versions of the aEEG.

To make a Compressed Spectral Array (CSA), the power spectra across each 10-second window of the preprocessed EEG was computed. From each spectrum, the power across each clinical band – delta, theta, alpha, and beta was computed from both channels. In addition, the spectral slope across the delta to beta (0.5-30Hz) was computed using Matlab's spectralSlope function.

4.3.3 Artifact Removal and Data Scaling

To reduce the effect of electrical and movement artifact on seizure detection, any epoch with an aEEG lower margin value below $0.001\mu V$ or with total power below $0.001\mu V^2$ in the 0.5-30Hz band was removed. Any epoch with an upper margin or total power greater than five times the median value for the patient was removed. Finally, each feature was scaled per patient using the median and interquartile range such that each feature had an interquartile range of one.

4.3.4 Model Training, Testing, and Post-Processing

Patients were split into three groups for 3-fold cross validation using a stratified K-Fold to ensure the proportion of seizure and non-seizure patients was equivalent across all three-fold groups. Three-fold cross validation was chosen because the dataset was relatively small (40,251 epochs total with 39,335 epochs following artifact removal). Following artifact removal, synthetic seizure data was added to the training data using Adaptive Synthetic Oversampling Technique (ADASYN) for imbalanced classification [151,152].

Training and testing were done using three supervised ML algorithms: Random Forest (RF), Support Vector Machine (SVM, and an Artificial Neural Network (ANN). The RF, SVM, and ANN classifiers were selected for their ability to assess both linear and potentially nonlinear decision boundaries between seizure and nonseizure epochs. The RF classifier was generated using MATLAB's fitcensemble function using the bagging method with all 11 feature variables selected for sampling per tree. An SVM classifier with a Radial Basis Function kernel was deployed using MATLAB's fitcsvm function [153]. SVM using an RBF kernel has been shown to be effective for neonatal seizure detection with a different feature set [58]. The predictors were standardized using weighted standardization to improve performance [153]. For the ANN, a feedforward neural network consisted of three fully connected layers, with a ReLu activation layer applied to the first fully connected layer and a SoftMax function applied to the final activation layer [154]. Z-score feature standardization was selected for each ANN via hyperparameter tuning. The ANN was created and deployed using MATLAB's fitcnet classifier [154].

Since neonatal seizures are often focal, two models were generated per classifier for training and testing – one was trained and tested on features from the left electrode and the other

model on the right electrode. Following testing, seizure prediction probability scores were smoothed using a 12-point (120 second) moving average filter. The maximum filtered probability scores across both electrodes were then compared to consensus seizure labels (Figure 4.1E) and used to assess algorithm performance via receiver operating characteristic curve (ROC) analysis. The optimal seizure probability threshold per classifier was determined using MATLAB's perfcurve function for ROC analysis.

Feature importance analysis was conducted using the RF Out-of Bag permutation entropy via MATLAB's oobPermutedPredictorImportance function [155,156]. This method uses the out-of-bag error, which measures the RF classifier's performance on samples not used to train each decision tree in the random forest. Features are scored by permuting each feature variable and assessing whether removing the feature increases or decreases the error. More important features cause greater increases in the error. The mean importance score was taken between the two aEEGs.

4.4 Results

4.4.1 Patient-Independent Performance

The RF and ANN classifiers performed better than did the SVM classifier (Figure 4.2A; mean AUC performance across all folds was 0.8, 0.71, and 0.8 for the RF, SVM and ANN classifiers respectively). The dataset contained 39,335 non-artifact epochs total, and out of those epochs, 4,224 were seizure epochs. Since the data was imbalanced, a precision-recall curve was also used to assess performance (Figure 4.2B). When compared to the mean PR-AUC of 0.11 for a nonsense classifier, the mean PR-AUC of each classifier was 0.46, 0.28, and 0.5 respectively.

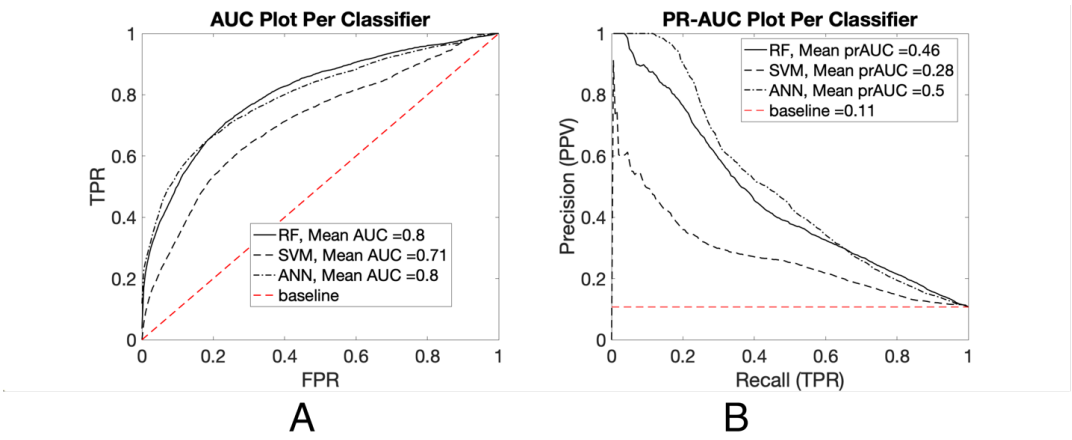

Figure 4.2 Patient-independent performance of the aEEG-CSA detection algorithm across the RF, SVM, and ANN classifiers. A. Performance measured by ROC AUC for RF (solid line), SVM (dashed line), ANN (dash-dot line). The red dashed line represents chance level of classification. B. Performance measured by precision-recall curve.

4.4.2 Feature Importance for Seizure Classification

To determine which features contributed most to classification, we estimated importance scores for the RF classifier. The mean out-of-bag permutation importance scores were largest for CSA Beta power, the aEEG Lower Margin, and CSA alpha power among. These three features, and their association with seizure activity in one case, can be seen in Figure 4.1 panels B, C, and D.

4.4.3 Per Patient Performance Assessments

Algorithm performance was also assessed across individual patients. The mean accuracy scores mean for the RF, SVM, and ANN classifiers were 0.91,0.90, and 0.92 respectively (Figure 4.4A). These accuracy scores were determined using the optimal patient-independent probability threshold for each classifier based on the ROC curve. These probability thresholds were 0.56, 0.75, 0.73 for the RF, SVM, and ANN classifiers respectively.

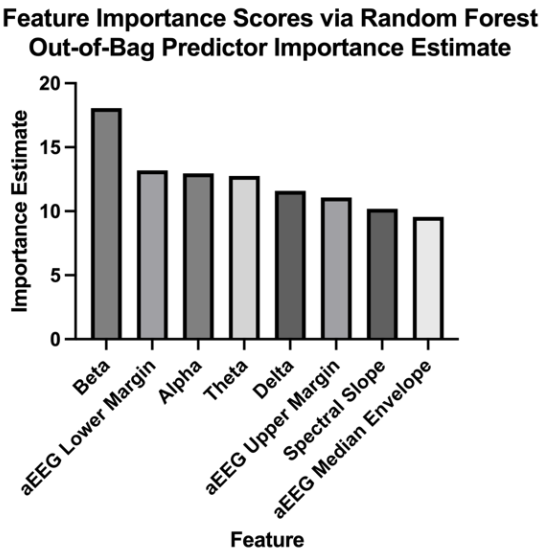

Figure 4.3 11 Extracted features scored for importance using RF out-of-bag permutation importance estimate. Rankings show that the CSA beta power, aEEG lower margin, and CSA alpha power were the three most important features for classification.

Within the group of 79 neonatal EEG used to train and test the classifiers, there were various disease etiologies that triggered monitoring. Hypoxic Ischemic Encephalopathy (HIE) due to birth asphyxia was the largest subgroup with 35 patients; 24 had seizures. To ascertain if the median accuracy scores were significantly different from those without HIE, multiple Wilcoxon signed rank tests were applied to compare median accuracy for each classifier. P-values were corrected for multiple comparison using the Bonferroni-Dunn correction (Figure 4.4B). Infants with HIE due to birth asphyxia diagnosis had significantly lower accuracy scores than those without birth asphyxia.

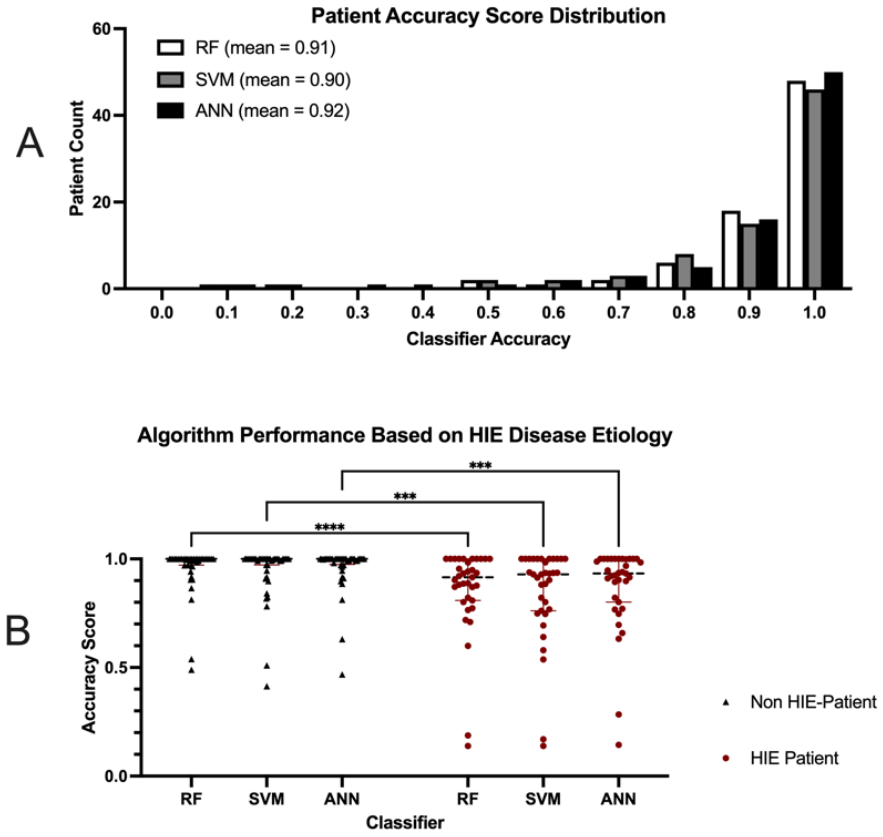

Figure 4.4 A. Histogram of accuracy scores for all 79 patients using patient independent seizure probability threshold. B. Comparing median accuracy scores for each classifier based on asphyxia disease etiology. Black triangles are non-asphyxia patients (n=44) and red circles are patients with HIE due to birth asphyxia (n=35). HIE patients have significantly lower median accuracy scores (median = 0.92, 0.93, 0.93 for RF, SVM, and ANN) than non-HIE patients (median =1 for all three classifiers).

4.4.4 Effect of Seizure Duration on Algorithm Performance

Seizure duration has been cited as a major cause of poor manual detection by clinicians[34,157]. To determine if seizure duration is also a determinant of detection accuracy with the automated algorithms, the effect of seizure duration on performance was tested using the 39 seizure patients included in this dataset. The 39 seizure patients were separated into four quartiles based on average seizure duration (Figure 4.5). Durations ranged from 10 seconds to 918 seconds with a median of 87 seconds. There were 10 patients in each group except group 2, where there were only nine

patients. Using the Kruskall-Wallis Test, the median AUC of each group was evaluated. While Figure 4.5 shows a trend of increasing AUC across groups with higher seizure durations, AUCs were only significantly different for groups 1 and group 4 when the ANN classifier was used.

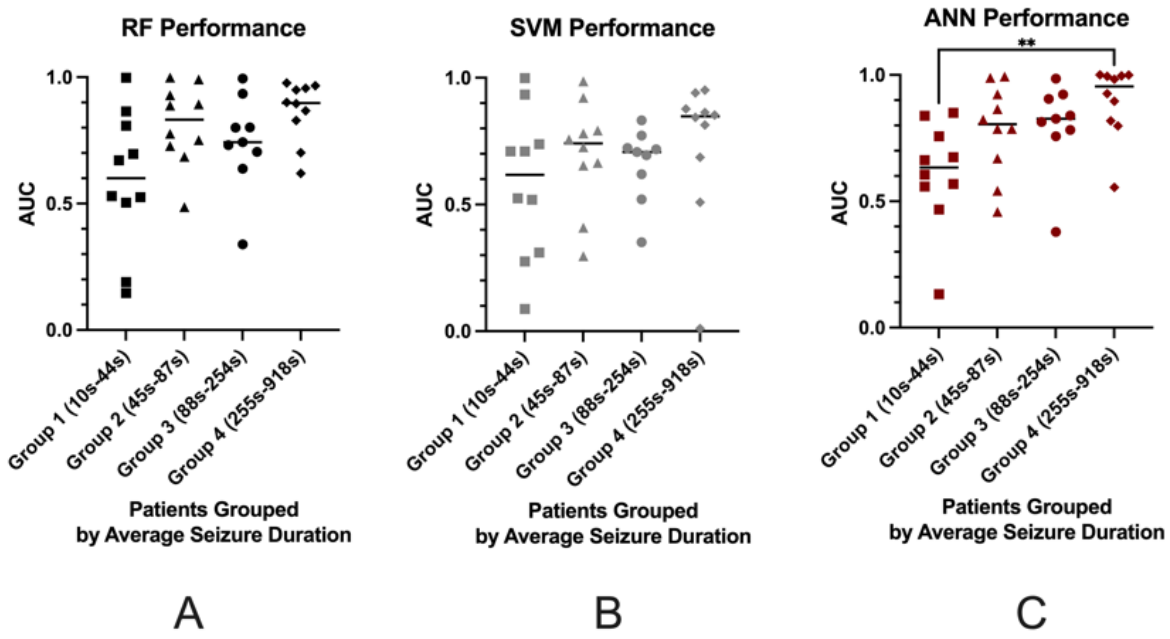

Figure 4.5 Algorithm performance based on seizure duration by grouping the 39 seizure patients by average seizure duration. Each classifier shows a trend of increasing AUC as the average seizure duration increases. Panels A, B, and C show the results of the RF, SVM, and ANN classifiers, respectively. Only the ANN showed a significant difference between the median AUC group 1 (n=10) and group 4 (n=10) based on the Kruskal-Wallis test.

4.5 Discussion

This study generated a neonatal seizure detection algorithm that combined of features from the aEEG and CSA. The findings have several implications. First, it showed that neonatal seizure detection can be done with the existing qEEG trends despite the problem of low seizure resolution during manual implementation. Zhang and Ding et al 2013 used a 5-minute window to calculate

the aEEG [150]; the good seizure detection rate with 10-second windows implies that the clinical aEEG would struggle to resolve seizures below five minutes, which represents most neonatal seizures[158]. Indeed, shorter duration seizures (<30 seconds) are most often missed entirely on aEEG [34,157]. In the case of the models used in this study, Figure 4.5 shows that AUC appears to increase with patient average seizure, though this association was only significant between patients with the shortest average duration (Group 1) and those with the longest average duration (Group 4).

Additionally, the CSA findings are particularly interesting because they show help highlight frequency bands that may be of importance for seizure detection. In a previous study, peri-ictal alpha activity was associated with seizure risk and reoccurrence in a small prospective cohort of cooled neonates with hypoxic ischemic encephalopathy (HIE) due to birth asphyxia, the primary cause of neonatal seizures [159]. This study shows that both alpha and beta activity does have importance in neonatal seizure detection and should be further studied for their utility in identifying infants with seizures and potentially HIE infants specifically during cooling.

Finally, the left and right centroparietal electrodes (C3-P3 and C4-P4) were chosen for training and testing because they have the greatest sensitivity for capturing seizures [29,160,161] and are part of the standard set used for expedited neonatal EEG interpretation. Since the two-channel electrode display is the most limited standard setup, it can even be used in both under-resourced settings and cases where neonates are too unstable for more electrodes. This study shows that an aEEG-CSA algorithm can perform well despite the challenges of manual assessment using limited electrodes such as slight reductions in manual seizure detection [162–164].

4.5.1 Potential for Improving Sensitivity and Specificity

A major challenge in neonatal seizure detection is not only improving algorithm sensitivity but also improving specificity. Low specificity leads to a high false positive rate. The next steps will be to improve algorithm false positive rates. In comparison to other studies that have several hours of nonseizure EEG data per patient [165,166], the amount of true negative nonseizure data to learn from was often less than an hour. Thus, there was greater potential for false positives. This challenge may be rectified with a larger dataset and longer overall recording times per patient.

The assessments presented here were stringent in establishing the confusion matrix for seizure detection. A true positive was defined as a window of time that was correctly assessed as a seizure. A true negative was a nonseizure window period correctly defined as nonseizure. A false positive was a nonseizure window incorrectly defined as a seizure window, and a false negative was a seizure window incorrectly defined as a nonseizure window. Loosening the definition of a true positive to that of a Good Detection where a seizure is considered captured if any single epoch or any prespecified number of epochs within the seizure was flagged as seizure will serve to increase the sensitivity of the algorithm [55,58]. Additionally, allowing for tolerance around the seizure boundary times so epochs classified immediately before consensus label are not being flagged as a false positive. These modifications will be carried out in future studies with careful focus on the clinical utility of the true positive definition and allowable limits of the seizure boundary.

4.6 Conclusions

This is a proof-of-concept study to assess whether a combination of extracted clinical aEEG and CSA features can be used to capture time periods with seizure occurrence in a heterogeneous cohort of neonates. This study demonstrated that a combination of aEEG and CSA features can be used to capture periods of neonatal seizure activity with an AUC of 0.8 for the Random Forest and ANN classifiers. The CSA beta power, aEEG lower margin and CSA alpha power were the three most efficacious features for seizure detection among the seizure patients. Finally, this study also showed that performances were higher in infants with shorter duration seizures and infants who did not have HIE due to birth asphyxia.

CHAPTER 5

VALIDATION OF aEEG-CSA ENSEMBLE SEIZURE DETECTION ALGORITHM ON HIE NEONATES DURING THERAPEUTIC HYPOTHERMIA

5.1 Abstract

Objective To validate a neonatal seizure detection algorithm that is based on extracted clinical features of the aEEG and CSA on a cohort of neonatal patients with HIE.

Methods

A seizure detection algorithm was designed using aEEG margin features, CSA Berger Band features, trained on a public dataset of 79 neonatal EEGs with three supervised machine learning classifiers. It was subsequently tested on an inhouse cohort of 23 neonates with asphyxia whose EEGs were collected during hypothermia therapy.

Results

The trained Random Forest Classifier, Support Vector Machines and Artificial Neural Network classifiers had an AUC of 0.79, 0.79, and 0.76 and an average accuracy of 0.85, 0.86, and 0.85 respectively. Beta power, aEEG lower margin amplitude, and alpha power were the most important features for classification.

Conclusion

A neonatal seizure detection algorithm that uses a combination of aEEG and CSA clinical features can capture seizures in HIE patients. Performance was not improved by training on non-HIE patients.

5.2 Introduction

Hypoxic Ischemic Encephalopathy (HIE) is the primary cause of neonatal seizures [167]. HIE is not only responsible for 60 to 80% of neonatal seizures, but it also represents a major global disease burden since birth asphyxia is associated with about a quarter of all neonatal deaths [56,167,168]. Electroencephalography (EEG) is necessary to capture and treat neonatal seizures, but because EEG interpretation is time consuming, it slows down clinical management during the time sensitive neonatal period. Clinicians are increasingly turning towards quantitative electroencephalography (qEEG) tools to expedite seizure detection. Thus, there is a need for sensitive and specific seizure detection algorithms especially in infants with HIE to expedite clinical management and potentially reduce the morbidity and mortality associated with neonatal HIE.

The objective of this study is to validate a neonatal seizure detection algorithm using EEG recordings from a cohort of 23 term neonates with HIE, while they received therapeutic hypothermia in the first few days of life. This algorithm implements clinical features extracted from the two widely used bedside qEEG trending tools, the amplitude-integrated electroencephalography (aEEG) and the Compressed Spectral Array (CSA). Currently, these tools require manual assessment by clinicians to find EEG periods that might be suspicious for seizure activity. In this study, the clinical features of aEEG and CSA have been extracted and used to generate a seizure detection algorithm and quantify seizure burden.

Seizure burden is particularly important to capture accurately during therapeutic hypothermia treatment for HIE because high seizure burden has been associated with poor neurocognitive outcome, and infants' brains are most vulnerable to seizure activity within the first

few days of life [11,108]. Thus, there is a need for sensitive and specific algorithms that can detect seizures especially in HIE patients. The goal of this work is to evaluate the aEEG-CSA algorithm's ability to detect seizures in HIE patients during therapeutic hypothermia when the algorithm is trained on all patients in the public dataset or only on the subset of birth asphyxia patients in the public dataset.

5.3 Methods

5.3.1 Datasets for Training and Testing

EEG recordings from a cohort of 23 term neonates with HIE were collected retrospectively to generate a dataset for testing (Table 5.1). All recordings were taken during hypothermia therapy within the first 48 hours of life and extracted from Natus Neuroworks EEG software. For the testing set, EEG recordings were annotated by a pediatric neurologist (JH) with extensive expertise in neonatal epilepsy and over ten years of experience. Seizure annotations for the testing cohort were also cross-referenced with the neonatal EEG chart report taken during hypothermia therapy and read by other experienced pediatric neurologists at the UChicago Comer's children's hospital.

The training dataset was taken from a public dataset of neonatal EEG recordings from Helsinki University [149]. This dataset was annotated by three reviewers. The consensus annotation label was used for training in this study. Birth asphyxia was the underlying disease etiology for 35 patients in the dataset.

Patient Number	**Gestational Age (Weeks)**	**Seizure Burden (min)**	**Recording Duration (min)**
1	40.3	10.6	45.3
2	35.6	46.0	60.7
3	40.1	0	60.1
4	37.0	0.97	41.9
5	40.0	3.40	59.8
6	40.3	37.6	46.5
7	40.0	24.2	60.4
8	41.0	68.6	145.8
9	38.6	0	60.2
10	38.0	0	60.1
11	40.4	60.3	239.2
12	38.6	0	60.1
13	38.7	0	60.1
14	38.1	0	60.1
15	39.0	0	60.1
16	40.4	0	60.1
17	39.9	0	60.1
18	38.9	0	60.1
19	38.4	0	60.1
20	41.1	26.5	193.5
21	39.3	0	60.0
22	39.1	96.9	367.1
23	40.4	0	60.1

Table 5.1: Gestational age, total seizure burden represented by minutes of annotated seizure activity, and recording length for patients in HIE inhouse validation group.

5.3.2 Preprocessing, aEEG-CSA Neonatal Seizure Detection Algorithm

All EEG recordings used were preprocessed by bandpass filtering from 0.5 to 40 Hz using a sixth-order Butterworth filter and the Matlab function filtfilt to prevent phase distortion. EEGs were placed in the bipolar montage using centroparietal electrodes C3-P3 and C4-P4 (Figure 5.1A). These electrodes are most sensitive for capturing neonatal seizures [29,33,34]. Computations were performed using MATLAB 2022b and GraphPad Prism.

To generate the neonatal seizure detection algorithm, 11 clinical features were extracted from the two centroparietal electrodes. Each feature was computed across 10-second nonoverlapping EEG windows. Six aEEG clinical features were collected after computing a generic aEEG using the procedure adapted from Zhang and Ding et al 2013 [37], with 10-second windows used for feature extraction instead of the five minutes windows used in the original publication. These features are the aEEG upper margin, lower margin, and median envelope for two versions of the aEEG (aEEG A and aEEG B). These versions were generated by implementing different proprietary domain algorithm parameters extracted from studies published in the literature (Table 4.1) [48]. For the CSA, five spectral features were collected. These features are the power across each of the Berger Bands (delta, theta, alpha, beta) and the spectral slope.

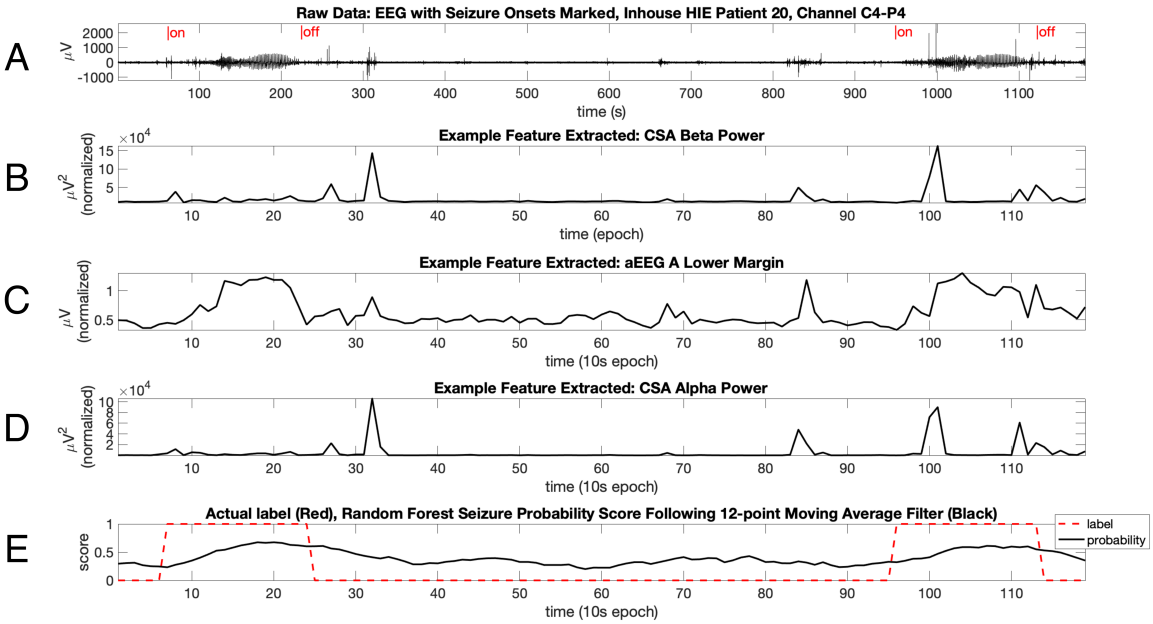

Figure 5.1 A. EEG, features, and output from inhouse HIE Patient 20. Channel C4-P4 plotted with EEG band-pass filtered from 0.5-40Hz. Seizure onsets and offsets marked using clinician label. Panels B, C, and D shows top three features of importance plotted for the same EEG epoch. Each features shown increases around seizure epoch. E. RF Classifier seizure probability scores following training and testing in black and true labels in red.

5.3.3 Artifact Removal and Data Normalization

Artifact removal was conducted using the extracted features. Epochs were removed if their aEEG lower margin values were less than $0.001\mu V$ or total power from 0.5-30Hz was less than $0.001\mu V^2$. Upper threshold artifacts were removed per patient. This process was calculating the mean upper margin and mean total power from 0.5-30Hz across all the epochs of a given patient. Epochs were removed if their upper margin values or total power were greater than five times either of these mean values. Finally, scaling was also done per patient. Scaling was accomplished by subtracting each feature by its median and dividing by the interquartile range.

5.3.4 Model Training, Testing, and Post-Processing

Random Forest (RF), Support Vector Machines (SVM), and feedforward Artificial Neural Network (ANN) were used for training and testing on MATLAB. To address the problem of class imbalance, synthetic seizure data was added to the training data using Adaptive synthetic Sampling [169]. The RF classifier was deployed using the MATLAB fitcensemble function with bagging method selected and using 100 decision trees. To facilitate feature selection assessments, all 11 variables were included for sampling across every decision tree. The SVM was applied using the fitcsvm function using a radial basis function kernel and feature standardization was selected [170]. The ANN was implemented using the fitcnet function with the default neural network architecture used [154]. Z-score feature standardization was selected for using hyperparameter tuning.

Two trained models were generated per classifier. One for the left electrode and the other for the right. Each model was tested and the maximum probability score across both models was used to evaluate algorithm performance. All classifier probability scores were smoothed using a 12-epoch moving average window filter (Figure 5.1E). This window length is based on the average seizure duration of the training data which was 115 seconds in duration since each epoch in this study represented 10-seconds of EEG. It also falls within the range of the average neonatal seizure duration that spans anywhere between one to three minutes [171]. Additionally, ROC curves and patient accuracy scores were assessed using the optimal patient independent probability threshold for each classifier extracted using MATLAB's perfcurve function.

5.3.5 Training using Entire dataset vs Subset of Asphyxia Patients

The entire training, testing, and model evaluation pipeline was conducted twice. The first training and testing assessment was done to evaluate algorithm performance when all 79 patients in the external dataset were used for training. Of these patients 39 had seizures. Testing was then conducted on the 23 HIE holdout patients from an inhouse dataset. The second training and testing assessment was done using models generated and trained on only a subset 35 patient in the external dataset who had birth asphyxia and subsequent Hypoxic Ischemic Injury. Within HIE subset patients, 24 had seizures. Testing was again done on the 23 HIE inhouse patients. These assessments were conducted to see if performance was similar on the HIE validation set by training on patients with the same disease etiology, birth asphyxia, despite having 14 fewer seizure patients for training.

5.3.6 Model Evaluation and Feature Importance Analysis

To promote model explainability, feature importance was also assessed using the out-of bag-error [156]. The out of bag error is the error of the RF classifier when trees are assessed on samples that were not used to train them [156]. A feature's importance can be scored by randomly permuting the feature from each decision tree when assessing the out-of-bag error and seeing its influence on the error. More important features will increase the error when they are absent. This assessment was done using MATLAB's oobPermutedPredictorImportance function [155]. Since there are two aEEGs, importance scores were averaged across the same aEEG features to get one score for each of the three aEEG features: the lower margin, upper margin, and the median envelope.

5.4. Results

5.4.1 Patient-Independent Performance

The patient-independent performance of the algorithm was first tested using models trained on all 79 patients in the external dataset. All three classifiers had an area under the curve (AUC) of 0.76 (Figure 5.2Ai). When the precision-recall AUC (PR-AUC) curve was assessed, performances were more varied with PR-AUC scores of 0.4, 0.41, and 0.47 for the RF, SVM and ANN classifiers respectively (Figure 5.2Aii). These PR-AUC values represent between a 111% to 147% increase over a no skill random classifier represented by a pr-AUC of 0.19, which is the ratio of true positive epochs to total epochs.

When the patient independent performance was analyzed following training on the subset of 35 patients with HIE from birth asphyxia, performances were similar to when the full 79 patients were used for training. The AUC of RF, SVM, and ANN classifiers were 0.79, 0.79, and 0.76 (Figure 5.2Bi.). Additionally, the PR-AUC scores were 0.44, 0.46, and 0.46 which presents between 132% to 142% increase of a no skill random classifier (Figure 5.2Bii).

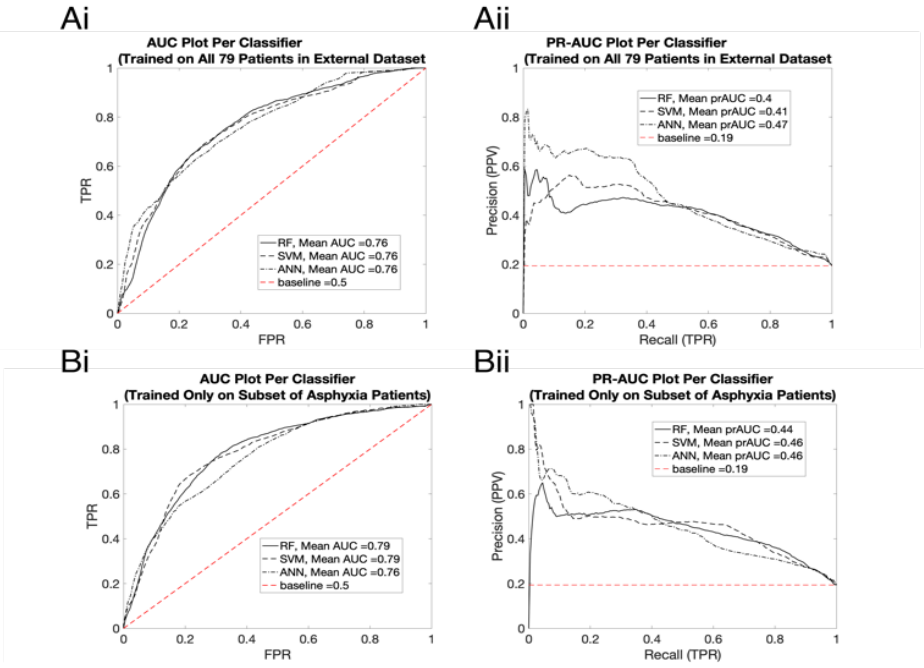

Figure 5.2 Patient-independent performances of aEEG and CSA detection algorithm using the RF, SVM, and ANN classifiers. A. Performance using all 79 patients in the external dataset for training, 39 of whom had seizures. Ai. All three classifiers have 0.76 AUC. Aii. PR-AUC is highest in the ANN at 0.47 147% above baseline performance of 0.19 PR-AUC. B. Algorithm performance following training using only the 35 asphyxia patients within the external dataset, 24 of whom had seizures. Bi. RF and SVM have the largest AUC. Bii. PR-AUC scores; SVM and ANN having the highest AUC.

5.4.2 Per Patient Performance Assessments

To ascertain if patient performance was significantly changed on a per patient basis by training on only the subset of birth asphyxia patients, the distribution of patient accuracy scores was computed for both cases. For classifiers trained on the entire external dataset, the average patient accuracy scores (Figure 5.3, black triangles) were 0.85, 0.85, and 0.86 for the RF, SVM, and ANN respectively. Accuracy scores were comparable when classifiers were trained on only the subset of birth asphyxia patients (Figure 5.3, red circles): 0.86, 0.85, and 0.85 for the RF, SVM and ANN. Using Wilcoxon signed rank test and a Bonferroni-Dunn correction for multiple comparisons, it was confirmed that the accuracy for each classifier was not significantly different

when classifiers were trained on the entire external dataset or just the asphyxia subset (Figure 5.3, p>0.990, p.0.99, p>0.904 for RF, SVM, and ANN respectively).

There were 343 seizures in the training dataset, and 235 of them were from birth asphyxia patients. Additionally, there were 654.3 total minutes of seizure training data and 457.8 of them were from asphyxia patients.

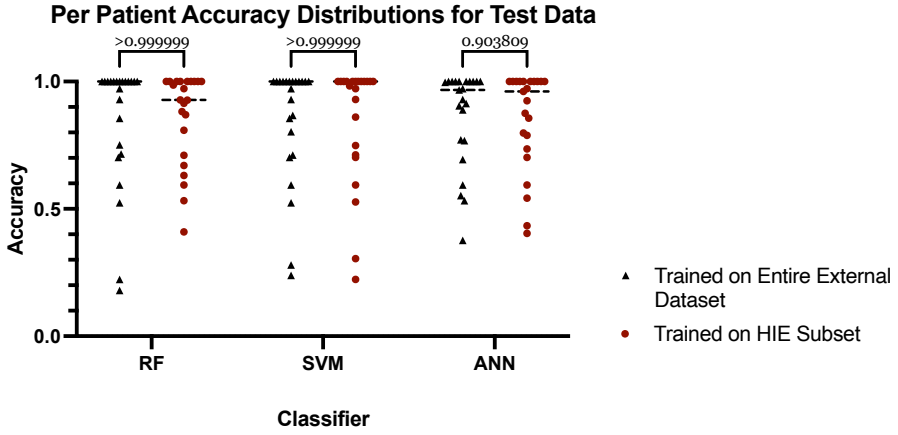

Figure 5.3 Distribution of accuracy scores for 23 HIE patients in the test set. Scores were calculated using the optimal threshold for each classifier computed from each classifier's the patient independent ROC from Figure 5.1. Black triangles represent patient accuracy scores after algorithm training on the external dataset and red circles represent accuracy scores after training on only the HIE subset. P-values of p>0.999, p>0.999 and p>0.904 indicate no significant difference.

5.4.3 Feature Importance for Seizure Classification

To determine the relative importance of features on seizure detection algorithm performance, features were scored using Random Forest Out-of-Bag permuted predictor importance. The top three features for both models trained on the entire dataset and those trained on only on the asphyxia subset were CSA Beta Power, aEEG Lower Margin, and the CSA Alpha power (Figure 5.4). When these features are plotted over time, they are found to increase with

seizure activity (Figure 5.1-B, C, and D). In the subset of models trained only on birth asphyxia patients, CSA beta power decreases in importance and was replaced by the aEEG margin as the most important feature.

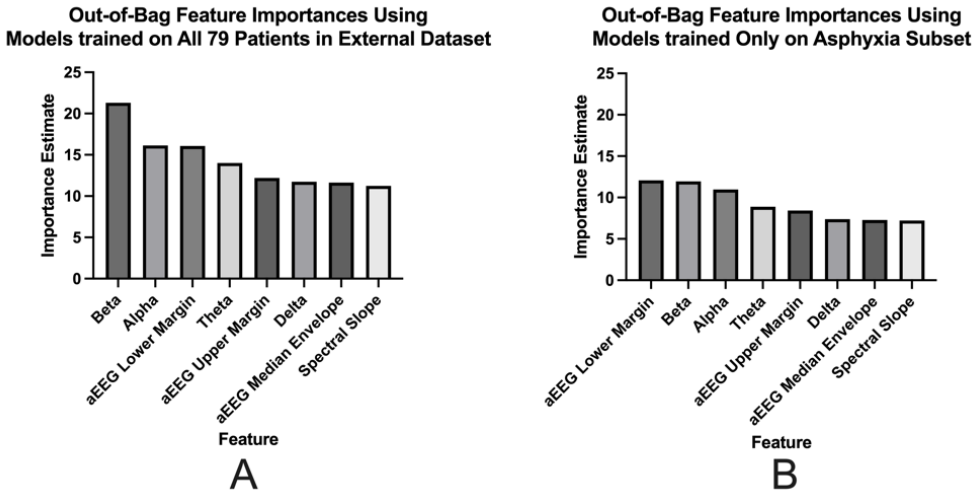

Figure 5.4 Feature importance scores. A. Importance scores when training data consists of all 79 patients in the external dataset. Beta power is the most important. B. Importance scores when training data only consists of the 35 asphyxia patients. The aEEG lower margin is the most important.

5.5 Discussion

This study validates an algorithm used for neonatal seizure detection on an independent cohort of 23 infants with HIE. This algorithm was initially assessed in a previous study using only patients in the external dataset via K-fold cross validation. The patient independent AUC reported in the previous study was 0.8, 0.71, and 0.8 for the RF, SVM, and ANN classifiers respectively. The previous study also showed that the algorithm had significantly lower accuracy across the 35 HIE patients compared to the other 44 patients with various etiologies. That finding motivated further assessment of algorithm performance on an independent HIE cohort.

The aEEG-CSA algorithm can detect seizures with an AUC ranging from 0.76 to 0.79. The study also showed that training models on patients with similar disease etiologies does not appear to decrease model performance, despite reducing the total amount of training data. This consistency in performance may have been because 70% of the seizure recordings in the Helsinki dataset used for training occurred in the subset of patients with birth asphyxia. However, there was also a large decrease in the amount of nonseizure data, since training on only the 35 asphyxia patients represents a 56% decrease in the number of patients for training. A reduction in precision and accuracy might have been expected. This decrease was not seen. Follow-up work with a larger cohort is necessary to ascertain the true benefit of more disease-specific training sets.

Beta Power, aEEG lower margin, and alpha power were the three most important features for seizure detection. The relative importance of beta power decreased while aEEG lower margin and alpha activity remained the same when only the asphyxia subset was used to train (Figure 5.4B). Further studies are necessary to ascertain if beta power importance is due to fast activity such as muscle activity or if they relate purely to nonmovement based seizure activity and if this feature is more pronounced among HIE patients.

5.6 Conclusions

The aEEG-CSA algorithm can detect seizures with reasonable sensitivity and specificity. Features most important for seizure detection are the Beta Power, aEEG lower margin, and alpha power. Beta and alpha power may play a role in classification of neonatal seizure activity specifically in infants with HIE. Training models on a smaller subset of patients but with same disease etiology, HIE due to birth asphyxia, does not decrease algorithm performance.

Chapter 6

DISCUSSION

6.1 Overview

The overall objective of this body of work is to generate clinically translatable algorithms for neonatal hypoxic ischemic encephalopathy using the two most widely used neonatal brain imaging modalities, EEG and MRI. The focus was also placed on using clinically relevant features taken from EEG and MRI because they are part of standard care.

The first objective was to assess which features from EEG and MRI would have the most relevance when assessing outcomes in infants with HIE. To accomplish this objective, a systematic review was conducted. From this review, EEG and MRI features shown to be consistently associated with outcomes were amalgamated. This review also revealed the importance of accurate measurements of seizure burden in assessing outcomes since high seizure burden was consistently associated with poor outcomes in neonates. This finding corroborated the importance of sensitive and specific seizure detection to generate accurate measurements of neonatal seizure burden.

The second objective was to generate a multimodal model for outcome prognostication in neonatal patients with HIE. To accomplish this objective, a cohort of patients with neonatal HIE was amalgamated via a retrospective chart review. This initial cohort, was whittled down to include patients who received hypothermia therapy, were full term, had neonatal encephalopathy specifically due to hypoxic injury no other encephalopathy confounders, received 72-hour therapeutic hypothermia, received EEG monitoring during hypothermia therapy, and finally received MRI within 11 days. This inclusion criteria generated a cohort of 23 patients whose diagnosis reflected the current HIE disease etiology and who received care that reflects the current

standard of care for neonatal HIE. Outcome data from this cohort of patients could then be assessed.

The multimodal algorithm generated results based on outcome information between the 3-to-6-month interval. This interval reflects the general post-neonatal follow-up period where standard neurological tests like fidgety movement are assessed[172]. This period also serves to address the gap of follow-up loss that may skew the effects of standard long-term outcome metrics like the BSID[173,174]. Results showed the combination of insula injury and absence of EEG state change generated a statistically significant, well-fit multimodal model for outcomes. Results also showed that insula injury and basal ganglia injury were independent and significant predictors of outcome in this cohort of infants and generated a potentially better predictive model. Finally, results also showed that white matter injury may play a role in term neonatal outcome and may have an association with seizure activity.

The third objective was to address the other important features associated with outcome in neonatal HIE patients which was seizure burden. The problem of lack of adequate measurements for seizure burden was addressed by generating a seizure detection algorithm using long-term qEEG monitoring tools that are part of standard care for neonatal patients. These tools are aEEG and CSA. The aEEG algorithm was computationally modeled and visual-clinical features were extracted across 10-second windows. This assessment window allows for better seizure resolution compared to traditional aEEG. This same process was applied for extracting CSA features – especially generating a CSA and extracting visual-clinical features from each power spectrum across 10 second windows.

Additionally, an automated artifact removal procedure was also implemented using aEEG and CSA to remove any windows that contained excessively high amplitude artifact and

excessively low amplitude artifacts. Then, these features were coupled with supervised machine learning classifiers and trained and tested on a cohort of neonatal patients with seizures from a variety of disease etiologies. Finally, the detection algorithm was validated using the previous more homogenous cohort of 23 neonatal patients with HIE.

6.2 Generating a Multimodal Detection Algorithm Using EEG and MRI

One of the primary interests of this study was to generate a multimodal detection algorithm. This objective was pursued to extract information from the current prognostic tools used as standard of care – the EEG and MRI. Prognostic assessments currently focus on a single modality when assessing outcomes. The objective was to go beyond this pattern by finding useful features from both EEG and MRI. A major limitation of this study was the sample size of the HIE cohort used. Since only 23 patients fit the inclusion criteria, only two features could be implemented. Follow-up studies with a larger cohort may facilitate the use of additional EEG and MRI features and improve algorithm performance and generalizability.

6.3 Challenges of the Early Prognostic Algorithm and Future Directions

The first major challenge to using the generated prognostic algorithm is the assessment of outcomes at three to six months. This challenge exists because clinicians are making assessments prior to the emergence of important developmental milestones like walking and talking which largely emerge after the first year of life. Thus, an argument might be made that these algorithms may not accurately gauge neurocognitive development. Additionally, the argument can be made that an infant who presents poorly on neurological tests at three to six months may recovery well after that time thus, poor outcome may be overestimated. Finally, since outcomes involve early

clinical tests such as the neurological exam, another argument concerning the inherent ambiguity of early clinical tests can also be made.

In response, the goal of the prognostic algorithm proposed in this thesis is to have a tool that bridges the follow-up gap but also maintains relatively high accuracy. The goal is to capture infants who would normally have had poor outcome even at the 18-to-24-month period but fall into the cohort of infants >40% of which are lost to follow-up after year one[22,23]. Thus, this study looked to assess outcomes of infants, who normally would have had poor outcomes but are not represented among the population of patients surveyed in long-term follow-up tests like the BSID.

To address the issue of the three to six-month period potentially being less prognostically useful, the time range was chosen based on standard follow-up and based on the time range of importance for neurological tests that do have prognostic efficacy such as the fidgety movement assessment. This assessment is also part of standard early follow-up and studies have shown the absence of fidgety movement between three to five months on neurological test is associated with poor developmental outcome and cerebral palsy[172].

To address the issue of ambiguity of early clinical tests, the study sought to not only rely on neurological tests, but also on the presence of unambiguous and potentially prognostically relevant markers like hearing screen failure, on abnormal visual tracking, and on presence of post-neonatal epilepsy via use of anti-epileptic medication, which are not ambiguous.

The prognostic study detailed in chapter three also opens many avenues for future assessments. One possible future direction involves ascertaining whether there is a strong correlation between patients included in the study who showed evidence of poor outcome within three to six months and those who have poor long-term outcomes. This assessment will serve to demonstrate how well the short-come outcome criteria captured patients who would have been

considered developmentally delayed by traditional outcome scoring metrics like the BSID. This assessment may also reveal whether the selected short-term outcome criteria correlate with either cognitive or motor delay or merely composite developmental delay scores. This stratification is particularly important since the current outcome criteria showed strong associations with both basal ganglia and cortical injury locations.

6.4 Future Directions of MRI Assessments for Term Neonatal HIE Pathogenesis

Currently, injury to the basal ganglia has been the primary focus of neonatal HIE pathogenesis. This is because basal ganglia injury pattern has been found to be associated with early markers of dysfunction and worsened neuromotor outcomes after 30 months based on the BSID.[175] This contrasts with neonates with watershed injury patterns. This injury pattern involves injuries of white matter in the watershed zones such as the periventricular white matter. Under more severe hypoxic conditions, these lesions may extend to cortical gray matter areas.[175] These areas would include structures like the insular cortex. Injuries and changes in volume in these areas have also been associated with both long-term intellectual deficits during adolescence and language deficits at above 30 months of age[176,177]. Interestingly, Barkovich et al 1998 found that a basal-ganglia watershed pattern between predicted outcome between 3 to 12 months[115]. This finding is consistent with what was seen in chapter three, where the basal-ganglia injury severity and insula injury combined, where both associated strongly predictive of outcomes. The study listed in chapter three went a step further by assessing injury in a particular cortical location combined with the basal ganglia as associated with poor outcome. A future direction would be to score the severity of insula injury. Another would also be to ascertain significance of the

combination of insula and basal ganglia injury with a larger cohort to see if neonates with insula injury independently of basal ganglia injury had more of an intermediate outcome score.

In addition, the other pattern that varied significantly with outcome was the periventricular white matter injury severity pattern. After further assessments, it was also found that white matter injury severity varied significantly with markers of both neonatal and post-neonatal epilepsy in the dataset of term neonates. Firstly, white matter injury was associated with the number epilepsy medications a patient was on during therapeutic hypothermia. White matter injury was also significantly associated with seizures between 24 and 26 hours of life. Finally, white matter injury was correlated significantly with anti-epileptic medication usage at the three-to-six-month follow-up period. Periventricular white matter injury has been primarily noted in preterm neonates as the primary injury pattern. The reasoning for this lies in the proliferation of $O4^+$ premyelinating oligodendrocytes after 20 weeks gestational.[178] These oligodendrocytes are highly sensitive to hypoxia and oxidative stress[178,179]. This sensitivity leads to selective degeneration of these cells in preterm HIE[179].

The mechanism of white matter injury in term neonates has not been as clearly elucidated. A previous study looked specifically at term neonates and found that gestational age in their cohort correlated with non-cystic white matter injury severity[180]. In the cohort included in the study detailed in chapter three, gestational age did not correlate with white matter injury severity. This discrepancy may have been due to chapter three's study focusing on scoring diffuse white matter injury using the Rutherford Scoring Metric. Another study published by Kim et al 2020 found that diffuse periventricular white matter injuries in a cohort of 13 neonates was seen in neonates with seizures. Despite this finding, Kim et al did not include any neonates with HIE[181]. Mao et al 2016 also showed that white matter hyperintensities are associated with epilepsy – albeit in a cohort of

adults[182]. They found that exposure to enzyme-inducing antiepileptic drugs (EIAEDs) was associated with both more extensive hyperintensities and new epilepsy diagnosis. They mentioned that the underlying mechanism for more extensive white matter hyperintensities in epilepsy may come from increases in the permeability of the blood-brain-barrier seen in epilepsy[183]. Finally, other studies have shown that cases of rotavirus infection have been associated with leukoencephalopathy and subsequent neonatal seizures. Although cursory chart review does not point to a viral basis for patients with severe white matter injury, more detailed chart review is necessary to ascertain seizures and white matter injury are not due to concomitant viral infection.

6.5 Towards a Purely Computational Prognostic Model for Outcome Assessment

The prognostic study showed that a combination of EEG and MRI markers can be used to predict early outcome. The future objective would be to computationally extract significant features from EEG and MRI to generally a purely computational tool. The benefit of such a tool would lie in its ability to be used to bridge the gap for potentially under-resourced centers that may not have consistent access to skilled neurophysiologists for efficient interpretation of EEG and MRI. Additionally, computational assessment may help overcome the challenge of interrater reliability between clinicians. Better consensus on the presence or absence of certain features will allow for more consistent prognostic assessments which may help increase algorithm sensitivity.

Despite these benefits, computational assessments still present many challenges. While features such as interburst interval duration have already been computed, a more challenging endeavor would be computing complex features like EEG state change. A potential answer and future assessment would be to compute more general features like delta power over the course of several hours to assess for sleep wake cycles. Kota et al 2020 found that median delta power during

therapeutic hypothermia was significantly lower between infants with poor outcome vs good outcomes. Poor outcome in this study was classified by death or adverse MRI injury. A follow-up study would be to model state changes by assessing the rate of change in delta power during cooling and evaluate if these varied significantly between infants with good versus poor outcomes. Another assessment may also be to try to capture specific graphoelements in the EEG that are associated with good outcome. Since term Neonatal EEG has graphoelements like frontal sharps and enconches frontales, which present as highly symmetric EEG deflections in the frontal electrodes, these patterns may be extracted using computational assessments like EEG phase locking index or EEG brain symmetry index specially using the delta band.

Additionally, MRI features may also be extracted computationally to promote greater sensitivity and overcome the challenge of interrater reliability seen in visual MRI assessments. The systematic review detailed in chapter two showed that lesion volume can be assessed using pixel density and can vary significantly with poor outcome. Additionally, the ADC within particular brain areas can also serve as a computational feature. Low ADC values reveal restricted diffusion and indicate pathologic processes like cytotoxic edema which are characteristic of encephalopathy due to energy failure[184]. Low ADC values have already been shown in the basal ganglia and thalamus in neonatal HIE. One downside of focus on quantitative DWI assessments in the basal ganglia is the possibility of basal ganglia and thalamic lesion being more prone to pseudonormalization in comparison to cortical areas[185]. Thus, future studies may focus on assessing ADC values in specific cortical areas such as the insular cortex that might have the potential for prognostic efficacy that will not be hampered by pseudonormalization.

6.6 Neonatal Seizure Detection: Reducing the False Positive Rate

A major challenge in generating neonatal seizure detection algorithms is reducing the false positive rate. This assessment is particularly important for preventing alarm fatigue. Alarm fatigue is potentially detrimental for patients because a high false alarm rate may lead to clinicians becoming so desensitized that they may ignore even true instances of a seizure. It was previously noted that increases in periictal power were associated with seizures. While these changes often occurred during the seizure, sometimes they occurred directly before or after the seizure. These changes may increase the false positive rate by causing the algorithm to flag epochs directly before or after the seizure label. A future approach would be to implement another parameter like seizure onset and offset delay tolerance. This parameter may help to improve the false positive rate.

6.7 aEEG Feature Engineering based on Feature Importance

Alpha power, beta power, and the aEEG lower margin activity were shown to be the most important features for detection. A future approach may be to modify the underlying aEEG lower margin algorithm to extract a potentially more sensitive lower margin by using an asymmetric band-ass filter that spans the alpha and beta band instead of the 2-15Hz standard bandpass filter. This modification may allow the algorithm to merge importance of trends in the alpha power and trends in the lower margin to produce potentially better lower margin feature. This metric could be tested to prove non-inferiority to the standard aEEG algorithm.

6.8 Personalized Seizure Detection

Fully patient-dependent seizure detection algorithms have been shown to perform better than patient-independent algorithms[186]. This performance difference may be due to differences in

the seizure probability threshold per patient. Additionally, other factors such as disease etiology, gestational age, and hours of life may play a role in generating seizure propensity. For these reasons, a future approach would be to improve algorithm performance using both patient independent and patient dependent seizure labels. A previous study addressed this challenge by combining a patient-independent, SVM-based model and a patient-dependent generative model[186]. They used the SVM to label new unseen data; these SVM labels were then used for real-time adaptation of a generative gaussian mixture model. The benefit of an ensemble approach such as this would be to allow for continuous adaptation of new patient specific information. Implementing a generative model in combination with a discriminative classifier like SVM may help to calibrate individual patient seizure probabilities to represent their evolving propensity for seizure activity.

Finally, the challenge of poorer performance in HIE patients must also be addressed. The data presented in this thesis suggests that training HIE patients with those with similar disease etiology does not decrease algorithm performance. More work needs to be done on assessing if there are unique time and frequency feature combinations that can help distinguish HIE patients from non-hie patients from the EEG. Better understanding of these patterns may help demystify if training based on disease etiology is indeed superior.

6.9 Conclusions

This body of work focused on generating algorithms to address the most common challenges faced in neonatal HIE, which are sensitive and specific outcome prognostication and accurate seizure detection. In chapter two, a systematic literature search was done to amalgamate early electrographic and imaging features that were associated with differential outcome in neonatal HIE. For the EEG, these features were the EEG background pattern, interburst interval

amplitudes, sleep wake cycle presence, and seizure burden during therapeutic hypothermia. For the MRI, these features were the MRI injury location, severity, and lack of week two pseudonormalization. These features were then implemented in chapter three to generate a multimodal detection algorithm using a cohort of HIE patients. It was found that a combination of EEG state change and cortical insula injury score could be predict early outcome. It was also found that periventricular white matter injury is also associated with poor outcomes in term infants with HIE. Finally, in chapters four and five, the problem of accurately capturing seizures was addressed. In chapter it was found that an algorithm that combined clinical features from the amplitude-integrated EEG and compressed spectral array could be used to detect neonatal seizures in a cohort of 79 patients with various disease etiologies. It was also discovered that HIE patients performed more poorly. In chapter four, algorithm performance was then validated on the inhouse cohort of neonates with HIE, and it was found that performance on this cohort was the same on this holdout cohort when patients were trained on only the 35 birth asphyxia patients in comparison to when they were trained on all 79 patients.

REFERENCES

1. Kurinczuk, J. J., White-Koning, M. & Badawi, N. Epidemiology of neonatal encephalopathy and hypoxic–ischaemic encephalopathy. *Early Hum. Dev.* **86**, 329–338 (2010).

2. Finder, M. *et al*. Two-Year Neurodevelopmental Outcomes After Mild Hypoxic Ischemic Encephalopathy in the Era of Therapeutic Hypothermia. *JAMA Pediatr.* **174**, 48–55 (2020).

3. Allen, K. A. & Brandon, D. H. Hypoxic Ischemic Encephalopathy: Pathophysiology and Experimental Treatments. *Newborn Infant Nurs. Rev. NAINR* **11**, 125–133 (2011).

4. Fatemi, A., Wilson, M. A. & Johnston, M. V. Hypoxic Ischemic Encephalopathy in the Term Infant. *Clin. Perinatol.* **36**, 835–vii (2009).

5. Arnesen, T. & Nord, E. The value of DALY life: problems with ethics and validity of disability adjusted life years. *BMJ* **319**, 1423–1425 (1999).

6. Murray, C. J. & Lopez, A. D. Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *The Lancet* **349**, 1436–1442 (1997).

7. Lee, A. C. *et al*. Intrapartum-related neonatal encephalopathy incidence and impairment at regional and global levels for 2010 with trends from 1990. *Pediatr. Res.* **74**, 50–72 (2013).

8. Kleuskens, D. G. *et al*. Pathophysiology of Cerebral Hyperperfusion in Term Neonates With Hypoxic-Ischemic Encephalopathy: A Systematic Review for Future Research. *Front. Pediatr.* **9**, (2021).

9. Nair, J. & Kumar, V. H. S. Current and Emerging Therapies in the Management of Hypoxic Ischemic Encephalopathy in Neonates. *Children* **5**, 99 (2018).

10. Kharoshankaya, L. *et al*. Seizure burden and neurodevelopmental outcome in neonates with hypoxic–ischemic encephalopathy. *Dev. Med. Child Neurol.* **58**, 1242–1248 (2016).

11. Kahle, K. T. & Staley, K. J. Neonatal Seizures and Neuronal Transmembrane Ion Transport. in *Jasper's Basic Mechanisms of the Epilepsies* (eds. Noebels, J. L., Avoli, M., Rogawski, M. A., Olsen, R. W. & Delgado-Escueta, A. V.) (National Center for Biotechnology Information (US), 2012).

12. Cotten, C. M. & Shankaran, S. Hypothermia for hypoxic–ischemic encephalopathy. *Expert Rev. Obstet. Gynecol.* **5**, 227–239 (2010).

13. Azzopardi, D. V. *et al*. Moderate Hypothermia to Treat Perinatal Asphyxial Encephalopathy. *N. Engl. J. Med.* **361**, 1349–1358 (2009).

14. Jacobs, S. E. *et al*. Cooling for newborns with hypoxic ischaemic encephalopathy. *Cochrane Database Syst. Rev.* **2013**, CD003311 (2013).

15. Herrera, T. I. *et al*. Outcomes of preterm infants treated with hypothermia for hypoxic-ischemic encephalopathy. *Early Hum. Dev.* **125**, 1–7 (2018).

16. Rao, R. *et al*. Safety and Short-Term Outcomes of Therapeutic Hypothermia in Preterm Neonates 34-35 Weeks Gestational Age with Hypoxic-Ischemic Encephalopathy. *J. Pediatr.* **183**, 37–42 (2017).

17. Lademann, H., Abshagen, K., Janning, A., Däbritz, J. & Olbertz, D. Long-Term Outcome after Asphyxia and Therapeutic Hypothermia in Late Preterm Infants: A Pilot Study. *Healthcare* **9**, 994 (2021).

18. NICHD Neonatal Research Network. *A Randomized Trial of Targeted Temperature Management With Whole Body Hypothermia For Moderate And Severe Hypoxic-Ischemic Encephalopathy In Premature Infants 33-35 Weeks Gestational Age*. https://clinicaltrials.gov/ct2/show/NCT01793129 (2022).

19. Robertson, G. J. Bayley Scales of Infant and Toddler Development. in *The Corsini Encyclopedia of Psychology* 1–2 (John Wiley & Sons, Ltd, 2010). doi:10.1002/9780470479216.corpsy0111.

20. Jary, S., Whitelaw, A., Walløe, L. & Thoresen, M. Comparison of Bayley-2 and Bayley-3 scores at 18 months in term infants following neonatal encephalopathy and therapeutic hypothermia. *Dev. Med. Child Neurol.* **55**, 1053–1059 (2013).

21. Anderson, P. J. *et al*. Underestimation of developmental delay by the new Bayley-III Scale. *Arch. Pediatr. Adolesc. Med.* **164**, 352–356 (2010).

22. Swearingen, C., Simpson, P., Cabacungan, E. & Cohen, S. Social disparities negatively impact neonatal follow-up clinic attendance of premature infants discharged from the neonatal intensive care unit. *J. Perinatol.* **40**, 790–797 (2020).

23. L. Orton, J., McGinley, J. L., Fox, L. M. & Spittle, A. J. Challenges of neurodevelopmental follow-up for extremely preterm infants at two years. *Early Hum. Dev.* **91**, 689–694 (2015).

24. Kirschstein, T. & Köhling, R. What is the source of the EEG? *Clin. EEG Neurosci.* **40**, 146–149 (2009).

25. Odabaee, M. *et al*. Neonatal EEG at scalp is focal and implies high skull conductivity in realistic neonatal head models. *NeuroImage* **96**, 73–80 (2014).

26. Hansman, C. F. Growth of Interorbital Distance and Skull Thickness as Observed in Roentgenographic Measurements. *Radiology* **86**, 87–96 (1966).

27. Shellhaas, R. A. *et al*. The American Clinical Neurophysiology Society's Guideline on Continuous Electroencephalography Monitoring in Neonates. *J. Clin. Neurophysiol.* **28**, 611–617 (2011).

28. Takanashi, J. *et al*. Brain MR Imaging in Neonatal Hyperammonemic Encephalopathy Resulting from Proximal Urea Cycle Disorders. *AJNR Am. J. Neuroradiol.* **24**, 1184–1187 (2003).

29. Wusthoff, C. J., Shellhaas, R. A. & Clancy, R. R. Limitations of single-channel EEG on the forehead for neonatal seizure detection. *J. Perinatol.* **29**, 237–242 (2009).

30. Tsuchida, T. N. *et al*. American Clinical Neurophysiology Society Standardized EEG Terminology and Categorization for the Description of Continuous EEG Monitoring in Neonates: Report of the American Clinical Neurophysiology Society Critical Care Monitoring Committee. *J. Clin. Neurophysiol.* **30**, (2013).

31. Dall, T. M. *et al*. Supply and demand analysis of the current and future US neurology workforce. *Neurology* **81**, 470–478 (2013).

32. Roychoudhury, S. *et al*. Implementation of Neonatal Neurocritical Care Program Improved Short-Term Outcomes in Neonates With Moderate-to-Severe Hypoxic Ischemic Encephalopathy. *Pediatr. Neurol.* **101**, 64–70 (2019).

33. Dilena, R. *et al*. Consensus protocol for EEG and amplitude-integrated EEG assessment and monitoring in neonates. *Clin. Neurophysiol.* **132**, 886–903 (2021).

34. Shellhaas, R. A., Soaita, A. I. & Clancy, R. R. Sensitivity of amplitude-integrated electroencephalography for neonatal seizure detection. *Pediatrics* **120**, 770–777 (2007).

35. Maynard, D., Prior, P. F. & Scott, D. F. Device for continuous monitoring of cerebral activity in resuscitated patients. *Br. Med. J.* **4**, 545–546 (1969).

36. Vries, L. S. de & Hellström-Westas, L. Role of cerebral function monitoring in the newborn. *Arch. Dis. Child. - Fetal Neonatal Ed.* **90**, F201-FF207 (2005).

37. Zhang, D. & Ding, H. Calculation of compact amplitude-integrated EEG tracing and upper and lower margins using raw EEG data. *Health (N. Y.)* **5**, 885–891 (2013).

38. Glass, H. C., Wusthoff, C. J. & Shellhaas, R. A. Amplitude Integrated EEG: The Child Neurologist's Perspective. *J. Child Neurol.* **28**, 1342–1350 (2013).

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

40. Csekő, A. J. *et al*. Accuracy of amplitude-integrated electroencephalography in the prediction of neurodevelopmental outcome in asphyxiated infants receiving hypothermia treatment. *Acta Paediatr. Oslo Nor. 1992* **102**, 707–711 (2013).

41. Sewell, E. K. *et al*. Evolution of Amplitude-Integrated Electroencephalogram as a Predictor of Outcome in Term Encephalopathic Neonates Receiving Therapeutic Hypothermia. *Am. J. Perinatol.* **35**, 277–285 (2018).

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

43. Shah, D. K. *et al*. Accuracy of bedside electroencephalographic monitoring in comparison with simultaneous continuous conventional electroencephalography for seizure detection in term infants. *Pediatrics* **121**, 1146–1154 (2008).

44. Mastrangelo, M. *et al*. Acute neonatal encephalopathy and seizures recurrence: a combined aEEG/EEG study. *Seizure* **22**, 703–707 (2013).

45. Rakshasbhuvankar, A., Paul, S., Nagarajan, L., Ghosh, S. & Rao, S. Amplitude-integrated EEG for detection of neonatal seizures: a systematic review. *Seizure* **33**, 90–98 (2015).

46. Kadivar, M., Moghadam, E. M., Shervin Badv, R., Sangsari, R. & Saeedy, M. A Comparison Of Conventional Electroencephalography With Amplitude-Integrated EEG In Detection Of Neonatal Seizures. *Med. Devices Auckl. NZ* **12**, 489–496 (2019).

47. Werther, T. *et al*. Are All Amplitude-Integrated Electroencephalogram Systems Equal? *Neonatology* **112**, 394–401 (2017).

48. Werther, T. *et al*. Are All Amplitude-Integrated Electroencephalogram Systems Equal? *Neonatology* **112**, 394–401 (2017).

49. Merchant, N. & Azzopardi, D. Early predictors of outcome in infants treated with hypothermia for hypoxic–ischaemic encephalopathy. *Dev. Med. Child Neurol.* **57**, 8–16 (2015).

50. del Río, R. *et al*. Amplitude Integrated Electroencephalogram as a Prognostic Tool in Neonates with Hypoxic-Ischemic Encephalopathy: A Systematic Review. *PLoS ONE* **11**, (2016).

51. Yuan, X. *et al*. Prognostic value of amplitude-integrated EEG in neonates with high risk of neurological sequelae. *Ann. Clin. Transl. Neurol.* **7**, 210–218 (2020).

52. van Drongelen, W. Chapter 7 - 1-D and 2-D Fourier Transform Applications. in *Signal Processing for Neuroscientists (Second Edition)* (ed. van Drongelen, W.) 119–152 (Academic Press, 2018). doi:10.1016/B978-0-12-810482-8.00007-2.

53. *Quantitative EEG analysis methods and clinical applications*. (Artech House, 2009).

54. Williamson, C. A., Wahlster, S., Shafi, M. M. & Westover, M. B. Sensitivity of compressed spectral arrays for detecting seizures in acutely ill adults. *Neurocrit. Care* **20**, 32–39 (2014).

55. Gotman, J., Flanagan, D., Zhang, J. & Rosenblatt, B. Automatic seizure detection in the newborn: methods and initial evaluation. *Electroencephalogr. Clin. Neurophysiol.* **103**, 356–362 (1997).

56. Kang, S. K. & Kadam, S. D. Neonatal Seizures: Impact on Neurodevelopmental Outcomes. *Front. Pediatr.* **3**, (2015).

57. Pavel, A. M. *et al.* Neonatal Seizure Management: Is the Timing of Treatment Critical? *J. Pediatr.* **243**, 61-68.e2 (2022).

58. Temko, A., Thomas, E., Marnane, W., Lightbody, G. & Boylan, G. EEG-based neonatal seizure detection with Support Vector Machines. *Clin. Neurophysiol.* **122**, 464–473 (2011).

59. Pavel, A. M. *et al.* A machine-learning algorithm for neonatal seizure recognition: a multicentre, randomised, controlled trial. *Lancet Child Adolesc. Health* **4**, 740–749 (2020).

60. Lommen, C. *et al.* An algorithm for the automatic detection of seizures in neonatal amplitude-integrated EEG. *Acta Paediatr.* **96**, 674–680 (2007).

61. Lawrence, R., Mathur, A., Nguyen The Tich, S., Zempel, J. & Inder, T. A pilot study of continuous limited-channel aEEG in term infants with encephalopathy. *J. Pediatr.* **154**, 835-841.e1 (2009).

62. Rakshasbhuvankar, A., Rao, S., Palumbo, L., Ghosh, S. & Nagarajan, L. Amplitude Integrated Electroencephalography Compared With Conventional Video EEG for Neonatal Seizure Detection: A Diagnostic Accuracy Study. *J. Child Neurol.* **32**, 815–822 (2017).

63. Alger, J. R. Magnetic Resonance Imaging [MRI]. in *Encyclopedia of the Human Brain* (ed. Ramachandran, V. S.) 729–744 (Academic Press, 2002). doi:10.1016/B0-12-227210-2/00222-3.

64. Grover, V. P. B. *et al.* Magnetic Resonance Imaging: Principles and Techniques: Lessons for Clinicians. *J. Clin. Exp. Hepatol.* **5**, 246–255 (2015).

65. Jisa, K. A., Clarey, D. D. & Peeples, E. S. Magnetic Resonance Imaging Findings of Term and Preterm Hypoxic-Ischemic Encephalopathy: A Review of Relevant Animal Models and Correlation to Human Imaging. *Open Neuroimaging J.* **12**, 55–65 (2018).

66. Sevick, R. J. *et al.* Cytotoxic brain edema: assessment with diffusion-weighted MR imaging. *Radiology* **185**, 687–690 (1992).

67. de Vries, L. S. & Groenendaal, F. Patterns of neonatal hypoxic–ischaemic brain injury. *Neuroradiology* **52**, 555–566 (2010).

68. Varghese, B. *et al*. Magnetic resonance imaging spectrum of perinatal hypoxic-ischemic brain injury. *Indian J. Radiol. Imaging* **26**, 316–327 (2016).

69. Cabaj, A., Bekiesińska-Figatowska, M. & Mądzik, J. MRI patterns of hypoxic-ischemic brain injury in preterm and full term infants – classical and less common MR findings. *Pol. J. Radiol.* **77**, 71–76 (2012).

70. Barkovich, A. J. & Sargent, S. K. Profound asphyxia in the premature infant: imaging findings. *AJNR Am. J. Neuroradiol.* **16**, 1837–1846 (1995).

71. Hasegawa, M. *et al*. Development of myelination in the human fetal and infant cerebrum: a myelin basic protein immunohistochemical study. *Brain Dev.* **14**, 1–6 (1992).

72. Talos, D. M. *et al*. Developmental regulation of α-amino-3-hydroxy-5-methyl-4-isoxazole-propionic acid receptor subunit expression in forebrain and relationship to regional susceptibility to hypoxic/ischemic injury. I. Rodent cerebral white matter and cortex. *J. Comp. Neurol.* **497**, 42–60 (2006).

73. McQuillen, P. S. & Ferriero, D. M. Selective vulnerability in the developing central nervous system. *Pediatr. Neurol.* **30**, 227–235 (2004).

74. Chen, W., Wang, Y., Cao, G., Chen, G. & Gu, Q. A random forest model based classification scheme for neonatal amplitude-integrated EEG. *Biomed. Eng. OnLine* **13**, S4 (2014).

75. Wang, J., Ju, R., Chen, Y., Liu, G. & Yi, Z. Automated diagnosis of neonatal encephalopathy on aEEG using deep neural networks. *Neurocomputing* **398**, 95–107 (2020).

76. O'Shea, A., Lightbody, G., Boylan, G. & Temko, A. Neonatal Seizure Detection using Convolutional Neural Networks. *ArXiv170905849 Cs Stat* (2017).

77. Temko, A. & Lightbody, G. Detecting Neonatal Seizures With Computer Algorithms. *J. Clin. Neurophysiol. Off. Publ. Am. Electroencephalogr. Soc.* **33**, 394–402 (2016).

78. Dereymaeker, A. *et al*. Automated EEG background analysis to identify neonates with hypoxic-ischemic encephalopathy treated with hypothermia at risk for adverse outcome: A pilot study. *Pediatr. Neonatol.* **60**, 50–58 (2019).

79. Abbasi. Applications of advanced signal processing and machine learning in the neonatal hypoxic-ischemic electroencephalogram. https://www.nrronline.org/article.asp?issn=1673-5374;year=2020;volume=15;issue=2;spage=222;epage=231;aulast=Abbasi.

80. Kota, S. *et al*. Prognostic Value of Continuous Electroencephalogram Delta Power in Neonates with Hypoxic Ischemic Encephalopathy. *J. Child Neurol*. **35**, 517–525 (2020).

81. Ansari, A. H. *et al*. Neonatal Seizure Detection Using Deep Convolutional Neural Networks. *Int. J. Neural Syst*. **29**, 1850011 (2019).

82. Tanveer, M. A., Khan, M. J., Sajid, H. & Naseer, N. Convolutional neural networks ensemble model for neonatal seizure detection. *J. Neurosci. Methods* **358**, 109197 (2021).

83. Tonekaboni, S., Joshi, S., McCradden, M. D. & Goldenberg, A. What Clinicians Want: Contextualizing Explainable Machine Learning for Clinical End Use. in *Machine Learning for Healthcare Conference* 359–380 (PMLR, 2019).

84. Gavrilov, A., Jordache, A., Vasdani, M. & Deng, J. Preventing Model Overfitting and Underfitting in Convolutional Neural Networks. *Int. J. Softw. Sci. Comput. Intell*. **10**, 19–28 (2018).

85. Iwata, O. *et al*. "Therapeutic time window" duration decreases with increasing severity of cerebral hypoxia–ischaemia under normothermia and delayed hypothermia in newborn piglets. *Brain Res*. **1154**, 173–180 (2007).

86. Ching, S., Purdon, P. L., Vijayan, S., Kopell, N. J. & Brown, E. N. A neurophysiological–metabolic model for burst suppression. *Proc. Natl. Acad. Sci*. **109**, 3095–3100 (2012).

87. Page, M. J. *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* n71 (2021) doi:10.1136/bmj.n71.

88. Huang, X., Lin, J. & Demner-Fushman, D. Evaluation of PICO as a Knowledge Representation for Clinical Questions. *AMIA. Annu. Symp. Proc*. **2006**, 359–363 (2006).

89. Covidence - Better systematic review management. *Covidence* https://www.covidence.org/.

90. Balasundaram, P. & Avulakunta, I. D. Bayley Scales Of Infant and Toddler Development. in *StatPearls* (StatPearls Publishing, 2022).

91. Barnett, A. L. *et al*. Can the Griffiths scales predict neuromotor and perceptual-motor impairment in term infants with neonatal encephalopathy? *Arch. Dis. Child*. **89**, 637–643 (2004).

92. Palisano, R. *et al*. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Dev. Med. Child Neurol*. **39**, 214–223 (1997).

93. Trivedi, S. B. *et al*. A validated clinical MRI injury scoring system in neonatal hypoxic-ischemic encephalopathy. *Pediatr. Radiol*. **47**, 1491–1499 (2017).

94. Rusli, E. R. M. *et al*. Neonatal hypoxic encephalopathy: Correlation between post-cooling brain MRI findings and 2 years neurodevelopmental outcome. *Indian J. Radiol. Imaging* **29**, 350–355 (2019).

95. Bos, A. F. Bayley-II or Bayley-III: what do the scores tell us? *Dev. Med. Child Neurol.* **55**, 978–979 (2013).

96. Dereymaeker, A. *et al*. Automated EEG background analysis to identify neonates with hypoxic-ischemic encephalopathy treated with hypothermia at risk for adverse outcome: A pilot study. *Pediatr. Neonatol.* **60**, 50–58 (2019).

97. Fitzgerald, M. P., Massey, S. L., Fung, F. W., Kessler, S. K. & Abend, N. S. High electroencephalographic seizure exposure is associated with unfavorable outcomes in neonates with hypoxic-ischemic encephalopathy. *Seizure* **61**, 221–226 (2018).

98. Koskela, T. *et al*. Prognostic value of neonatal EEG following therapeutic hypothermia in survivors of hypoxic-ischemic encephalopathy. *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.* **132**, 2091–2100 (2021).

99. Leroy-Terquem, E. *et al*. Abnormal Interhemispheric Synchrony in Neonatal Hypoxic-Ischemic Encephalopathy: A Retrospective Pilot Study. *Neonatology* **112**, 359–364 (2017).

100. Takenouchi, T. *et al*. Delayed onset of sleep-wake cycling with favorable outcome in hypothermic-treated neonates with encephalopathy. *J. Pediatr.* **159**, 232–237 (2011).

101. Chang, P. D., Chow, D. S., Alber, A., Lin, Y.-K. & Youn, Y. A. Predictive Values of Location and Volumetric MRI Injury Patterns for Neurodevelopmental Outcomes in Hypoxic-Ischemic Encephalopathy Neonates. *Brain Sci.* **10**, E991 (2020).

102. Chintalapati, K., Miao, H., Mathur, A., Neil, J. & Aravamuthan, B. R. Objective and Clinically Feasible Analysis of Diffusion MRI Data can Help Predict Dystonia After Neonatal Brain Injury. *Pediatr. Neurol.* **118**, 6–11 (2021).

103. Hayakawa, K. *et al*. Diffusion pseudonormalization and clinical outcome in term neonates with hypoxic–ischemic encephalopathy. *Pediatr. Radiol.* **48**, 865–874 (2018).

104. Jung, D. E., Ritacco, D. G., Nordli, D. R., Koh, S. & Venkatesan, C. Early Anatomical Injury Patterns Predict Epilepsy in Head Cooled Neonates with Hypoxic Ischemic Encephalopathy. *Pediatr. Neurol.* **53**, 135–140 (2015).

105. Lakatos, A. *et al*. Neurodevelopmental effect of intracranial hemorrhage observed in hypoxic ischemic brain injury in hypothermia-treated asphyxiated neonates - an MRI study. *BMC Pediatr.* **19**, 430 (2019).

106. Mastrangelo, M. *et al*. Early Post-cooling Brain Magnetic Resonance for the Prediction of Neurodevelopmental Outcome in Newborns with Hypoxic-Ischemic Encephalopathy. *J. Pediatr. Neurosci*. **14**, 191–202 (2019).

107. Takenouchi, T., Heier, L. A., Engel, M. & Perlman, J. M. Restricted diffusion in the corpus callosum in hypoxic-ischemic encephalopathy. *Pediatr. Neurol*. **43**, 190–196 (2010).

108. Basti, C. *et al*. Seizure burden and neurodevelopmental outcome in newborns with hypoxic-ischemic encephalopathy treated with therapeutic hypothermia: A single center observational study. *Seizure* **83**, 154–159 (2020).

109. Lin, Y.-K., Hwang-Bo, S., Seo, Y.-M. & Youn, Y.-A. Clinical seizures and unfavorable brain MRI patterns in neonates with hypoxic ischemic encephalopathy. *Medicine (Baltimore)* **100**, e25118 (2021).

110. Peeples, E. S. *et al*. Predictive Models of Neurodevelopmental Outcomes After Neonatal Hypoxic-Ischemic Encephalopathy. *Pediatrics* **147**, e2020022962 (2021).

111. Weeke, L. C. *et al*. Role of EEG background activity, seizure burden and MRI in predicting neurodevelopmental outcome in full-term infants with hypoxic-ischaemic encephalopathy in the era of therapeutic hypothermia. *Eur. J. Paediatr. Neurol. EJPN Off. J. Eur. Paediatr. Neurol. Soc*. **20**, 855–864 (2016).

112. Louis, E. K. S. *et al*. *The Developmental EEG: Premature, Neonatal, Infant, and Children. Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants [Internet]* (American Epilepsy Society, 2016).

113. Shankaran, S. *et al*. Brain injury following trial of hypothermia for neonatal hypoxic–ischaemic encephalopathy. *Arch. Dis. Child. - Fetal Neonatal Ed*. **97**, F398–F404 (2012).

114. Bednarek, N. *et al*. Impact of therapeutic hypothermia on MRI diffusion changes in neonatal encephalopathy. *Neurology* **78**, 1420–1427 (2012).

115. Barkovich, A. J. *et al*. Prediction of neuromotor outcome in perinatal asphyxia: evaluation of MR scoring systems. *AJNR Am. J. Neuroradiol*. **19**, 143–149 (1998).

116. Shankaran, S. *et al*. Whole-Body Hypothermia for Neonates with Hypoxic–Ischemic Encephalopathy. *N. Engl. J. Med*. **353**, 1574–1584 (2005).

117. Rutherford, M., Pennock, J., Schwieso, J., Cowan, F. & Dubowitz, L. Hypoxic-ischaemic encephalopathy: early and late magnetic resonance imaging findings in relation to outcome. *Arch. Dis. Child. - Fetal Neonatal Ed*. **75**, F145–F151 (1996).

118. Parmentier, C. E. J., de Vries, L. S. & Groenendaal, F. Magnetic Resonance Imaging in (Near-)Term Infants with Hypoxic-Ischemic Encephalopathy. *Diagnostics* **12**, 645 (2022).

119. Corbo, E. T. *et al*. The effect of whole-body cooling on brain metabolism following perinatal hypoxic–ischemic injury. *Pediatr. Res*. **71**, 85–92 (2012).

120. Massaro, A. N. *et al*. Brain Perfusion in Encephalopathic Newborns after Therapeutic Hypothermia. *AJNR Am. J. Neuroradiol*. **34**, 1649–1655 (2013).

121. Singh, E. *et al*. Improving access to magnetic resonance imaging for the newborn. *J. Neonatal Nurs*. **29**, 199–202 (2023).

122. El-Ayouty, M. *et al*. Relationship between electroencephalography and magnetic resonance imaging findings after hypoxic-ischemic encephalopathy at term. *Am. J. Perinatol*. **24**, 467–473 (2007).

123. Jain, S. V. *et al*. Prediction of Neonatal Seizures in Hypoxic-Ischemic Encephalopathy Using Electroencephalograph Power Analyses. *Pediatr. Neurol*. **67**, 64-70.e2 (2017).

124. Craig, A. K., Gerwin, R., Bainter, J., Evans, S. & James, C. Exploring parent expectations of neonatal therapeutic hypothermia. *J. Perinatol. Off. J. Calif. Perinat. Assoc*. **38**, 857–864 (2018).

125. Volpe, J. J. *Neurology of the Newborn E-Book*. (Elsevier Health Sciences, 2008).

126. Giampietri, M. & Biver, P. Recovery of aEEG Patterns at 24 Hours of Hypothermia Predicts Good Neurodevelopmental Outcome. *J. Neonatal Biol*. **5**, (2016).

127. De Wispelaere, L. ATT. *et al*. Electroencephalography and brain magnetic resonance imaging in asphyxia comparing cooled and non-cooled infants. *Eur. J. Paediatr. Neurol*. **23**, 181–190 (2019).

128. Lemyre, B. & Chau, V. Hypothermia for newborns with hypoxic-ischemic encephalopathy. *Paediatr. Child Health* **23**, 285–291 (2018).

129. Rutherford, M. *et al*. Assessment of brain tissue injury after moderate hypothermia in neonates with hypoxic–ischaemic encephalopathy: a nested substudy of a randomised controlled trial. *Lancet Neurol*. **9**, 39–45 (2010).

130. Hemmert, G. A. J., Schons, L. M., Wieseke, J. & Schimmelpfennig, H. Log-likelihood-based Pseudo-R2 in Logistic Regression: Deriving Sample-sensitive Benchmarks. *Sociol. Methods Res*. **47**, 507–531 (2018).

131. Tsuchida, T. N. *et al*. American clinical neurophysiology society standardized EEG terminology and categorization for the description of continuous EEG monitoring in neonates: report of the American Clinical Neurophysiology Society critical care monitoring committee. *J. Clin. Neurophysiol. Off. Publ. Am. Electroencephalogr. Soc*. **30**, 161–173 (2013).

132. Robertson, C. M. & Perlman, M. Follow-up of the term infant after hypoxic-ischemic encephalopathy. *Paediatr. Child Health* **11**, 278–282 (2006).

133. Acun, C. *et al*. Trends of neonatal hypoxic-ischemic encephalopathy prevalence and associated risk factors in the United States, 2010 to 2018. *Am. J. Obstet. Gynecol.* **227**, 751.e1-751.e10 (2022).

134. Pisani, F. *et al*. Neonatal seizures and postneonatal epilepsy: a 7-y follow-up study. *Pediatr. Res.* **72**, 186–193 (2012).

135. Cerebral Palsy. *National Institute of Neurological Disorders and Stroke* https://www.ninds.nih.gov/health-information/disorders/cerebral-palsy.

136. Mercuri, E. *et al*. Visual function in term infants with hypoxic-ischaemic insults: correlation with neurodevelopment at 2 years of age. *Arch. Dis. Child. Fetal Neonatal Ed.* **80**, F99–F104 (1999).

137. Chen, D.-Y., Lee, I.-C., Wang, X.-A. & Wong, S.-H. Early Biomarkers and Hearing Impairments in Patients with Neonatal Hypoxic–Ischemic Encephalopathy. *Diagnostics* **11**, 2056 (2021).

138. Allen, K. A. & Brandon, D. H. Hypoxic Ischemic Encephalopathy: Pathophysiology and Experimental Treatments. *Newborn Infant Nurs. Rev. NAINR* **11**, 125–133 (2011).

139. Deshpande, P., Jain, A. & McNamara, P. J. Effect of Phenobarbitone on Amplitude-Integrated Electroencephalography in Neonates with Hypoxic-Ischemic Encephalopathy during Hypothermia. *Neonatology* **117**, 721–728 (2020).

140. Padiyar, S., Nusairat, L., Kadri, A., Abu-Shaweesh, J. & Aly, H. Neonatal seizures in the U.S. National Inpatient Population: Prevalence and outcomes. *Pediatr. Neonatol.* **61**, 300–305 (2020).

141. Bassan, H. *et al*. Neonatal Seizures: Dilemmas in Workup and Management. *Pediatr. Neurol.* **38**, 415–421 (2008).

142. Offringa, M. & Kalish, B. T. Subclinical Electrographic Seizures in the Newborn—Is More Treatment Better? *JAMA Netw. Open* **4**, e2140677 (2021).

143. Werner, R. M. & Polsky, D. Comparing the supply of pediatric subspecialists and child neurologists. *J. Pediatr.* **146**, 20–25 (2005).

144. Zhang, D. *et al*. The Prognostic Value of Amplitude-Integrated EEG in Full-Term Neonates with Seizures. *PLoS ONE* **8**, (2013).

145. Rakshasbhuvankar, A., Paul, S., Nagarajan, L., Ghosh, S. & Rao, S. Amplitude-integrated EEG for detection of neonatal seizures: a systematic review. *Seizure* **33**, 90–98 (2015).

146. Pensirikul, A. D. *et al*. Density spectral array for seizure identification in critically ill children. *J. Clin. Neurophysiol. Off. Publ. Am. Electroencephalogr. Soc.* **30**, 371–375 (2013).

147. Amorim, E. *et al*. Performance of spectrogram-based seizure identification of adult EEGs by critical care nurses and neurophysiologists. *J. Clin. Neurophysiol. Off. Publ. Am. Electroencephalogr. Soc.* **34**, 359–364 (2017).

148. Sharma, S., Nunes, M. & Alkhachroum, A. Adult Critical Care Electroencephalography Monitoring for Seizures: A Narrative Review. *Front. Neurol.* **13**, (2022).

149. Stevenson, N. J., Tapani, K., Lauronen, L. & Vanhatalo, S. A dataset of neonatal EEG recordings with seizure annotations. *Sci. Data* **6**, 190039 (2019).

150. Zhang, D. & Ding, H. Calculation of compact amplitude-integrated EEG tracing and upper and lower margins using raw EEG data. *Health (N. Y.)* **5**, 885–891 (2013).

151. He, H., Bai, Y., Garcia, E. A. & Li, S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)* 1322–1328 (2008). doi:10.1109/IJCNN.2008.4633969.

152. ADASYN (improves class balance, extension of SMOTE). https://www.mathworks.com/matlabcentral/fileexchange/50541-adasyn-improves-class-balance-extension-of-smote.

153. Train support vector machine (SVM) classifier for one-class and binary classification - MATLAB fitcsvm. https://www.mathworks.com/help/stats/fitcsvm.html#bt9w6j6-Standardize.

154. Train neural network classification model - MATLAB fitcnet. https://www.mathworks.com/help/stats/fitcnet.html#mw_a9fa4524-aecf-4d9b-95ef-58914f7780f3.

155. Predictor importance estimates by permutation of out-of-bag predictor observations for random forest of classification trees - MATLAB. https://www.mathworks.com/help/stats/classificationbaggedensemble.oobpermutedpredictorimportance.html.

156. Altmann, A., Toloşi, L., Sander, O. & Lengauer, T. Permutation importance: a corrected feature importance measure. *Bioinformatics* **26**, 1340–1347 (2010).

157. Amplitude-integrated EEG Classification and Interpretation in Preterm and Term Infants | American Academy of Pediatrics. https://neoreviews.aappublications.org/content/7/2/e76?sso=1&sso_redirect_count=4&nfstatus=401&nfstatus=401&nftoken=00000000-0000-0000-0000-000000000000&nftoken=00000000-0000-0000-0000-000000000000&nfstatusdescription=ERROR%3A%20No%20local%20token&nfstatusdescription=ERROR%3A%20No%20local%20token&utm_source=TrendMD&utm_medium=TrendMD&utm_campaign=NeoRev_TrendMD_0.

158. Panayiotopoulos, C. P. *Neonatal Seizures and Neonatal Syndromes*. *The Epilepsies: Seizures, Syndromes and Management* (Bladon Medical Publishing, 2005).

159. Major, P. *et al*. Periictal activity in cooled asphyxiated neonates with seizures. *Seizure* **47**, 13–16 (2017).

160. Shellhaas, R. A. & Clancy, R. R. Characterization of neonatal seizures by conventional EEG and single-channel EEG. *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.* **118**, 2156–2161 (2007).

161. Bruns, N. *et al*. Application of an Amplitude-integrated EEG Monitor (Cerebral Function Monitor) to Neonates. *J. Vis. Exp. JoVE* (2017) doi:10.3791/55985.

162. Tacke, M. *et al*. Effects of a reduction of the number of electrodes in the EEG montage on the number of identified seizure patterns. *Sci. Rep.* **12**, 4621 (2022).

163. Parameswaran, S. *et al*. COMPARISON OF VARIOUS EEG ELECTRODE PLACEMENT SYSTEMS TO DETECT EPILEPTIFORM ABNORMALITIES IN INFANTS. *Malang Neurol. J.* **7**, 30–33 (2021).

164. Stevenson, N. J., Lauronen, L. & Vanhatalo, S. The effect of reducing EEG electrode number on the visual interpretation of the human expert for neonatal seizure detection. *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.* **129**, 265–270 (2018).

165. Temko, A., Thomas, E., Marnane, W., Lightbody, G. & Boylan, G. EEG-based neonatal seizure detection with Support Vector Machines. *Clin. Neurophysiol.* **122**, 464 (2011).

166. Temko, A., Marnane, W., Boylan, G. & Lightbody, G. Clinical implementation of a neonatal seizure detection algorithm. *Decis. Support Syst.* **70**, 86–96 (2015).

167. Panayiotopoulos, C. P. Neonatal Seizures and Neonatal Syndromes. in *The Epilepsies: Seizures, Syndromes and Management* (Bladon Medical Publishing, 2005).

168. Black, R. E. *et al*. Global, regional, and national causes of child mortality in 2008: a systematic analysis. *The Lancet* **375**, 1969–1987 (2010).

169. Haibo He, Yang Bai, Garcia, E. A., & Shutao Li. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)* 1322–1328 (IEEE, 2008). doi:10.1109/IJCNN.2008.4633969.

170. Train support vector machine (SVM) classifier for one-class and binary classification - MATLAB fitcsvm. https://www.mathworks.com/help/stats/fitcsvm.html.

171. Abend, N. S. & Wusthoff, C. J. Neonatal Seizures and Status Epilepticus. *J. Clin. Neurophysiol. Off. Publ. Am. Electroencephalogr. Soc.* **29**, 441–448 (2012).

172. Einspieler, C., Peharz, R. & Marschik, P. B. Fidgety movements – tiny in appearance, but huge in impact. *J. Pediatr. (Rio J.)* **92**, S64–S70 (2016).

173. Callanan, C. *et al*. Children followed with difficulty: how do they differ? *J. Paediatr. Child Health* **37**, 152–156 (2001).

174. Tin, W., Fritz, S., Wariyar, U. & Hey, E. Outcome of very preterm birth: children reviewed with ease at 2 years differ from those followed up with difficulty. *Arch. Dis. Child. Fetal Neonatal Ed.* **79**, F83–F87 (1998).

175. Miller, S. P. *et al*. Patterns of brain injury in term neonatal encephalopathy. *J. Pediatr.* **146**, 453–460 (2005).

176. Shapiro, K. A. *et al*. Early changes in brain structure correlate with language outcomes in children with neonatal encephalopathy. *NeuroImage Clin.* **15**, 572–580 (2017).

177. Lee, B. L. *et al*. Long-term cognitive outcomes in term newborns with watershed injury caused by neonatal encephalopathy. *Pediatr. Res.* **92**, 505–512 (2022).

178. Motavaf, M. & Piao, X. Oligodendrocyte Development and Implication in Perinatal White Matter Injury. *Front. Cell. Neurosci.* **15**, (2021).

179. Back, S. A. White Matter Injury in the Preterm Infant: Pathology and Mechanisms. *Acta Neuropathol. (Berl.)* **134**, 331–349 (2017).

180. Li, A. M. *et al*. White Matter Injury in Term Newborns With Neonatal Encephalopathy. *Pediatr. Res.* **65**, 85–89 (2009).

181. Clinical and Imaging Findings of Neonatal Seizures Presenting as Diffuse Cerebral White Matter Abnormality on Diffusion-Weighted Imaging without any Structural or Metabolic Etiology. *J. Korean Soc. Radiol. Taehan Ŏngsang Ŭihakhoe Chi* **81**, 1412–1423 (2020).

182. Mao, Y. *et al*. White Matter Hyperintensities on Brain Magnetic Resonance Imaging in People with Epilepsy: A Hospital-Based Study. *CNS Neurosci. Ther.* **22**, 758–763 (2016).

183. Marchi, N. & Lerner-Natoli, M. Cerebrovascular remodeling and epilepsy. *Neurosci. Rev. J. Bringing Neurobiol. Neurol. Psychiatry* **19**, 304–312 (2013).

184. Moritani, T., Smoker, W. R. K., Sato, Y., Numaguchi, Y. & Westesson, P.-L. A. Diffusion-Weighted Imaging of Acute Excitotoxic Brain Injury. *AJNR Am. J. Neuroradiol.* **26**, 216–228 (2005).

185. Forbes, K. P. N., Pipe, J. G. & Bird, R. Neonatal Hypoxic-ischemic Encephalopathy: Detection with Diffusion-weighted MR Imaging. *AJNR Am. J. Neuroradiol.* **21**, 1490–1496 (2000).

186. Temko, A. *et al.* Toward a Personalized Real-Time Diagnosis in Neonatal Seizure Detection. *IEEE J. Transl. Eng. Health Med.* **5**, 1–14 (2017).