A PILOT STUDY USING MACHINE LEARNING FOR CLASSIFICATION OF PAIN-RELATED VERSUS NON-PAIN-RELATED ELECTROENCEPHALOGRAPHIC ACTIVITY IN PRETERM INFANTS

LOJAIN HAMWI

A THESIS SUBMITTED TO

THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF ARTS

GRADUATE PROGRAM IN CLINICAL-DEVELOPMENTAL PSYCHOLOGY

YORK UNIVERSITY

TORONTO, ONTARIO

JULY 2024

© Lojain Hamwi, 2024

Abstract

Effective pain assessment and management are crucial to mitigate both immediate and long-term consequences of prolonged NICU stays. Accurately assessing pain in premature infants is challenging due to their inability to verbally communicate their pain, the potential judgement bias by caregivers, the lack of specificity in current pain assessment tools and time constraints in a busy hospital environment. This pilot study explores a machine learning approach to support pain assessment in neonatal care using cortical activity.

The current study aims to test machine learning models that autonomously distinguishes non-pain related from pain-related cortical activity. The present dataset includes 72 preterm infants (27 females), born between 24- and 36-weeks gestational age, from two NICUs: Mount Sinai Hospital (Toronto, Canada) and University College London Hospital (London, UK). The primary outcome was to assess the accuracy of various machine learning models (XGBoost, Support Vector Machines, Random Forest, Logistic Regression, Convolutional Neural Networks) in distinguishing EEG features within a one-second pre-lance epoch (non-pain related) from a one-second post-lance epoch (pain-related). Performance metrics varied across post-menstrual age groups, reflecting developmental differences in EEG patterns. Machine learning algorithms can autonomously distinguish the one-second epoch immediately following a heel lance from the one-second epoch immediately preceding the procedure in preterm infants. Moreover, the performance of these algorithms improves with increasing postmenstrual age, demonstrating greater accuracy and reliability in older infants. This study provides a foundation for developing an autonomous and accurate tool for pain assessment in neonatal patients that can improve pain management practices in NICUs.

Keywords: Infant Pain Assessment; Preterm Infants; Machine Learning; Neonatal Intensive Care Unit; Brain Activity

Acknowledgements

In the name of God, the Most Gracious, the Most Merciful

I begin by expressing profound gratitude to God for His wisdom and the divine orchestration that has guided every step of my academic endeavors. Any achievement I attain stands as a testament to His blessings and guidance.

To my incredible supervisor, Dr. Rebecca Pillai Riddell, I express my deepest appreciation for your exceptional guidance and mentorship. Your passion for research and commitment to excellence have been a tremendous source of motivation. The belief you have shown in my abilities and the opportunities you provided me have been instrumental to the success of this work.

To my committee member, Dr. Steven Wang, I am thankful for your contributions and expertise. Thank you, Dr. Hang Du, for your collaboration and insightful feedback. I wish to thank the Sinai research team—Leslie Johannsson, Dr. Carol Cheng, Dr. Vibhuti Shah, and the Research Assistants—whose dedication enabled this research to be possible. I also extend my sincere appreciation to Juliette Champaud, Dr. Mohamed Rupawala, Dr. Lara Pierce, and Dr. Priyanka Ghosh for introducing me to the world of EEG data. Learning from you has been invaluable. A special thanks to Affan Jilani and Vadim Tyuryaev, my keen tutors, for your instrumental role in my learning of new methodologies and for your resolute support in this project.

To the faculty and staff in the Department of Psychology, thank you for your invaluable assistance throughout my Master's degree. I am grateful for the funding sources that allowed me to pursue my graduate school studies: The Lillian Meighen and Don Wright Foundation, the Social Sciences and Humanities Research Council, and the Ontario Graduate Scholarship.

To my OUCH Lab sisters—Oana Bucsea, Ilana Shiff, Sara Jasim, Haleh Hashemi, and Estreya Cohen—I am profoundly grateful for your steadfast support and inspiration. Your friendship has made this academic journey truly special.

To my many wonderful friends, I am deeply grateful for your continuous encouragement. Thank you for all the FaceTime calls, for listening, for sharing your laughter, and for always being my cheerleaders.

And finally, my warmest thanks go to my beautiful family, to whom I dedicate this thesis. To my mom, thank you for your sacrifices, wisdom, and guidance. Your unwavering support and encouragement have been the foundation of all my achievements, and you remain my constant source of strength. To my dad, thank you for being my biggest fan and for your hard work, which has allowed me to pursue my goals. To my siblings, Marwan, Mohanad, and Lean, who have unconditionally supported my academic endeavors, celebrated my successes, and reminded me to take breaks, thank you.

Publication Disclosure

The study comprising this thesis is under review for publication. The manuscript citation is provided below.

Hamwi, L., Du, H., Jasim, S., Wang, X., Shah, V., Cheng, C., Fabrizi, L., Meek, J., Fitzgerald, M., Racine, N., Stedman, I., & Pillai Riddell, R*. (submitted). Machine learning classification of EEG responses of pain versus non-pain-related stimulus in preterm infants.

* Corresponding author. Email: rpr@yorku.ca

Table of Contents

Abstract ii
Acknowledgements iii
Publication Disclosure iv
Table of Contents v
List of Tables vi
List of Figures vii
Chapter 1: Introduction 1
Effects of Pain Exposure on Preterm Infants 2
Current Preterm Infant Pain Assessment Approaches 4
Electroencephalography as An Assessment Tool 7
Machine Learning in Infant Pain Assessment 9
Ethical, Legal, and Social Implications of AI in the NICU 14
References 16
Chapter 2: Machine Learning Classification of Pain Versus Non-Pain-Related Electroencephalographic Activity in Preterm Infants 29
Abstract 30
Introduction 32
Method 35
Participants 35
Procedure 35
Signal Acquisition. 36
Preprocessing. 36
Feature Extraction. 37
Feature Selection. 37
Classification. 38
Performance Evaluation. 38
Results 39
Youngest PMA Group (≤ 31 weeks, 6 days) 39
Middle PMA Group (32 to 33 weeks, 6 days) 40
Oldest PMA Group (≥ 34 weeks) 40
Discussion 40
Performance across PMA groups 41
References 45
Supplementary Material 58
Chapter 3: Concluding Notes 68
Limitations and Future Directions 69
Summary 70
References 72

List of Tables

Table 1. Demographic Data of Participants 54
Table 2. Important Features identified by XGBoost and Random Forest 55
Table 3. Summary of Metrics of Binary Classification Across Preterm Infants PMA groups..... 56

List of Figures

Figure 1: Overview of Current Study Procedure. 57

Chapter 1: Introduction

Approximately 8% of Canadian children are born prematurely (less than 37 weeks gestational age) each year[1]. These preterm infants often require extended care in the Neonatal Intensive Care Unit (NICU), where they are subjected to an average of 12–17 painful procedures daily within their first 14 days of life[2,3]. Frequent early pain exposure has been linked to adverse effects on the development of the central nervous system, leading to structural and functional cortical changes[4]. These changes are associated with cognitive and motor difficulties during infancy and school age, as well as poorer neurodevelopmental outcomes later in life, including altered pain perception and stress regulation[5-9].

Effective pain assessment and management are crucial to mitigate both the immediate and long-term consequences of prolonged NICU stays[10]. However, the major challenge lies in objectively assessing pain in non-verbal infants. Current neonatal pain assessment scales rely on behavioral (e.g., facial expressions, crying patterns) and/or physiological responses (e.g., heart rate, cortisol concentrations) as indirect pain indicators[2,11,12]. Despite their value, these methods face significant challenges, particularly in premature infants. In response, researchers worldwide are exploring the use of artificial intelligence (AI) technologies to enhance neonatal pain assessment. AI technologies hold the promise of improving the accuracy, feasibility, and comprehensiveness of infant pain assessments. This study builds on these advancements by proposing a novel AI-driven approach to refine pain assessment methods in neonatal care using cortical activity.

Effects of Pain Exposure on Preterm Infants

Preterm birth and the subsequent stay in the Neonatal Intensive Care Unit (NICU) can have profound impacts on the developing brain of neonates. The NICU environment, with its various stressors, contributes to regional alterations in brain structure and function. These

changes include decreased frontal and parietal brain width, altered diffusion measures, and functional connectivity in the temporal lobes, which are linked to abnormalities in motor behavior on neurobehavioral examinations[13].

Frequent early exposure to pain in neonates further exacerbates these impacts, leading to structural and functional cortical changes, particularly in regions involved in somatosensory processing[14]. This is especially pronounced in extremely premature neonates, where early pain is associated with slower thalamic macrostructural growth and volume losses localized to somatosensory regions[8]. Additionally, greater exposure to neonatal procedural pain is associated with reduced white matter fractional anisotropy (FA) and reduced subcortical grey matter NAA/choline ratios, indicating impaired brain development[4].

Increased exposure to pain in early life is also linked to the development of attention-deficit disorders, atypical behaviors such as hypervigilance and exaggerated startle responses, and sensorimotor impairments in children, including impaired vision or hearing, cerebral palsy, and impaired intellect observed into childhood and adulthood[9]. Prolonged neonatal pain triggers excessive N-methyl-D-aspartate activation, leading to excitotoxic damage in the brain. Over the long term, this is associated with reduced brain size in the frontal and parietal regions and the development of stress-related psychosocial disabilities (hypervigilance, exaggerated startle responses, anxiety)[15-18], affecting somatosensory processing by lowering pain thresholds and increasing pain sensitivity. Furthermore, Giordano and colleagues demonstrated that very preterm infants who experienced higher levels of pain in the NICU showed significant deficits in mental and motor development at a corrected age of 12 months[19]. These developmental differences persisted even after controlling for major preterm birth morbidities. While not the focus of the current project, it is critical to note that these painful procedures often occur outside

of the typical primary social context of parents (i.e., parent separation due to NICU hospitalization), which compounds the impact of painful procedures on the developing neonate[20]. Early-life pain and stress exposure have profound long-term consequences on the neurodevelopmental and psychosocial trajectories of preterm infants. These findings underscore the critical need for pain management in neonatal care, which is predicated on accurate pain assessment.

Current Preterm Infant Pain Assessment Approaches

Accurately assessing pain in preterm infants is crucial for delivering the specialized care they require during this vulnerable stage of their development. A recent systematic review offered a thorough analysis of currently available tools for evaluating neonatal pain[21]. Existing scales were developed to assess pain, with a smaller number evaluating sedation. As of today, 65 scales, all of which underwent validity testing, can be used to measure various types of pain including acute, prolonged, and postoperative pain. Specifically, 37 scales were validated for preterm infants, with several also suitable for different age groups, ranging from preterm and term infants to toddlers. Moreover, more than half of these scales employ behavioral methods to evaluate pain, reflecting a common approach in assessing infant discomfort.

Research to date highlights the most evidenced behavioral indicators of pain in infants are facial expressions and body movements[22]. Prominent among these are the Neonatal Facial Coding System (NFCS)[23], Behavioral Indicators of Infant Pain (BIIP)[24] and The Neonatal Infant Pain Score (NIPS)[25]. These tools assess pain through observable behaviors including facial expressions (e.g., brow bulge, eye squeeze, nasolabial furrow), crying, body movements, and vital signs (e.g., heart rate, breathing patterns). Physiological measures have also been used as indicators of pain or stress assessment, though they are less frequently used on their own[6]. The

typical physiological indicators monitored at bedside in the NICU include heart rate, respiratory rate, oxygen saturation, and blood pressure[26]. The most commonly used scale for pain assessment in the NICU is the Premature Infant Pain Profile (PIPP-R)[27] and the Neonatal Pain, Agitation and Sedation Scale (N-PASS)[28], which utilize a combination of both behavioural and physiological measures.

Challenges and Limitations

While these tools are invaluable in clinical practice, they are not without their challenges. These include inherent judgement bias and the feasibility limitations associated from a busy hospital environment[29,30]. Additionally, many assessment tools often fail to consider critical infant-specific clinical data such as sex differences, gestational age, postmenstrual age, pain history, and concurrent illness, as well as the nature of the pain—whether it is acute-procedural, acute-prolonged, or chronic—which are crucial for accurate pain evaluation[2,6,31].

Furthermore, while some studies have shown significant positive correlations between multidimensional scale pain responses (e.g., PIPP scores) and increases in the magnitude of nociceptive cortical activity, the relationship between behavioral indicators and cortical responses is not consistently reliable[32-35]. This inconsistency is particularly evident under conditions such as high stress or following sucrose administration, where behavioral responses may not fully reflect the underlying changes in cortical activity detected by electroencephalographic features[33,36].

Pain assessments are also compromised by the intermittent nature of bedside caregiver assessments, which risk missing critical pain episodes. The variability in observer judgments can result in inconsistent treatments with potential long-term consequences[37]. Specific challenges in the NICU setting, such as staffing constraints, insufficient training for medical personnel, time

pressures, and a reluctance to modify entrenched practices, exacerbate these issues[38]. In scenarios that demand regular pain monitoring, such as acute postoperative care, relying solely on caregiver observations becomes both resource-intensive and inefficient[38].

While physiological measures like heart rate, heart rate variability, and oxygen saturation have been associated with infant pain, their utility is complicated by factors such as medication, stress, underlying pathological conditions, and other treatments that can affect readings[39]. Furthermore, despite their sensitivity to pain, these indicators lack specificity, making it difficult to distinguish between pain-related distress and non-pain-related distress. Their use remains largely confined to research settings, limiting their application in everyday clinical practice[40].

Finally, despite the proliferation of over 60 measures, no measures to date incorporate cortical measures. Research has shown that while behavioural and physiological measures are highly sensitive to pain, they are not specific to pain but rather are elevated when an organism is under high stress and distress, regardless of the source. Cortical measures provide a potential pathway for measurement that is specific to pain-related distress[41].

Emerging tools and techniques

In the realm of objective infant pain assessment, cortical activity as measured through electroencephalography (EEG) emerges as a promising approach for assessing infant pain. Given that traditional methods may not specifically reflect the pain state of preterm infants, examining the cortical activity of neonates during painful procedures provides a more precise insight into the real-time mechanisms involved in processing pain. Furthermore, incorporating an EEG-based indicator of pain in neonates into clinical practice is compelling as it could seamlessly be added to ongoing vital monitoring systems as a visual trend, allowing for real-time observation and management of neonatal distress[32].

Electroencephalography as An Assessment Tool

EEG is highly regarded as a non-invasive technique for evaluating neurological health and predicting neurological outcomes[42]. Technological advancements have now made it feasible to extend EEG monitoring over longer periods, particularly for preterm infants[43]. Therefore, conventional EEG is recognized for its prognostic value in identifying pain in infants. Various EEG methodologies have thus been developed to analyze how infants react to pain[44]. One significant analytic technique involves event-related potentials (ERPs)—distinct wave form activity triggered by sensory, cognitive, or motor events. These are measured using techniques such as principal component analysis (PCA) and peak-to-peak amplitude detection[36]. By quantifying ERPs, researchers can directly observe the cortical processing of painful stimuli, providing a more accurate assessment of neonatal pain compared to solely behavioral or autonomic responses[45,46].

Recent studies have leveraged ERPs evoked by nociceptive stimulation as a key index for neonatal pain assessment[35,47,48]. For instance, Jones et al. identified a specific nociceptive ERP component with a characteristic N3P3 waveform in response to heel lance procedures, noting these components around 500-700 milliseconds (ms) post-stimulus[48]. The development of a template pattern of nociceptive brain activity in infants at the electrode Cz (central midline electrode), contrasting the patterns of neonates' brain activity evoked by noxious and non-noxious stimulation, marks an important initial foundation towards eventually using cortical activity in pain response. Using this method, research by Hauck et al. and Hartley et al. demonstrate that the magnitude of brain activity evoked by a noxious heel lance significantly differs from that induced by non-noxious stimuli, particularly within the critical time window of 400 to 700 ms after stimulation[49,50]. These findings support the distinctive nature of cortical

activity in response to noxious versus non-noxious stimuli, underscoring the potential of EEG recordings as a robust indicator for pain assessment.

However, the predominant use of a single-channel, usually the Cz electrode site—the vertex electrode positioned on the midline of the scalp—may not fully capture the distributed nature of nociceptive brain activity, potentially limiting the comprehensiveness of assessments[50]. New directions in measuring cortical pain responses are moving towards microstate analyses[35,51]. These analyses involve using multichannel EEG to characterize time-varying cortical potentials, reducing complex spatial patterns to representative topographic maps. Bucsea et al. specifically found that pain-related facial activity levels are associated with differences in nociceptive brain activity patterns[35]. They further highlighted a separate, yet interrelated sequence of early activity that correlates with the intensity of the immediate behavioral response, suggesting a complex interaction between the cortical activation and the observable pain behavior. Their results indicated that cortical responses to pain occur independently of outward behaviors and may differ between full-term and preterm infants. This suggests that neonatal pain behaviors reflect only a portion of the overall cortical pain response, emphasizing the need for comprehensive, multi-channel EEG approaches.

Despite these advances, the manual, quantitative analysis of cortical brain activity via EEG is often hindered by its lack of robustness and the labor-intensive nature of visual inspections, which are susceptible to subjective bias and can be complicated by pain-induced motor movements[52,53]. In response, researchers are exploring alternative methods that incorporate advanced signal processing and machine learning techniques to improve neonatal pain assessment. These innovations aim to develop automated systems capable of recognizing patterns from complex data, thus promising a more effective and fully automated approach to

pain monitoring in clinical settings. Such advancements have the potential to transform neonatal care by providing real-time, accurate assessments of physiological indicators, enhancing pain management, and alleviating the burden of pain on preterm infants[6,54,55].

Machine Learning in Infant Pain Assessment

Recent systematic reviews of the application of machine learning techniques in the field of pain medicine demonstrate the wide range of applications of machine learning in pain medicine, including pain assessment, prediction, diagnosis, and treatment optimization[56-58]. In 2022, over 44 research studies have been reported to utilizing machine learning in their pain assessment[59].

Machine learning models have shown promise in accurately predicting pain intensity, identifying pain patterns, and classifying different types of pain. However, most of the existing techniques for automatic pain assessment are primarily geared towards adults, which may not perform as effectively when used with infants[60].

Although limited in practice, current evidence on the use of machine learning in the assessment and monitoring of infant pain also appears to be promising[61,62]. In particular, there is a growing number of facial expression-centered machine learning methods for automatic infant pain assessment that have demonstrated quite satisfactory performance, as summarized in a recent review[63,64]. The 15 studies found have demonstrated promising results, with accuracies ranging from 85% to nearly 100% across various settings. For instance, one study achieved perfect sensitivity and specificity in detecting procedural pain, while another reported an accuracy of 93.07% alongside notable AUC scores. These models utilize diverse methods and datasets, including convolutional neural networks and support vector machines, and are generally effective at distinguishing between neutral and painful expressions. Findings from a recent study

also found relevant facial regions to discriminate neonatal pain to AI-based model were different from those observed by health professionals and parents of neonates[65].

However, significant challenges persist in using facial expressions for pain assessment, beyond the lack of specificity to pain mentioned above that is a challenge to all behaviorally based measures[54,66]. Key limitations include difficulties in autodetecting facial points obscured by medical devices or low lighting in the NICU, and many studies have utilized static photographs or pre-recorded videos, which diminishes their practicality in real-world clinical settings. Additionally, hospitalized preterm behaviors are more subtle, making it difficult to discern pain intensity nuances that might necessitate treatment adjustments. Furthermore, due to their developmental stage, some infants have movement disorders or demonstrate physical exertion such as exhaustion or sedation, thus having a restricted capacity to express pain behaviorally[67]. These models require further validation across varied neonatal populations and clinical environments to improve their accuracy and minimize subjectivity in pain assessment. Thus, there is a need to create machine learning approaches that capitalize on other pain indicators to address gaps in the field[68].

Albeit few in number, some studies have utilized multimodal approaches that integrate multiple pain indicators to enhance the accuracy of infant pain assessment[69-71]. For example, Zamzmi and colleagues proposed an automated system that combines behavioral and physiological indicators to assess infants' pain, demonstrating significant efficacy[71]. In their findings, recognizing pain using single indicators—facial expression, body movement, or vital signs—yielded overall accuracies of 88%, 85%, and 82% respectively. However, combining these indicators into a multimodal approach increased the accuracy to 95%, underscoring the superior performance of this strategy. Similarly, Salekin et al. created a temporal multimodal AI-

based system that initially generates pain scores from video (face, body) and audio (crying sound) signals independently, known as the unimodal approach[69]. These scores are then combined using decision fusion to predict the final pain assessment. Their findings suggest that the multimodal approach for assessing postoperative pain in neonates achieved an accuracy of 79%, which was 6.33% higher than the unimodal approaches, making it a viable alternative to manual assessment.

Other studies underway have proposed using facial electromyography to record facial muscle activity-related infant pain, ECG to examine heart rate (HR) changes and HR variability, electrodermal activity (skin conductance) to measure catecholamine-induced palmar sweating, changes in oxygen saturations and skin perfusion, and EEG to assess brain activity in real time[72]. As noted in the ClinicalTrials.gov Registry, following data collection, Roué and colleagues plan on developing a machine-learning algorithm to test a sensor-fusion framework designed to integrate data from different sensor modalities[72]. To our knowledge, no updates have been released from that research team. Given the challenges of collecting data with preterm infants, it is not unexpected that initial approaches to attempting to use machine learning have focused on readily available indicators of pain – behaviour, cardiac, oxygen saturation. However, given their lack of specificity to pain, involving the brain in autonomous pain assessment is a critical future direction.

EEG in Machine Learning

Scalp-recorded EEG is a crucial tool for bedside assessment of newborn brain function, but its manual and visual interpretation can be subjective, complicating reliability[73]. Moreover, obtaining EEG data in hospitalized premature infants is extremely challenging given the emotional state of parents having most often given birth unexpectedly to a premature infant and

the fragile medical state of the infant. However, previous research has shown that automatic analysis of EEG in preterms can be reliable, feasible, and effective. For instance, automatic methods for detecting bursting patterns in both single-channel[74,75] and multi-channel EEGs[76-78] have been developed. While interobserver agreement among humans in behavioural measures typically shows high reliability, this agreement decreases significantly when annotating EEG features due to the complexity and subtlety of EEG patterns[79,80]. Machine learning algorithms often achieve performance levels comparable to or even exceeding those of human experts in identifying EEG features, demonstrating their potential to provide consistent and accurate results in EEG analysis. Leveraging algorithms in this context can help mitigate variability and improve the reliability of EEG feature annotation.

Furthermore, EEG-based algorithms for preterm neonates have demonstrated satisfactory results across various metrics such as Area Under the Curve, Kappa score, and percentage consistency. These studies cover a range of applications including artefact detection[81], seizure detection[82-84], estimation of functional brain maturation[85], quiet sleep staging[86], predicting neurodevelopmental outcomes[87] and the severity of adverse drug effects[88] on preterm infants.

To date, two automated EEG-based *pain* assessment algorithms exist for neonates[89,90]. The first study included 109 infants between 34 to 42 weeks' gestational age and utilized a random forest algorithm to discriminate between responses to a noxious (clinically required heel lance) and a non-noxious stimulus (control heel lance)[89]. The multimodal model, which incorporated heart rate, respiratory rate, and EEG data, achieved an accuracy of 81% and an AUC of 0.89 for all infants in the dataset. When focusing specifically on preterm infants (n = 21), the model's accuracy increased to 88% (AUC of 0.94). Despite the promising application of EEG-based machine learning in distinguishing neonatal pain, the noxious-evoked brain activity

was assessed using a single feature and a single channel, specifically the magnitude of activity at the Cz electrode. This limitation led to a drop in performance when only EEG data was utilized across age groups, resulting in an accuracy of 64% and an AUC of 0.75. Additionally, the process was not fully automated, as an observer was required to assess and quantify the presence of peaks. Furthermore, EEG traces that included artifacts were removed, which may not reflect real-time pain monitoring using EEG.

The second study from that research team included 144 infants ranging from 28 to 40 weeks PMA and used a bagged decision tree classification model, which was created based on observed noxious-evoked cerebral, limb withdrawal, cardiovascular, and facial expression responses[90]. This model achieved an accuracy of 76% (no AUC reported) in distinguishing between noxious and non-noxious conditions in the testing set and 74% in preterm infants within the same dataset. Additionally, the model performed well in an independent dataset with an accuracy of 84% overall and 83% in preterm infants. However, in this model, brain activity was also recorded using only the Cz electrode. Furthermore, sub-analyses of accuracy in four different postmenstrual age groups (28–31, 31–34, 34–37, and 37–40 weeks) were performed, but the sample size in some subgroups was too small to generate reliable results (n = 3 in the youngest age group).

Building on these initial efforts, there remains a need to incorporate multiple EEG features and channels to provide a more comprehensive analysis of pain-related cortical activity. Additionally, to enhance the accuracy and automation of EEG-based pain assessment in neonates, it is crucial to develop methodologies that can handle EEG traces with artifacts and utilize automated processes from the outset. This approach not only aims to improve the overall

accuracy but also ensures that the assessment can be performed in real-time, reflecting the actual conditions within the NICU environment.

Ethical, Legal, and Social Implications of AI in the NICU

Before implementing the current pilot study on an AI-driven approach to pain assessment in the NICU, a careful consideration of the ethical, legal, and social implications associated with using AI for pain assessment in such sensitive environments[91,92] was undertaken. From an ethical-legal perspective, an analysis of relevant literature suggested key themes were discerned as ensuring equitable distribution of AI benefits, addressing consent and data privacy concerns, maintaining robust regulatory oversight, and considering the potential impact on caregiver-infant interactions[91]. Understanding the social context of potential key knowledge users AI in the NICU was conducted through qualitative interviews with healthcare practitioners—including physicians and nurses—and parents at two tertiary neonatal intensive care units in London, UK, and Toronto, Canada, provided valuable perspectives (n = 40). The feedback revealed general support among both healthcare professionals and parents for integrating AI technologies into the NICU. While emphasizing the importance of shared decision-making and cautioning against over-reliance on automated systems, participants also raised concerns about the effectiveness and accuracy of these technologies and the emotional impact on caregivers of continuous knowledge of pain. Despite these reservations, both healthcare practitioners and primary caregivers expressed optimism about the potential of AI to continuously monitor pain and thereby enhance the quality of care. They recognized that such technology could lead to better, more responsive care for neonates.

Current Study

The objective of the present pilot study is to explore machine learning techniques to differentiate EEG features associated with pain resulting from routine heel lance (blood draw via lancet to heel) and a one-second epoch preceding the heel lance without any skin-breaking stimulation in hospitalized preterm infants.

The following research question will be addressed: How well can machine learning models (XGBoost, Support Vector Machines, Random Forest, Logistic Regression, Convolutional Neural Network) distinguish the EEG responses during the one-second post-lance epoch from the one-second epoch immediately prior to the heel lance procedure in different postmenstrual age (PMA) groups at the time of study (youngest group: PMA ≤ 31 weeks, 6 days; middle group: PMA 32 to 33 weeks, 6 days; and oldest group: PMA ≥ 34 weeks)?

It is hypothesized that machine learning algorithms can accurately distinguish the one-second epoch immediately following a heel lance from the one-second epoch immediately preceding the procedure in preterm infants. Different machine learning models, including XGBoost, Support Vector Machines (SVM), Random Forest (RF), Logistic Regression (LR), and Convolutional Neural Networks (CNN), are anticipated to demonstrate varying levels of classification metrics (e.g., accuracy, precision, etc.) and summary performance metrics (AUC) in assessing pain. Additionally, it is hypothesized that the performance of these models in discriminating the post-lance epoch will vary across different PMA groups, with potential differences in the EEG features and model performance metrics reflective of the developmental stage of the infants. To achieve this, real-time recordings of cortical activity through continuous EEG monitoring before and during the onset of a routine heel lance procedure will be used.

References

1. Government of Canada Statistics Canada. Live births, by weeks of gestation. Published September 26, 2023. Accessed May 9, 2024. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310042501
2. Hatfield LA, Murphy N, Karp K, Polomano RC. A Systematic Review of Behavioral and Environmental Interventions for Procedural Pain Management in Preterm Infants. *J Pediatr Nurs*. 2019;44:22-30. doi:10.1016/j.pedn.2018.10.004
3. Chawanpaiboon S, Vogel JP, Moller AB, et al. Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. *Lancet Glob Health*. 2019;7(1):e37-e46. doi:10.1016/S2214-109X(18)30451-0
4. Brummelte S, Grunau RE, Chau V, et al. Procedural pain and brain development in premature newborns. *Ann Neurol*. 2012;71(3):385-396. doi:10.1002/ana.22267
5. Vinall J, Pillai Riddell R, Greenberg S. The influence of culture on maternal soothing behaviours and infant pain expression in the immunization context. *Pain Res Manag*. 2011;16(4):234-238.
6. Llerena A, Tran K, Choudhary D, et al. Neonatal pain assessment: Do we have the right tools? *Frontiers in Pediatrics*. 2023;10. Accessed September 6, 2023. https://www.frontiersin.org/articles/10.3389/fped.2022.1022751
7. Nath N, Beltrano W, Haynes L, Dewey D, Bray S. Long-Term Effects of Preterm Birth on Children's Brain Structure: An Analysis of the Adolescent Brain Cognitive Development (ABCD) Study. *eNeuro*. 2023;10(6):ENEURO.0196-22.2023. doi:10.1523/ENEURO.0196-22.2023

8. Duerden EG, Grunau RE, Guo T, et al. Early Procedural Pain Is Associated with Regionally-Specific Alterations in Thalamic Development in Preterm Neonates. *J Neurosci*. 2018;38(4):878-886. doi:10.1523/JNEUROSCI.0867-17.2017

9. Williams MD, Lascelles BDX. Early Neonatal Pain—A Review of Clinical and Experimental Implications on Painful Conditions Later in Life. *Front Pediatr*. 2020;8. doi:10.3389/fped.2020.00030

10. Grunau RE, Miller JV, Chau CM. Long-term effects of pain in infants. *Oxford Textbook of Pediatric Pain*. Published online 2021:38.

11. Roué JM, Rioualen S, Gendras J, Misery L, Gouillou M, Sizun J. Multi-modal pain assessment: are near-infrared spectroscopy, skin conductance, salivary cortisol, physiologic parameters, and Neonatal Facial Coding System interrelated during venepuncture in healthy, term neonates? *Journal of Pain Research*. 2018;11:2257-2267. doi:10.2147/JPR.S165810

12. Stoye DQ, Boardman JP, Osmond C, et al. Saliva cortisol diurnal variation and stress responses in term and preterm infants. *Archives of Disease in Childhood - Fetal and Neonatal Edition*. 2022;107(5):558-564. doi:10.1136/archdischild-2021-321593

13. Smith GC, Gutovich J, Smyser C, et al. NICU Stress Is Associated with Brain Development in Preterm Infants. *Ann Neurol*. 2011;70(4):541-549. doi:10.1002/ana.22545

14. Grunau RE, Weinberg J, Whitfield MF. Neonatal Procedural Pain and Preterm Infant Cortisol Response to Novelty at 8 Months. *Pediatrics*. 2004;114(1):e77-e84. doi:10.1542/peds.114.1.e77

15. Bouza H. The impact of pain in the immature brain. *The Journal of Maternal-Fetal & Neonatal Medicine*. 2009;22(9):722-732. doi:10.3109/14767050902926962

16. Ranger M, Grunau RE. Early repetitive pain in preterm infants in relation to the developing brain. *Pain Manag*. 2014;4(1):57-67. doi:10.2217/pmt.13.61

17. Vinall J, Grunau RE. Impact of repeated procedural pain-related stress in infants born very preterm. *Pediatr Res*. 2014;75(5):584-587. doi:10.1038/pr.2014.16

18. Lilla M, Stadelman-Diaw C, Ramelet AS. Prolonged pain in neonates: retrospective analysis. *Rech Soins Infirm*. 2013;(115):36-48.

19. Giordano V, Deindl P, Gal E, et al. Pain and neurodevelopmental outcomes of infants born very preterm. *Developmental Medicine & Child Neurology*. 2023;65(8):1043-1052. doi:10.1111/dmcn.15505

20. Shiff I, Bucsea O, Pillai Riddell R. Psychosocial and Neurobiological Vulnerabilities of the Hospitalized Preterm Infant and Relevant Non-pharmacological Pain Mitigation Strategies. *Front Pediatr*. 2021;9. doi:10.3389/fped.2021.568755

21. Giordano V, Edobor J, Deindl P, et al. Pain and Sedation Scales for Neonatal and Pediatric Patients in a Preverbal Stage of Development: A Systematic Review. *JAMA Pediatr*. 2019;173(12):1186-1197. doi:10.1001/jamapediatrics.2019.3351

22. Hatfield LA, Ely EA. Measurement of Acute Pain in Infants: A Review of Behavioral and Physiological Variables. *Biological Research For Nursing*. 2015;17(1):100-111. doi:10.1177/1099800414531448

23. Grunau RV, Craig KD. Pain expression in neonates: Facial action and cry. *Pain*. 1987;28(3):395-410. doi:10.1016/0304-3959(87)90073-X

24. Holsti L, Grunau RE. Initial validation of the Behavioral Indicators of Infant Pain (BIIP). *Pain*. 2007;132(3):264-272. doi:10.1016/j.pain.2007.01.033

25. Lawrence J, Alcock D, McGrath P, Kay J, MacMurray SB, Dulberg C. The development of a tool to assess neonatal pain. *Neonatal Netw*. 1993;12(6):59-66.

26. Holsti L, Grunau RE, Shany E. Assessing Pain in Preterm Infants in the Neonatal Intensive Care Unit: Moving to a 'Brain-Oriented' Approach. *Pain Management*. 2011;1(2):171-179. doi:10.2217/pmt.10.19

27. Stevens BJ, Gibbins S, Yamada J, et al. The premature infant pain profile-revised (PIPP-R): initial validation and feasibility. *Clin J Pain*. 2014;30(3):238-243. doi:10.1097/AJP.0b013e3182906aed

28. Hummel P, Puchalski M, Creech SD, Weiss MG. Clinical reliability and validity of the N-PASS: neonatal pain, agitation and sedation scale with prolonged pain. *J Perinatol*. 2008;28(1):55-60. doi:10.1038/sj.jp.7211861

29. Pillai Riddell RR, Jasim S, Hamwi L. Out of the mouth of babes: a lot about pain has nothing to do with pain. *Pain*. 2022;163(Suppl 1):S117-S125. doi:10.1097/j.pain.0000000000002761

30. Boyle EM, Bradshaw J, Blake KI. Persistent pain in neonates: challenges in assessment without the aid of a clinical tool. *Acta Paediatrica*. 2018;107(1):63-67. doi:10.1111/apa.14081

31. Zhi R, Goldgof GZD, Ashmeade T, Sun Y. Automatic Infants' Pain Assessment by Dynamic Facial Representation: Effects of Profile View, Gestational Age, Gender, and Race. *J Clin Med*. 2018;7(7):173. doi:10.3390/jcm7070173

32. Norman E, Rosén I, Vanhatalo S, et al. Electroencephalographic Response to Procedural Pain in Healthy Term Newborn Infants. *Pediatr Res*. 2008;64(4):429-434. doi:10.1203/PDR.0b013e3181825487

33. Jones L, Laudiano-Dray MP, Whitehead K, et al. EEG, behavioural and physiological recordings following a painful procedure in human neonates. *Sci Data*. 2018;5(1):180248. doi:10.1038/sdata.2018.248

34. Slater R, Cantarella A, Franck L, Meek J, Fitzgerald M. How Well Do Clinical Pain Assessment Tools Reflect Pain in Infants? *PLOS Medicine*. 2008;5(6):e129. doi:10.1371/journal.pmed.0050129

35. Bucsea O, Rupawala M, Shiff I, et al. Clinical thresholds in pain-related facial activity linked to differences in cortical network activation in neonates. *Pain*. 2023;164(5):1039-1050. doi:10.1097/j.pain.0000000000002798

36. Slater R, Fabrizi L, Worley A, Meek J, Boyd S, Fitzgerald M. Premature infants display increased noxious-evoked neuronal activity in the brain compared to healthy age-matched term-born infants. *NeuroImage*. 2010;52(2):583-589. doi:10.1016/j.neuroimage.2010.04.253

37. Lötsch J, Ultsch A, Mayer B, Kringel D. Artificial intelligence and machine learning in pain research: a data scientometric analysis. *Pain Rep*. 2022;7(6):e1044. doi:10.1097/PR9.0000000000001044

38. Popowicz H, Mędrzycka-Dąbrowska W, Kwiecień-Jaguś K, Kamedulska A. Knowledge and Practices in Neonatal Pain Management of Nurses Employed in Hospitals with Different Levels of Referral—Multicenter Study. *Healthcare*. 2021;9(1):48. doi:10.3390/healthcare9010048

39. Ilhan E, Pacey V, Brown L, et al. Exploration and Validation of Behavioral Pain Measures and Physiological Pain Measures Factor Structure Extracted From the Pain Assessment Tool

Item Scores for Infants Admitted to Neonatal Intensive Care. *The Clinical Journal of Pain*. 2021;37(6):397. doi:10.1097/AJP.0000000000000931

40. Olsson E, Ahl H, Bengtsson K, et al. The use and reporting of neonatal pain scales: a systematic review of randomized trials. *PAIN*. 2021;162(2):353. doi:10.1097/j.pain.0000000000002046

41. Pillai Riddell R, Fitzgerald M, Slater R, Stevens B, Johnston C, Campbell-Yeo M. Using only behaviours to assess infant pain: a painful compromise? *Pain*. 2016;157(8):1579-1580. doi:10.1097/j.pain.0000000000000598

42. Miranda P, Cox C, Alexander M, Danev S, Lakey J. Overview of current diagnostic, prognostic, and therapeutic use of EEG and EEG-based markers of cognition, mental, and brain health. Integr Mol Med. 2019;6:1-10. doi:10.15761/IMM.1000378.

43. Schumacher EM, Westvik AS, Larsson PG, Lindemann R, Westvik J, Stiris TA. Feasibility of Long-Term Continuous EEG Monitoring During the First Days of Life in Preterm Infants: An Automated Quantification of the EEG Activity. *Pediatr Res*. 2011;69(7):413-417. doi:10.1203/PDR.0b013e31821267d2

44. Benoit B, Martin-Misener R, Newman A, Latimer M, Campbell-Yeo M. Neurophysiological assessment of acute pain in infants: a scoping review of research methods. *Acta Paediatrica*. 2017;106(7):1053-1066. doi:10.1111/apa.13839

45. Tracey I, Mantyh PW. The cerebral signature for pain perception and its modulation. *Neuron*. 2007;55(3):377-391. doi:10.1016/j.neuron.2007.07.012

46. Verriotis M, Fabrizi L, Lee A, Cooper RJ, Fitzgerald M, Meek J. Mapping Cortical Responses to Somatosensory Stimuli in Human Infants with Simultaneous Near-Infrared

Spectroscopy and Event-Related Potential Recording. *eNeuro*. 2016;3(2). doi:10.1523/ENEURO.0026-16.2016

47. Shiroshita Y, Kirimoto H, Ozawa M, et al. Can Event-Related Potentials Evoked by Heel Lance Assess Pain Processing in Neonates? A Systematic Review. *Children*. 2021;8(2):58. doi:10.3390/children8020058

48. Jones L, Laudiano-Dray MP, Whitehead K, et al. The impact of parental contact upon cortical noxious-related activity in human neonates. *European Journal of Pain*. 2021;25(1):149-159. doi:10.1002/ejp.1656

49. Hauck AGV, Vaart M van der, Adams E, et al. Effect of parental touch on relieving acute procedural pain in neonates and parental anxiety (Petal): a multicentre, randomised controlled trial in the UK. *The Lancet Child & Adolescent Health*. 2024;8(4):259-269. doi:10.1016/S2352-4642(23)00340-1

50. Hartley C, Duff EP, Green G, et al. Nociceptive brain activity as a measure of analgesic efficacy in infants. *Science Translational Medicine*. 2017;9(388):eaah6122. doi:10.1126/scitranslmed.aah6122

51. Rupawala M, Bucsea O, Laudiano-Dray MP, et al. A developmental shift in habituation to pain in human neonates. *Current Biology*. 2023;33(8):1397-1406.e5. doi:10.1016/j.cub.2023.02.071

52. San-Segundo R, Gil-Martín M, D'Haro-Enríquez LF, Pardo JM. Classification of epileptic EEG recordings using signal transforms and convolutional neural networks. *Computers in Biology and Medicine*. 2019;109:148-158. doi:10.1016/j.compbiomed.2019.04.031

53. Maimon N, Grunau RE, Cepeda IL, et al. Electroencephalographic Activity in Response to Procedural Pain in Preterm Infants Born at 28 and 33 Weeks Gestational Age. *The Clinical Journal of Pain*. 2013;29(12):1044. doi:10.1097/AJP.0b013e318284e525

54. Hughes JD, Chivers P, Hoti K. The Clinical Suitability of an Artificial Intelligence–Enabled Pain Assessment Tool for Use in Infants: Feasibility and Usability Evaluation Study. *Journal of Medical Internet Research*. 2023;25(1):e41992. doi:10.2196/41992

55. Cascella M, Schiavo D, Cuomo A, et al. Artificial Intelligence for Automatic Pain Assessment: Research Methods and Perspectives. *Pain Res Manag*. 2023;2023:6018736. doi:10.1155/2023/6018736

56. Zhang M, Zhu L, Lin SY, et al. Using artificial intelligence to improve pain assessment and pain management: a scoping review. *Journal of the American Medical Informatics Association*. 2023;30(3):570-587. doi:10.1093/jamia/ocac231

57. Gkikas S, Tsiknakis M. Automatic assessment of pain based on deep learning methods: A systematic review. *Computer Methods and Programs in Biomedicine*. 2023;231:107365. doi:10.1016/j.cmpb.2023.107365

58. Matsangidou M, Liampas A, Pittara M, Pattichi CS, Zis P. Machine Learning in Pain Medicine: An Up-To-Date Systematic Review. *Pain Ther*. 2021;10(2):1067-1084. doi:10.1007/s40122-021-00324-2

59. Mari T, Henderson J, Maden M, Nevitt S, Duarte R, Fallon N. Systematic Review of the Effectiveness of Machine Learning Algorithms for Classifying Pain Intensity, Phenotype or Treatment Outcomes Using Electroencephalogram Data. *The Journal of Pain*. 2022;23(3):349-369. doi:10.1016/j.jpain.2021.07.011

60. Zamzmi G, Goldgof D, Kasturi R, Sun Y, Ashmeade T. Machine-based Multimodal Pain Assessment Tool for Infants: A Review. Published online January 16, 2019. Accessed May 8, 2024. http://arxiv.org/abs/1607.00331

61. Cheng D, Liu D, Philpotts LL, et al. Current state of science in machine learning methods for automatic infant pain evaluation using facial expression information: study protocol of a systematic review and meta-analysis. *BMJ Open*. 2019;9(12):e030482. doi:10.1136/bmjopen-2019-030482

62. Matava C, Pankiv E, Ahumada L, Weingarten B, Simpao A. Artificial intelligence, machine learning and the pediatric airway. *Pediatric Anesthesia*. 2020;30(3):264-268. doi:10.1111/pan.13792

63. Heiderich TM, Carlini LP, Buzuti LF, et al. Face-based automatic pain assessment: challenges and perspectives in neonatal intensive care units. *Jornal de Pediatria*. 2023;99(6):546-560. doi:10.1016/j.jped.2023.05.005

64. Manworren RCB, Horner S, Joseph R, Dadar P, Kaduwela N. Performance Evaluation of a Supervised Machine Learning Pain Classification Model Developed by Neonatal Nurses. *Advances in Neonatal Care*. 2024;24(3):301. doi:10.1097/ANC.0000000000001145

65. Carlini LP, Coutrin G de AS, Ferreira LA, et al. Human vs machine towards neonatal pain assessment: A comprehensive analysis of the facial features extracted by health professionals, parents, and convolutional neural networks. *Artificial Intelligence in Medicine*. 2024;147:102724. doi:10.1016/j.artmed.2023.102724

66. Ahola Kohut S, Pillai Riddell R. Does the Neonatal Facial Coding System differentiate between infants experiencing pain-related and non-pain-related distress? *J Pain*. 2009;10(2):214-220. doi:10.1016/j.jpain.2008.08.010

67. Gibbins S, Stevens B, Beyene J, Chan PC, Bagg M, Asztalos E. Pain behaviours in Extremely Low Gestational Age infants. *Early Hum Dev*. 2008;84(7):451-458. doi:10.1016/j.earlhumdev.2007.12.007

68. Zamzmi G, Hsu LY, Li W, Sachdev V, Antani S. Harnessing Machine Intelligence in Automatic Echocardiogram Analysis: Current Status, Limitations, and Future Directions. *IEEE Reviews in Biomedical Engineering*. 2021;14:181-203. doi:10.1109/RBME.2020.2988295

69. Salekin MS, Zamzmi G, Goldgof D, Kasturi R, Ho T, Sun Y. Multimodal Spatio-Temporal Deep Learning Approach for Neonatal Postoperative Pain Assessment. *Comput Biol Med*. 2021;129:104150. doi:10.1016/j.compbiomed.2020.104150

70. Zamzmi G, Goldgof D, Kasturi R, Sun Y, Ashmeade T. Machine-based Multimodal Pain Assessment Tool for Infants: A Review. Published online January 16, 2019. Accessed September 28, 2023. http://arxiv.org/abs/1607.00331

71. Zamzmi G, Pai CY, Goldgof D, Kasturi R, Ashmeade T, Sun Y. An approach for automated multimodal analysis of infants' pain. *2016 23rd International Conference on Pattern Recognition (ICPR)*. Published online December 2016:4148-4153. doi:10.1109/ICPR.2016.7900284

72. Roué JM, Morag I, Haddad WM, Gholami B, Anand KJS. Using sensor-fusion and machine-learning algorithms to assess acute pain in non-verbal infants: a study protocol. *BMJ Open*. 2021;11(1):e039292. doi:10.1136/bmjopen-2020-039292

73. Stevenson NJ, Tataranno ML, Kaminska A, et al. Reliability and accuracy of EEG interpretation for estimating age in preterm infants. *Annals of Clinical and Translational Neurology*. 2020;7(9):1564-1573. doi:10.1002/acn3.51132

74. Palmu K, Wikström S, Hippeläinen E, Boylan G, Hellström-Westas L, Vanhatalo S. Detection of "EEG bursts" in the early preterm EEG: visual vs. automated detection. *Clin Neurophysiol*. 2010;121(7):1015-1022. doi:10.1016/j.clinph.2010.02.010

75. Navarro M, Porée M, Kuchenbuch M, Chavez M, Beuchée A, Carrault G. Multi-feature classifiers for burst detection in single EEG channels from preterm infants. *Journal of Neural Engineering*. 2017;14(4):046015. doi:10.1088/1741-2552/aa714a

76. Murphy K, Stevenson NJ, Goulding RM, et al. Automated analysis of multi-channel EEG in preterm infants. *Clin Neurophysiol*. 2015;126(9):1692-1702. doi:10.1016/j.clinph.2014.11.024

77. Murphy BM, Goulding RM, O'Toole JM. Detection of Transient Bursts in the EEG of Preterm Infants using Time-Frequency Distributions and Machine Learning. *Annu Int Conf IEEE Eng Med Biol Soc*. 2020;2020:1023-1026. doi:10.1109/EMBC44109.2020.9175154

78. O'Toole JM, Boylan GB. Quantitative Preterm EEG Analysis: The Need for Caution in Using Modern Data Science Techniques. *Front Pediatr*. 2019;7:174. doi:10.3389/fped.2019.00174

79. Rakshasbhuvankar AA, Wagh D, Athikarisamy SE, et al. Inter-rater reliability of amplitude-integrated EEG for the detection of neonatal seizures. *Early Human Development*. 2020;143:105011. doi:10.1016/j.earlhumdev.2020.105011

80. Massey SL, Shou H, Clancy R, et al. Interrater and Intrarater Agreement in Neonatal Electroencephalogram Background Scoring. *Journal of Clinical Neurophysiology*. 2019;36(1):1. doi:10.1097/WNP.0000000000000534

81. Marchant S, van der Vaart M, Pillay K, et al. A machine learning artefact detection method for single-channel infant event-related potential studies. *J Neural Eng*. Published online June 26, 2024. doi:10.1088/1741-2552/ad5c04

82. O'Shea A, Ahmed R, Lightbody G, et al. Deep Learning for EEG Seizure Detection in Preterm Infants. *Int J Neural Syst*. 2021;31(8):2150008. doi:10.1142/S0129065721500088

83. Ansari AH, Cherian PJ, Caicedo Dorado A, et al. Weighted Performance Metrics for Automatic Neonatal Seizure Detection Using Multiscored EEG Data. *IEEE Journal of Biomedical and Health Informatics*. 2018;22(4):1114-1123. doi:10.1109/JBHI.2017.2750769

84. Pavel AM, Rennie JM, Vries LS de, et al. A machine-learning algorithm for neonatal seizure recognition: a multicentre, randomised, controlled trial. *The Lancet Child & Adolescent Health*. 2020;4(10):740-749. doi:10.1016/S2352-4642(20)30239-X

85. Gschwandtner L, Hartmann M, Oberdorfer L, et al. Deep learning for estimation of functional brain maturation from EEG of premature neonates. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. 2020:104-107. doi:10.1109/EMBC44109.2020.9175380

86. Ansari AH, Pillay K, Dereymaeker A, et al. A Deep Shared Multi-Scale Inception Network Enables Accurate Neonatal Quiet Sleep Detection With Limited EEG Channels. *IEEE J Biomed Health Inform*. 2022;26(3):1023-1033. doi:10.1109/JBHI.2021.3101117

87. Routier L, Querne L, Ghostine-Ramadan G, et al. Predicting the Neurodevelopmental Outcome in Extremely Preterm Newborns Using a Multimodal Prognostic Model Including Brain Function Information. *JAMA Network Open*. 2023;6(3):e231590. doi:10.1001/jamanetworkopen.2023.1590

88. Hartley C, Baxter L, Moultrie F, et al. Predicting severity of adverse cardiorespiratory effects of morphine in premature infants: a post hoc analysis of Procedural Pain in Premature Infants trial data. *British Journal of Anaesthesia*. 2021;126(4):e133-e135. doi:10.1016/j.bja.2020.10.034

89. van der Vaart M, Duff E, Raafat N, Rogers R, Hartley C, Slater R. Multimodal pain assessment improves discrimination between noxious and non-noxious stimuli in infants. *Paediatr Neonatal Pain*. 2019;1(1):21-30. doi:10.1002/pne2.12007

90. van der Vaart M, Hartley C, Baxter L, et al. Premature infants display discriminable behavioral, physiological, and brain responses to noxious and nonnoxious stimuli. *Cerebral Cortex*. 2022;32(17):3799-3815. doi:10.1093/cercor/bhab449

91. Stedman I, Racine N, Wang X, et al. Artificial Intelligence for NICU Pain Assessment: Ethical, Legal and Social Implications. In preparation.

92. Racine N, Chow C, Hamwi L, et al. Health Care Professionals' and Parents' Perspectives on the Use of AI for Pain Monitoring in the Neonatal Intensive Care Unit: Multisite Qualitative Study. *JMIR AI*. 2024;3(1):e51535. doi:10.2196/51535

Chapter 2: Machine Learning Classification of Pain Versus Non-Pain-Related Electroencephalographic Activity in Preterm Infants

Note

Chapter 2 is from a manuscript currently under review at *JAMA Pediatrics*:

Hamwi, L., Du, H., Jasim, S., Wang, X., Shah, V., Cheng, C., Fabrizi, L., Meek, J., Fitzgerald, M., Racine, N., Stedman, I., & Pillai Riddell, R*. (submitted). Machine learning classification of EEG responses of pain versus non-pain-related stimulus in preterm infants.

Abstract

Importance: Unmanaged pain in hospitalized infants has serious long-term complications. Infants cannot self-report their pain and current assessment tools have significant limitations such as feasibility and lack of specificity, leading to potential under-treatment or over-treatment of pain.

Objective: To enhance pain assessment by exploring machine learning models that autonomously pinpoints cortical activity distinguishing pain-related versus non-pain related cortical responses in preterm infants.

Design: Data for this ongoing observational study was conducted between June 2015 and May 2024. Data was collected during a routine heel lance, while mother holds her child in skin-to-skin contact. Multiple machine learning models were attempted to analyze multi-channel EEG data from 1000 milliseconds before and after the heel lance.

Setting: Data was collected in two tertiary care NICUs: Mount Sinai Hospital (Toronto, Canada) and University College London Hospital (London, United Kingdom).

Participants: The current dataset comprises 72 preterm infants (27 female), born between 24- and 36-weeks gestational age. The mean gestational age at birth was 30.11 weeks, and the mean post-menstrual age (PMA) at study was 32.87 weeks.

Main Outcome(s) and Measure(s): The primary outcome was the accuracy of various machine learning models (XGBoost, Support Vector Machines, Random Forest, Logistic Regression, Convolutional Neural Network) in distinguishing one-second post-lance EEG epochs from one-second pre-lance epochs. Performance metrics included accuracy, precision, recall, specificity,

F1-score, and AUC. The study also examined how model performance varied across different post-menstrual age groups (≤ 31 weeks, 6 days; 32 to 33 weeks, 6 days; ≥34 weeks).

Results: The performance metrics varied across PMA groups. In the Oldest PMA group, SVM, Logistic Regression, and CNN models demonstrated the highest accuracy (90%) and AUC (0.96). In the Middle PMA group, XGBoost model achieved highest accuracy of 0.88, while the Random Forest model achieved an AUC of 1.00. For the Youngest PMA group, Random Forest had the highest accuracy (73%), although the SVM and CNN models had the highest AUC (0.7).

Conclusions and Relevance: Machine learning algorithms can autonomously distinguish the one-second epoch immediately following a heel lance from the one-second epoch immediately preceding the lance in preterm infants. The study provides a foundation for developing an autonomous, feasible, and accurate tool for pain assessment in neonatal patients, potentially improving pain management practices in NICUs.

Introduction

In 2020, approximately 13.4 million newborns were born prematurely (before 37 weeks gestational age) globally[1]. Preterm infants often require prolonged care in Neonatal Intensive Care Units (NICUs), undergoing an average of 12–17 painful procedures daily in their first two weeks of life[2,3]. Frequent pain exposure can adversely affect the central nervous system, leading to structural and functional cortical changes that correlate with cognitive and motor difficulties, and poorer neurodevelopmental outcomes[4-9].

Effective pain assessment and management are essential to mitigate the immediate and long-term consequences of extended NICU stays[10]. However, assessing pain in hospitalized infants is challenging due to the inability to communicate directly[11]. Current pain assessment tools rely on behavioral (e.g., facial expressions, crying) and physiological responses (e.g., heart rate, cortisol levels)[12-14]. Despite their utility, these methods face challenges in preterm infants, including lack of specificity to pain-related distress, staffing constraints, and insufficient training[15,16]. The need for improved pain assessment has driven research into artificial intelligence technologies to enhance the accuracy and feasibility of infant pain assessment.

Machine Learning in Infant Pain Assessment

The application of machine learning (ML) in pain medicine is expanding rapidly[17-19]. In recent years, there has been a growing body of literature focusing on the use of ML for infant pain assessment that have yielded promising results in predicting pain intensity, identifying pain patterns, and classifying various types of pain [20-22]. Particularly, facial expression-centered machine learning methods have shown high accuracy in detecting pain[23-25]. However, challenges exist specifically for preterm infants that include detecting facial points obscured by medical devices and discerning subtle pain intensity nuances.

Multimodal approaches combining behavioral and physiological indicators typically integrate data from sources such as facial expressions, heart rate, respiratory rate, and oxygen saturation to create a more comprehensive assessment of pain[26-28]. However, multimodal or unimodal approaches integrating facial indicators and/or physiological indicators available from routine bedside monitors are based on indicators that lack specificity to the infant pain response. To push the field forward, researchers must examine cortical indicators. Cortical indicators hold the potential to be the first pain-specific clinical indicator for infant pain-related versus non-pain-related distress assessment[29-31].

EEG in Machine Learning

While EEG is essential for the assessment of newborn brain function, its manual interpretation can be subjective and time-consuming, however automatic analysis holds great promise to be feasible and effective, especially for preterms[32-34]. EEG-based algorithms have demonstrated satisfactory results across various non-pain metrics. Applications include artefact detection, seizure detection, sleep staging, functional brain maturation estimation, and predicting neurodevelopmental outcomes[35-47].

To date, two automated EEG-based pain assessment algorithms currently exist for neonates[48,49]. The first study included 109 infants (34 to 42 weeks' gestational age) and employed a random forest algorithm to differentiate between a clinically required heel lance and a control heel lance[48]. This multimodal model, incorporating heart rate, respiratory rate, and EEG data from the Cz electrode, achieved an accuracy of 81% and an AUC of 0.89 for all infants. For preterm infants (n = 21), the model's accuracy increased to 88% (AUC of 0.94). Assessing noxious-evoked brain activity using only peak-to-peak magnitude across a single EEG channel led to further reduced performance with EEG data alone (accuracy of 64%, AUC of 0.75). The

process required human assessment of peaks and excluded artifact-affected EEG traces, limiting real-time applicability.

The second study included 144 infants (28 to 40 weeks PMA) and used a bagged decision tree model based on noxious-evoked cerebral, limb withdrawal, cardiovascular, and facial expression responses. This model achieved an accuracy of 76% (no AUC reported) in distinguishing between noxious and non-noxious conditions in the testing set and 74% in preterm infants within the same dataset. Additionally, the model performed well in an independent dataset with an accuracy of 84% overall and 83% in preterm infants. However, brain activity was recorded using the Cz electrode only and sub-analyses across different PMA groups were limited by small sample sizes in some subgroups (n = 3 in the youngest group).

Considering these limitations, there is a need to incorporate multiple EEG features and channels for a more comprehensive analysis of pain-related cortical activity. Additionally, developing methodologies that handle EEG artifacts and utilize automated processes from the outset is crucial for enhancing accuracy and real-time applicability. Finally, integrating continuous cortical activity and conducting studies in clinical NICU preterm settings are crucial for enhancing applicability[50-52].

Current Study

The objective of the present study is to explore machine learning techniques to distinguish the EEG response associated with pain resulting from routine heel lance (blood draw via lancet to heel) and a one-second epoch preceding the heel lance without any skin-breaking stimulation in hospitalized preterm infants.

Method

To evaluate the feasibility of this approach, an observational study was conducted involving clinically required painful procedures in hospitalized preterm newborns. Data collection took place at two tertiary care NICUs: Mount Sinai Hospital (Toronto, Canada) and Elizabeth Garett Anderson Obstetric Wing, University College London Hospital (London, United Kingdom) between June 2015 and May 2024.

Participants

The dataset comprises 72 preterm infants (27 females), born between 24- and 36-weeks gestational age (see Table 1 for demographic information) at one of the two study sites. Infants were recruited only after the child was medically stable. Exclusion criteria included congenital malformations, administration of analgesics or sedatives at the time of the study, evidence of perinatal hypoxia/ischemia, hormonal imbalances, or any cardiac or neurological abnormalities. Additionally, to help ensure a non-pain related distress state pre-lance, infants with diaper rash or excoriated buttocks were not included.

Procedure

Across sites, the timing of the study was always standardized around the infants' scheduled feed time, leading up to their routine blood draw (see Figure 1). The data collection process was standardized between sites and the same trained senior neonatal research nurse at each site (who met and synchronized procedures in person) was present for every study. EEG leads were placed on the infants' scalp while in the cot or incubator. Position of infant was also always standardized. The diaper-clad infant was transitioned from the cot to skin-to-skin contact with the mother, who was seated comfortably with the child against her bare chest. Within one hour following this transfer, the infant received a routine heel lance performed by the research

nurse. The exact moment of heel lance was synchronized with the ongoing EEG recording using an accelerometer mounted on the lancet to detect the blade's vibrations. A custom-made trigger box marked the EEG at the exact moment the blade was released.

Data Processing

Signal Acquisition. Data from 10 electrodes (F7, F8, T7, T8, P7, P8, O1, O2, Cz, Pz) were recorded according to a subset of the international 10/20 electrode placement system (eFigure 1 in Supplement). A soft bonnet was placed on the scalp to secure the electrodes.

EEG signals collected in the UK were sampled at a rate of 2000 Hz and recorded using CURRY. Canadian data were sampled at a rate of 256 Hz and recorded using Moberg CNS Envision. To enhance the signal, a differential amplifier amplified the electrical activity detected by each active electrode relative to the reference electrode (Fz). The amplified signal then underwent filtering with an anti-aliasing filter to remove unwanted frequencies. The filtered signal was then converted from analog to digital format using an analog-to-digital converter. The exact timing of the lance was extracted from the accelerometer trigger annotation indicating the onset of the stimulus and was labeled “1” and whereas the other epoch was labelled as “0”.

Preprocessing. EEG data was preprocessed using the Harvard Automated Processing Pipeline for Electroencephalography *HAPPILEE* pipeline[53] (see pipeline in eFigure2 in Supplement) in MATLAB[54]. Raw data were filtered with a notch filter (48–52 Hz for data from the UK; 58–62 Hz for Canadian data) to remove power line noise. Further noise was removed using independent component analysis. Wavelet thresholding with a soft empirical Bayesian level-dependent threshold was applied using coiflet wavelets optimized for EEG data[53]. Data from the UK was down-sampled to 250Hz to match the sampling rate of the Canadian data. All EEG data was re-referenced to the common average. To reflect real-world NICU conditions,

where EEG signals often contain artifacts, only basic artifact removal was performed without thorough manual inspection. Instead, artifacts present in long-term EEG recordings were incorporated into the data-driven training process, an approach proven effective in developing ML algorithms for similar applications[41].

Feature Extraction. The most informative features of neonatal EEG signals were compiled using the NEURAL software package[55] (v0.3.1) within the MATLAB environment (MathWorks Inc.). The software generates 30 features grouped into four categories (see eTable 1 in Supplement):

1. **Amplitude**: absolute amplitude and envelope of EEG signal; range EEG;
2. **Spectral**: absolute and relative spectral power, spectral entropy (Wiener and Shannon), spectral differences, spectral edge frequency, and fractal dimension;
3. **Connectivity**: coherence, cross-correlation, and brain symmetry index;
4. **Inter-burst interval**: summary measures based on the inter-burst interval annotation.

Features were calculated at four frequency bands (FB1: 0.5–3Hz; FB2: 3–8Hz; FB3: 8–15Hz; and FB4: 15–30 Hz) over the entirety of the data in 2-second epochs with 50% overlap. This process yielded nearly 1200 features (10 channels x 4 frequency bands x 30 features) per second. To create a balanced dataset, only the second before the heel lance and the second immediately after were exported from each baby for further analysis.

Feature Selection. To reduce multicollinearity, Pearson correlation coefficients were calculated between features. For highly correlated pairs ($|r| > 0.8$), significance testing was conducted to retain the more significantly different feature between the pain and no-pain groups. Finally, a tree-based feature selection using XGBoost identified the most important features, which were then normalized using a standard scaler prior to machine learning analysis.

Classification. Multiple algorithms were evaluated for an epoch-by-epoch binary classification of non-noxious stimulus vs. noxious stimulus: support vector machines (SVM), random forests (RF), logistic regressions (LR), convolutional neural networks (CNN), and gradient boosting (XGBoost) (see Supplement for a description of each algorithm). Reduced features across epochs were grouped into a single panda DataFrame, and further analyzed in using custom scripts and libraries such as NumPy, SciPy, and Scikit-learn.

Given the rapid brain development in infants, model performances were explored by splitting the data into three datasets based on post-menstrual age (PMA) at time of study: Youngest group: PMA ≤ 31 weeks, 6 days (n = 27); Middle group: PMA 32 to 33 weeks, 6 days (n = 20); and Oldest group: PMA ≥ 34 weeks (n = 25). This stratification was important because the infant brain's structural and functional properties change exponentially early in development. Different stages of maturation might exhibit distinct EEG patterns in response to pain, thus requiring tailored models for each developmental stage to improve classification accuracy[56].

To obtain the optimal model for predicting pain presence based on extracted EEG features, a grid search was conducted with various gradient boosting hyperparameters (see Supplementary Material for the hyperparameter tuning parameters tested).

Performance Evaluation. The full dataset was randomly split into an 80/20 train/test dataset. In the training set, a ten-fold cross-validation was conducted to assess model performance and generalization. This involved dividing the data into ten equal subsets or "folds." The model was trained and evaluated ten times, in each iteration, a different fold served as the validation set, while the remaining nine folds formed the training set.

Various evaluation metrics such as accuracy, precision, recall, and specificity were calculated to evaluate the models' performance on the binary classification task (see eTable 2 in

Supplement for details). In addition to the classification metrics, a summary performance metric, the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC), was calculated to assess how well each model distinguished between classes.

Results

The most influential features identified by the XGBoost and RF models varied across the PMA groups (see Table 2 for the top features used by these models for each PMA group). Similarly, the optimal parameters identified through grid search for XGBoost varied across the PMA groups. In contrast, the key parameters used in each model for the CNN, SVM, RF, and LR classifiers were set to their default values. The detailed hyperparameter configurations for each model are provided in eTable 3 in Supplementary Materials. The performance of several machine learning models in classifying the noxious stimulus versus non-noxious stimulus in preterm infants was evaluated across three PMA groups. The classification results are summarized in Table 3.

Youngest PMA Group (≤ 31 weeks, 6 days)

The overall accuracy of predicting the true pain and true non-pain epochs ranged between 55% and 73%, with the RF model performing the best in this age group. Precision values ranged from 0.50 (CNN) to 0.75 (RF), indicating variability in correctly identifying true pain epochs from those predicted as pain. Recall, which measures the model's ability to correctly identify all actual pain epochs, was at 0.6 across all models. The highest specificity was 0.83 for the RF model, indicating its strength in correctly identifying non-pain epochs. The SVM and CNN model demonstrated the highest performance with AUCs of 0.70. This was followed by the XGBoost model, also with an AUC of 0.67. The RF and LR models showed slightly lower performance with AUCs of 0.63 and 0.60, respectively.

Middle PMA Group (32 to 33 weeks, 6 days)

The overall accuracy ranged from 75% to 88%, with the XGBoost model achieving the highest accuracy of 0.88 for the Middle PMA group. All models achieved a precision of 1.00. Recall varied, with all models achieving 0.50, except for XGBoost, which achieved a true positive rate of 0.75. Specificity was also consistently perfect across models (1.00). The RF model achieved the highest AUC value of 1.00, indicating perfect performance. The LR model followed with an AUC of 0.94. The XGBoost, SVM, and CNN models performed similarly with an AUC of 0.88.

Oldest PMA Group (≥ 34 weeks)

The overall accuracy ranged from 70% to 90%, with the SVM, LR, and CNN models all achieving the highest accuracy of 0.90. Precision values were uniformly high, with all models achieving a precision of 1.00. Recall was also high, with SVM, LR, and CNN models reaching 0.80, indicating strong performance in identifying true pain epochs. Interestingly, XGBoost's recall was significantly lower in this age group compared to the two younger PMA groups, highlighting a potential challenge in accurately identifying all true pain epochs in the oldest infants using this model. Specificity was perfect (1.00) across all models. The SVM, LR, and CNN models all achieved the highest AUC value of 0.96. The RF model had a slightly lower AUC of 0.92, while the XGBoost model had an AUC of 0.88.

Discussion

Our study demonstrated that evoked cortical responses to noxious and non-noxious stimuli can be discriminated in preterm infants, in line with previous research. First, we will discuss the overall accuracies and AUC values of multimodal models of van der Vaart and colleagues[48,49].

In the Youngest PMA group, our RF model achieved an accuracy of 73%, outperforming the existing EEG-only RF model reported by van der Vaart et al. in 2019[48], which had an accuracy of 64%. However, the AUC for our models ranged from 0.60 to 0.70, slightly lower than their reported AUC of 0.75. Despite this, our SVM model demonstrated competitive performance with an AUC of 0.70, closely matching their results. In the Middle PMA group, our RF model achieved an exceptional AUC of 1.00, surpassing the AUC of 0.89 reported by van der Vaart et al.[48]. Additionally, our LR model performed very well with an AUC of 0.94. For the Oldest PMA group, our SVM, LR, and CNN models achieved high accuracies of 90% and an AUC of 0.96, significantly outperforming the reported accuracies of 76% and an AUC of 0.82 in their multimodal models[48].

Compared to van der Vaart and colleagues' full model in 2022[49], which achieved an accuracy of 81% and an AUC of 0.90 in the test set, our models, particularly in the Middle and Oldest PMA groups, not only match but in some cases surpass the performance with high accuracies and AUC values. The comprehensive analysis using multiple EEG features and channels likely contributes to this improved accuracy and robustness in distinguishing pain-related cortical activity. Moreover, our approach takes a more autonomous approach in that no stage in data pre-processing or analysis required human judgment.

Performance across PMA groups

Thresholds for clinically acceptable standards for evaluation metrics can vary depending on the context and the specific medical decision being made[57]. In the preterm infant pain context, the primary concern is to ensure that all true pain epochs are identified, making recall, also known as the sensitivity or True Positive Rate, the most important metric to consider. Recall measures the proportion of actual pain epochs correctly identified by the model. Falsely flagging

a non-painful epoch as pain is less problematic since a graduated pain management approach can be employed in the NICU. That is, non-pharmacological methods such as handholding and kangaroo care (i.e., skin-to-skin contact), as well as sucrose and feeding, can provide significant pain relief before considering more intensive interventions like analgesics or sedatives.

Our results demonstrate a recall rate that ranges from 0.6 to 0.8. In the Youngest PMA group, all models yielded a recall of 0.6. For the Middle PMA group, XGBoost generated a recall rate of 0.75, while in the Oldest PMA group, LR, CNN, and SVM achieved a recall of 0.8. The significant differences in model performance across PMA groups underscore the importance of considering developmental stages when assessing pain in preterm infants. Infants at different PMA stages exhibit distinct EEG patterns, reflecting the ongoing development of their nervous systems[56,58]. The youngest PMA group had the lowest performance across all models and metrics, highlighting the challenges in accurately classifying pain epochs in this group. This is likely due to the heterogeneity in cortical development among infants between 28 and 32 weeks, consistent with cortical growth trajectories reported in previous literature[59].

This finding—that model performance increases with PMA—was consistent across all metrics. The accuracy of predicting a true pain versus non-pain epoch increased drastically to 88% with XGBoost in the Middle PMA group and 90% with SVM, LR, and CNN in the Oldest PMA group. The proportion of epochs predicted as pain that were actually true pain epochs (precision), as well as the models' ability to correctly identify true non-pain epochs (specificity), were perfectly classified across the models in the Middle and Oldest PMA groups.

Additionally, the AUC increased to 0.94 in the Middle PMA group and 0.96 in the Oldest PMA group. These improvements highlight the enhanced ability of the models to accurately classify pain and non-pain epochs as infants' brains develop. Considering clinical applicability,

this suggests that age-specific models implemented in ML tools may be necessary for more accurate pain assessment, as they capture a more nuanced understanding of the developmental differences within the first weeks of life.

Furthermore, the varying strengths of different models across PMA groups indicate that no single model is universally superior. For example, RF excelled in the Youngest PMA group, while XGBoost performed exceptionally well in the Middle PMA group, and SVM, LR and CNN equally showed the best performance in the Oldest PMA group. This variability highlights the need for a tailored approach, potentially combining multiple models to leverage their respective strengths.

Additionally, our comprehensive feature selection process using tree-based methods like XGBoost helps reduce overfitting and enhances model robustness, providing a reliable foundation for future research and clinical application. Furthermore, we implemented balanced training data across different PMA groups, employing 10-fold cross-validation techniques, and performing extensive feature selection using tree-based methods to mitigate potential AI-related bias and inaccuracy.

Limitations and Future Directions

While our study focused exclusively on clinically-required heel lance procedures, which limits the generalizability of our findings to other pain conditions such as chronic or post-operative pain, the preliminary classification models developed provide a solid foundation for distinguishing brain activity during skin-breaking procedures versus otherwise. Future work will include more instances of non-pain related distress epochs, such as those time-stamped when the infant was crying or receiving continuous positive airway pressure (CPAP) adjustments. This will help to enhance the comprehensiveness and applicability of our models to a broader range of

pain conditions. Our future work will also incorporate multimodal assessments, including heart rate, respiration rate, oxygen saturation, and facial behavioral coding via the Neonatal Facial Coding System (NFCS).

Conclusion

Results from this study demonstrate the feasibility of using machine learning to discriminate preterm infants' cortical responses to noxious and innocuous stimuli using EEG. By leveraging advanced machine learning techniques and ensuring a more automated and robust feature extraction process, we address several limitations found in existing studies. This is a step towards autonomous real-time pain assessments in NICUs, improving the accuracy and consistency of pain management for the most vulnerable patients.

References

1. Ohuma EO, Moller AB, Bradley E, et al. National, regional, and global estimates of preterm birth in 2020, with trends from 2010: a systematic analysis. *The Lancet*. 2023;402(10409):1261-1271. doi:10.1016/S0140-6736(23)00878-4
2. Hatfield LA, Murphy N, Karp K, Polomano RC. A Systematic Review of Behavioral and Environmental Interventions for Procedural Pain Management in Preterm Infants. *J Pediatr Nurs*. 2019;44:22-30. doi:10.1016/j.pedn.2018.10.004
3. Chawanpaiboon S, Vogel JP, Moller AB, et al. Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. *Lancet Glob Health*. 2019;7(1):e37-e46. doi:10.1016/S2214-109X(18)30451-0
4. Brummelte S, Grunau RE, Chau V, et al. Procedural pain and brain development in premature newborns. *Ann Neurol*. 2012;71(3):385-396. doi:10.1002/ana.22267
5. Vinall J, Grunau RE. Impact of repeated procedural pain-related stress in infants born very preterm. *Pediatr Res*. 2014;75(5):584-587. doi:10.1038/pr.2014.16
6. Nath N, Beltrano W, Haynes L, Dewey D, Bray S. Long-Term Effects of Preterm Birth on Children's Brain Structure: An Analysis of the Adolescent Brain Cognitive Development (ABCD) Study. *eNeuro*. 2023;10(6):ENEURO.0196-22.2023. doi:10.1523/ENEURO.0196-22.2023
7. Duerden EG, Grunau RE, Guo T, et al. Early Procedural Pain Is Associated with Regionally-Specific Alterations in Thalamic Development in Preterm Neonates. *J Neurosci*. 2018;38(4):878-886. doi:10.1523/JNEUROSCI.0867-17.2017

8. Williams MD, Lascelles BDX. Early Neonatal Pain—A Review of Clinical and Experimental Implications on Painful Conditions Later in Life. *Front Pediatr*. 2020;8. doi:10.3389/fped.2020.00030

9. Grunau RE, Miller JV, Chau CM. Long-term effects of pain in infants. *Oxford Textbook of Pediatric Pain*. Published online 2021:38.

10. Llerena A, Tran K, Choudhary D, et al. Neonatal pain assessment: Do we have the right tools? *Frontiers in Pediatrics*. 2023;10. Accessed September 6, 2023. https://www.frontiersin.org/articles/10.3389/fped.2022.1022751

11. Shiff I, Bucsea O, Pillai Riddell R. Psychosocial and Neurobiological Vulnerabilities of the Hospitalized Preterm Infant and Relevant Non-pharmacological Pain Mitigation Strategies. *Front Pediatr*. 2021;9. doi:10.3389/fped.2021.568755

12. Hatfield LA, Ely EA. Measurement of Acute Pain in Infants: A Review of Behavioral and Physiological Variables. *Biological Research For Nursing*. 2015;17(1):100-111. doi:10.1177/1099800414531448

13. Roué JM, Rioualen S, Gendras J, Misery L, Gouillou M, Sizun J. Multi-modal pain assessment: are near-infrared spectroscopy, skin conductance, salivary cortisol, physiologic parameters, and Neonatal Facial Coding System interrelated during venepuncture in healthy, term neonates? *Journal of Pain Research*. 2018;11:2257-2267. doi:10.2147/JPR.S165810

14. Stoye DQ, Boardman JP, Osmond C, et al. Saliva cortisol diurnal variation and stress responses in term and preterm infants. *Archives of Disease in Childhood - Fetal and Neonatal Edition*. 2022;107(5):558-564. doi:10.1136/archdischild-2021-321593

15. Pillai Riddell RR, Jasim S, Hamwi L. Out of the mouth of babes: a lot about pain has nothing to do with pain. *Pain*. 2022;163(Suppl 1):S117-S125. doi:10.1097/j.pain.0000000000002761

16. Boyle EM, Bradshaw J, Blake KI. Persistent pain in neonates: challenges in assessment without the aid of a clinical tool. *Acta Paediatrica*. 2018;107(1):63-67. doi:10.1111/apa.14081

17. Zhang M, Zhu L, Lin SY, et al. Using artificial intelligence to improve pain assessment and pain management: a scoping review. *Journal of the American Medical Informatics Association*. 2023;30(3):570-587. doi:10.1093/jamia/ocac231

18. Gkikas S, Tsiknakis M. Automatic assessment of pain based on deep learning methods: A systematic review. *Computer Methods and Programs in Biomedicine*. 2023;231:107365. doi:10.1016/j.cmpb.2023.107365

19. Matsangidou M, Liampas A, Pittara M, Pattichi CS, Zis P. Machine Learning in Pain Medicine: An Up-To-Date Systematic Review. *Pain Ther*. 2021;10(2):1067-1084. doi:10.1007/s40122-021-00324-2

20. Mari T, Henderson J, Maden M, Nevitt S, Duarte R, Fallon N. Systematic Review of the Effectiveness of Machine Learning Algorithms for Classifying Pain Intensity, Phenotype or Treatment Outcomes Using Electroencephalogram Data. *The Journal of Pain*. 2022;23(3):349-369. doi:10.1016/j.jpain.2021.07.011

21. Matava C, Pankiv E, Ahumada L, Weingarten B, Simpao A. Artificial intelligence, machine learning and the pediatric airway. *Pediatric Anesthesia*. 2020;30(3):264-268. doi:10.1111/pan.13792

22. Hughes JD, Chivers P, Hoti K. The Clinical Suitability of an Artificial Intelligence–Enabled Pain Assessment Tool for Use in Infants: Feasibility and Usability Evaluation Study. *Journal of Medical Internet Research*. 2023;25(1):e41992. doi:10.2196/41992

23. Heiderich TM, Carlini LP, Buzuti LF, et al. Face-based automatic pain assessment: challenges and perspectives in neonatal intensive care units. *Jornal de Pediatria*. 2023;99(6):546-560. doi:10.1016/j.jped.2023.05.005

24. Manworren RCB, Horner S, Joseph R, Dadar P, Kaduwela N. Performance Evaluation of a Supervised Machine Learning Pain Classification Model Developed by Neonatal Nurses. *Advances in Neonatal Care*. 2024;24(3):301. doi:10.1097/ANC.0000000000001145

25. Carlini LP, Coutrin G de AS, Ferreira LA, et al. Human vs machine towards neonatal pain assessment: A comprehensive analysis of the facial features extracted by health professionals, parents, and convolutional neural networks. *Artificial Intelligence in Medicine*. 2024;147:102724. doi:10.1016/j.artmed.2023.102724

26. Salekin MS, Zamzmi G, Goldgof D, Kasturi R, Ho T, Sun Y. Multimodal Spatio-Temporal Deep Learning Approach for Neonatal Postoperative Pain Assessment. *Comput Biol Med*. 2021;129:104150. doi:10.1016/j.compbiomed.2020.104150

27. Zamzmi G, Goldgof D, Kasturi R, Sun Y, Ashmeade T. Machine-based Multimodal Pain Assessment Tool for Infants: A Review. Published online January 16, 2019. Accessed September 28, 2023. http://arxiv.org/abs/1607.00331

28. Zamzmi G, Pai CY, Goldgof D, Kasturi R, Ashmeade T, Sun Y. An approach for automated multimodal analysis of infants' pain. *2016 23rd International Conference on Pattern Recognition (ICPR)*. Published online December 2016:4148-4153. doi:10.1109/ICPR.2016.7900284

29. Verriotis M, Fabrizi L, Lee A, Cooper RJ, Fitzgerald M, Meek J. Mapping Cortical Responses to Somatosensory Stimuli in Human Infants with Simultaneous Near-Infrared Spectroscopy and Event-Related Potential Recording. *eNeuro*. 2016;3(2). doi:10.1523/ENEURO.0026-16.2016

30. Shiroshita Y, Kirimoto H, Ozawa M, et al. Can Event-Related Potentials Evoked by Heel Lance Assess Pain Processing in Neonates? A Systematic Review. *Children*. 2021;8(2):58. doi:10.3390/children8020058

31. Benoit B, Martin-Misener R, Newman A, Latimer M, Campbell-Yeo M. Neurophysiological assessment of acute pain in infants: a scoping review of research methods. *Acta Paediatrica*. 2017;106(7):1053-1066. doi:10.1111/apa.13839

32. San-Segundo R, Gil-Martín M, D'Haro-Enríquez LF, Pardo JM. Classification of epileptic EEG recordings using signal transforms and convolutional neural networks. *Computers in Biology and Medicine*. 2019;109:148-158. doi:10.1016/j.compbiomed.2019.04.031

33. Maimon N, Grunau RE, Cepeda IL, et al. Electroencephalographic Activity in Response to Procedural Pain in Preterm Infants Born at 28 and 33 Weeks Gestational Age. *The Clinical Journal of Pain*. 2013;29(12):1044. doi:10.1097/AJP.0b013e318284e525

34. Stevenson NJ, Tataranno ML, Kaminska A, et al. Reliability and accuracy of EEG interpretation for estimating age in preterm infants. *Annals of Clinical and Translational Neurology*. 2020;7(9):1564-1573. doi:10.1002/acn3.51132

35. Palmu K, Wikström S, Hippeläinen E, Boylan G, Hellström-Westas L, Vanhatalo S. Detection of "EEG bursts" in the early preterm EEG: visual vs. automated detection. *Clin Neurophysiol*. 2010;121(7):1015-1022. doi:10.1016/j.clinph.2010.02.010

36. Navarro M, Porée M, Kuchenbuch M, Chavez M, Beuchée A, Carrault G. Multi-feature classifiers for burst detection in single EEG channels from preterm infants. *Journal of Neural Engineering*. 2017;14(4):046015. doi:10.1088/1741-2552/aa714a

37. Murphy K, Stevenson NJ, Goulding RM, et al. Automated analysis of multi-channel EEG in preterm infants. *Clin Neurophysiol*. 2015;126(9):1692-1702. doi:10.1016/j.clinph.2014.11.024

38. Murphy BM, Goulding RM, O'Toole JM. Detection of Transient Bursts in the EEG of Preterm Infants using Time-Frequency Distributions and Machine Learning. *Annu Int Conf IEEE Eng Med Biol Soc*. 2020;2020:1023-1026. doi:10.1109/EMBC44109.2020.9175154

39. O'Toole JM, Boylan GB. Quantitative Preterm EEG Analysis: The Need for Caution in Using Modern Data Science Techniques. *Front Pediatr*. 2019;7:174. doi:10.3389/fped.2019.00174

40. Marchant S, van der Vaart M, Pillay K, et al. A machine learning artefact detection method for single-channel infant event-related potential studies. *J Neural Eng*. Published online June 26, 2024. doi:10.1088/1741-2552/ad5c04

41. O'Shea A, Ahmed R, Lightbody G, et al. Deep Learning for EEG Seizure Detection in Preterm Infants. *Int J Neural Syst*. 2021;31(8):2150008. doi:10.1142/S0129065721500088

42. Ansari AH, Cherian PJ, Caicedo Dorado A, et al. Weighted Performance Metrics for Automatic Neonatal Seizure Detection Using Multiscored EEG Data. *IEEE Journal of Biomedical and Health Informatics*. 2018;22(4):1114-1123. doi:10.1109/JBHI.2017.2750769

43. Pavel AM, Rennie JM, Vries LS de, et al. A machine-learning algorithm for neonatal seizure recognition: a multicentre, randomised, controlled trial. *The Lancet Child & Adolescent Health*. 2020;4(10):740-749. doi:10.1016/S2352-4642(20)30239-X

44. Gschwandtner L, Hartmann M, Oberdorfer L, et al. Deep learning for estimation of functional brain maturation from EEG of premature neonates. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. 2020:104-107. doi:10.1109/EMBC44109.2020.9175380

45. Ansari AH, Pillay K, Dereymaeker A, et al. A Deep Shared Multi-Scale Inception Network Enables Accurate Neonatal Quiet Sleep Detection With Limited EEG Channels. *IEEE J Biomed Health Inform*. 2022;26(3):1023-1033. doi:10.1109/JBHI.2021.3101117

46. Routier L, Querne L, Ghostine-Ramadan G, et al. Predicting the Neurodevelopmental Outcome in Extremely Preterm Newborns Using a Multimodal Prognostic Model Including Brain Function Information. *JAMA Network Open*. 2023;6(3):e231590. doi:10.1001/jamanetworkopen.2023.1590

47. Hartley C, Baxter L, Moultrie F, et al. Predicting severity of adverse cardiorespiratory effects of morphine in premature infants: a post hoc analysis of Procedural Pain in Premature Infants trial data. *British Journal of Anaesthesia*. 2021;126(4):e133-e135. doi:10.1016/j.bja.2020.10.034

48. van der Vaart M, Duff E, Raafat N, Rogers R, Hartley C, Slater R. Multimodal pain assessment improves discrimination between noxious and non-noxious stimuli in infants. *Paediatr Neonatal Pain*. 2019;1(1):21-30. doi:10.1002/pne2.12007

49. van der Vaart M, Hartley C, Baxter L, et al. Premature infants display discriminable behavioral, physiological, and brain responses to noxious and nonnoxious stimuli. *Cerebral Cortex*. 2022;32(17):3799-3815. doi:10.1093/cercor/bhab449

50. Baud O, Arzounian D, Bourel-Ponchel E. Continuous monitoring of neonatal cortical activity: A major step forward. *Cell Reports Medicine*. 2022;3(12):100864. doi:10.1016/j.xcrm.2022.100864

51. Montazeri S, Airaksinen M, Nevalainen P, et al. An automated bedside measure for monitoring neonatal cortical activity: a supervised deep learning-based electroencephalogram classifier with external cohort validation. *The Lancet Digital Health*. 2022;4(12):e884-e892. doi:10.1016/S2589-7500(22)00196-0

52. Dilena R, Raviglione F, Cantalupo G, et al. Consensus protocol for EEG and amplitude-integrated EEG assessment and monitoring in neonates. *Clin Neurophysiol*. 2021;132(4):886-903. doi:10.1016/j.clinph.2021.01.012

53. Lopez KL, Monachino AD, Morales S, Leach SC, Bowers ME, Gabard-Durnam LJ. HAPPILEE: HAPPE In Low Electrode Electroencephalography, a standardized pre-processing software for lower density recordings. *NeuroImage*. 2022;260:119390. doi:10.1016/j.neuroimage.2022.119390

54. MathWorks, Inc. MATLAB and Simulink. 2023b.

55. O'Toole JM, Boylan GB. NEURAL: quantitative features for newborn EEG using Matlab. Published online April 19, 2017. doi:10.48550/arXiv.1704.05694

56. Myers MM, Grieve PG, Izraelit A, et al. Developmental Profiles of Infant EEG: Overlap with Transient Cortical Circuits. *Clin Neurophysiol*. 2012;123(8):1502-1511. doi:10.1016/j.clinph.2011.11.264

57. Hicks SA, Strümke I, Thambawita V, et al. On evaluation metrics for medical applications of artificial intelligence. *Sci Rep*. 2022;12(1):5979. doi:10.1038/s41598-022-09954-8

58. Stevenson N, Oberdorfer L, Koolen N, et al. Functional maturation in preterm infants measured by serial recording of cortical activity. *Scientific Reports*. 2017;7. doi:10.1038/s41598-017-13537-3

59. Garcia KE, Robinson EC, Alexopoulos D, et al. Dynamic patterns of cortical expansion during folding of the preterm human brain. *Proceedings of the National Academy of Sciences*. 2018;115(12):3156-3161. doi:10.1073/pnas.1715451115

Table 1. Demographic Data of Participants

	Site		
	UCLH, UK (*n* = 27)	Mount Sinai, Canada (*n* = 45)	Combined (*n* = 72)
Mean Gestational age at birth (weeks) (SD)	31.37 (3.88)	29.35 (1.39)	30.11 (2.77)
Range of Gestational age at birth (weeks)	24 – 36.86	27 – 32.43	24 – 36.86
Mean Postmenstrual age at time of study (weeks) (SD)	33.76 (2.72)	32.34 (1.75)	32.87 (2.26)
Range of Gestational age at birth (weeks	29.86 – 38	28.71 – 36.29	28.71 – 38
Mean Postnatal age at time of study (days) (SD)	14.89 (15.51)	21.18 (8.83)	19.13 (12.01)
Range of Postnatal age at time of study (days)	0 – 61	7 – 41	0 – 61
Sex			
Female	12	15	27
Male	17	30	47
Birth weight (g)	1700.55 (790.2)	1168.85 (213.07)	1349.08 (546.57)
Apgar score (5 min)	8.75 (1.77)	8.42 (1.15)	8.53 (1.39)

Note: The Apgar score at 5 minutes evaluates a newborn's condition based on five criteria: Appearance (skin color), Pulse (heart rate), Grimace response (reflex irritability), Activity (muscle tone), and Respiration (breathing rate and effort). Each criterion is scored between 0 and 2, with a total score ranging from 0 to 10. A score of 7-10 indicates good health.

Table 2. Important Features identified by XGBoost and Random Forest

PMA Group	Features
Youngest	T7_FB1_spectral_flatness T8_FB1_amplitude_skew F7_FB1_spectral_flatness F8_FB3_spectral_flatness _FB3_connectivity_coh_freqmax T7_FB2_spectral_flatness T7_FB1_spectral_entropy F7_FB1_spectral_entropy
Middle	T8_FB2_spectral_flatness P8_FB1_spectral_flatness T7_FB3_spectral_flatness T7_FB3_spectral_entropy T8_FB2_spectral_entropy F7_FB2_spectral_relative_power
Oldest	T8_FB2_spectral_flatness T7_FB3_amplitude_skew O2_FB3_spectral_flatness O1_FB4_amplitude_skew P7_FB2_spectral_entropy P7_FB2_spectral_flatness O2_FB1_amplitude_skew O2_FB1_amplitude_kurtosis O2_FB3_spectral_entropy

Note: The feature names follow the format: Channel_FrequencyBand_FeatureType. For example, "T8_FB2_spectral_flatness" indicates that the feature was extracted from the T8 channel. "FB2" corresponds to the frequency band from 3 to 8 Hz, and "spectral_flatness" is the type of feature extracted from the NEURAL toolbox over the one-second epoch.

Table 3. Summary of Metrics of Binary Classification Across Preterm Infants PMA groups

PMA Group	Model	Accuracy	Precision	Recall	Specificity	F1-score	AUC
Youngest	XGBoost	0.64	0.6	0.6	0.67	0.60	0.67
	SVM	0.64	0.6	0.6	0.67	0.6	0.7
	RF	0.73	0.75	0.6	0.83	0.67	0.63
	LR	0.64	0.6	0.6	0.67	0.6	0.6
	CNN	0.55	0.5	0.6	0.67	0.55	0.7
Middle	XGBoost	0.88	1	0.75	1	0.86	0.88
	SVM	0.75	1	0.5	1	0.67	0.88
	RF	0.75	1	0.5	1	0.67	1
	LR	0.75	1	0.5	1	0.67	0.94
	CNN	0.75	1	0.5	1	0.67	0.88
Oldest	XGBoost	0.7	1	0.4	1	0.57	0.88
	SVM	0.9	1	0.8	1	0.89	0.96
	RF	0.8	1	0.6	1	0.75	0.92
	LR	0.9	1	0.8	1	0.89	0.96
	CNN	0.9	1	0.8	1	0.89	0.96

Note: Models examined: XGBoost, Support Vector Machines (SVM), Random Forest (RF), Logistic Regression (LR) and Convolutional Neural Networks (CNN). Youngest group: infants with a post-menstrual age (PMA) of 31 weeks, 6 days or less at the time of the study (n = 27). Middle group: infants with a PMA between 32 weeks and 33 weeks, 6 days at the time of the study (n = 20). Oldest group: infants with a PMA of 34 weeks or greater at the time of the study (n = 25). The dark grey shading highlights the model that achieved the best overall performance across each metric within a PMA group.

Figure 1: Overview of Current Study Procedure
The study involves EEG placement, transfer to skin-to-skin contact (STS), a clinical procedure (heel lance), and post-lance STS resting. The blue segment represents one second of non-noxious-related brain activity, while the red segment indicates one second of noxious-evoked brain activity immediately post-lance.



Supplementary Material

NEURAL Toolbox – Features

eTable 1
List of Features Generated by NEURAL Toolbox

Feature Category	Subfeature – MATLAB Code Name	Description
Amplitude	amplitude_total_power	time-domain signal: total power
	amplitude_SD	time-domain signal: standard deviation
	amplitude_skew	time-domain signal: skewness
	amplitude_kurtosis	time-domain signal: kurtosis
	amplitude_env_mean	envelope: mean value
	amplitude_env_SD	envelope: standard deviation
	rEEG_mean	range EEG: mean
	rEEG_median	range EEG: median
	rEEG_lower_margin	range EEG: lower margin (5th percentile)
	rEEG_upper_margin	range EEG: upper margin (95th percentile)
	rEEG_width	range EEG: upper margin - lower margin
	rEEG_SD	range EEG: standard deviation
	rEEG_CV	range EEG: coefficient of variation
	rEEG_asymmetry	range EEG: measure of skew about median
Spectral	spectral_power	spectral power: absolute

	spectral_relative_power	spectral power: relative (normalised to total spectral power)
	spectral_flatness	spectral entropy: Wiener (measure of spectral flatness)
	spectral_entropy	spectral entropy: Shannon
	spectral_diff	difference between consecutive short-time spectral estimates
	spectral_edge_frequency	spectral edge frequency: 95% of spectral power contained between 0.5 and fc Hz (cut-off frequency)
	FD	fractal dimension
Connectivity	connectivity_BSI	brain symmetry index
	connectivity_corr	correlation (Pearson) between envelopes of hemisphere-paired channels
	connectivity_coh_mean	coherence: mean value
	connectivity_coh_max	coherence: maximum value
	connectivity_coh_freqmax	coherence: frequency of maximum value
Inter-Burst	IBI_length_max	burst annotation: maximum (95th percentile) inter-burst interval
	IBI_length_median	burst annotation: median inter-burst interval
	IBI_burst_prc	burst annotation: burst percentage

IBI_burst_number	burst annotation: number of bursts

Performance Metrics

Confusion Matrix		**Actual**	
		Pain	**Non-pain**
Predicted	**Pain**	True positive (TP)	False positive (FP)
	Non-pain	False negative (FN)	True negative (TN)

eTable 2
Evaluation Metrics for Pain Epoch Classification

Metric	Equation	Definition
Accuracy	$\frac{TP + FN}{TP + TN + FP + FN}$	Out of all the epochs, how many did the model correctly identify as either pain or non-pain?
Precision	$\frac{TP}{TP + FP}$	Out of all the epochs predicted as pain epochs by the model, how many were actually true pain epochs?
Recall	$\frac{TP}{TP + FN}$	Out of all the actual pain epochs, how many did the model correctly identify as pain epochs?
Specificity	$\frac{TP}{TN + FP}$	Out of all the actual non-pain epochs, how many did the model correctly identify as non-pain epochs?
F1 score	$2 \times \frac{Precision \times Recall}{Precision + Recall}$	Performance metric that balances the trade-off between precision (not mislabeling non-pain epochs as pain) and recall (not missing any pain epochs).
AUC	Calculated from the ROC curve	The AUC measures the area under the ROC curve, which plots the True Positive Rate (Recall) against the False Positive Rate (1 - Specificity) at various threshold settings. The AUC represents the model's ability to distinguish between pain and non-pain epochs.

Machine Learning Models Tested

- **Support Vector Machines (SVM)**: SVMs are effective at finding the optimal hyperplane that separates classes in high-dimensional space, making them suitable for EEG data classification with complex patterns.
- **Gradient Boosting (XGBoost)**: XGBoost is a highly efficient and flexible boosting algorithm that sequentially improves model accuracy by correcting errors from previous iterations, making it effective for complex datasets.
- **Random Forests (RF)**: RFs are robust ensemble methods that reduce overfitting by averaging the results of multiple decision trees, which improves the generalization of the model to new data.
- **Logistic Regressions (LR)**: Logistic regression is a straightforward yet powerful algorithm for binary classification that provides probabilities for class membership, making it a good baseline for comparison.
- **Convolutional Neural Networks (CNN)**: CNNs excel at capturing spatial and temporal patterns in raw EEG data through convolutional layers, which can learn intricate features indicative of pain-related brain activity.

XGBoost Hypertuning Parameters

- Number of boosting rounds ('n_estimators'): [300, 500, 700, 1000, 1500, 2000, 2500]
- Learning rate ('learning_rate'): [0.1, 0.3, 0.5, 1.0] (step size shrinkage used to prevent overfitting)
- Subsample ('subsample'): [0.6, 0.8, 1.0] (fraction of samples used for training each tree)
- Maximum depth ('max_depth'): [3, 6, 9, 10, 12, 15] (maximum depth of a tree)
- Fraction of features used for splitting nodes ('colsample_bynode'): [0.5, 0.8]
- Minimum sum of instance weight needed in a child ('min_child_weight'): [1, 3, 5]

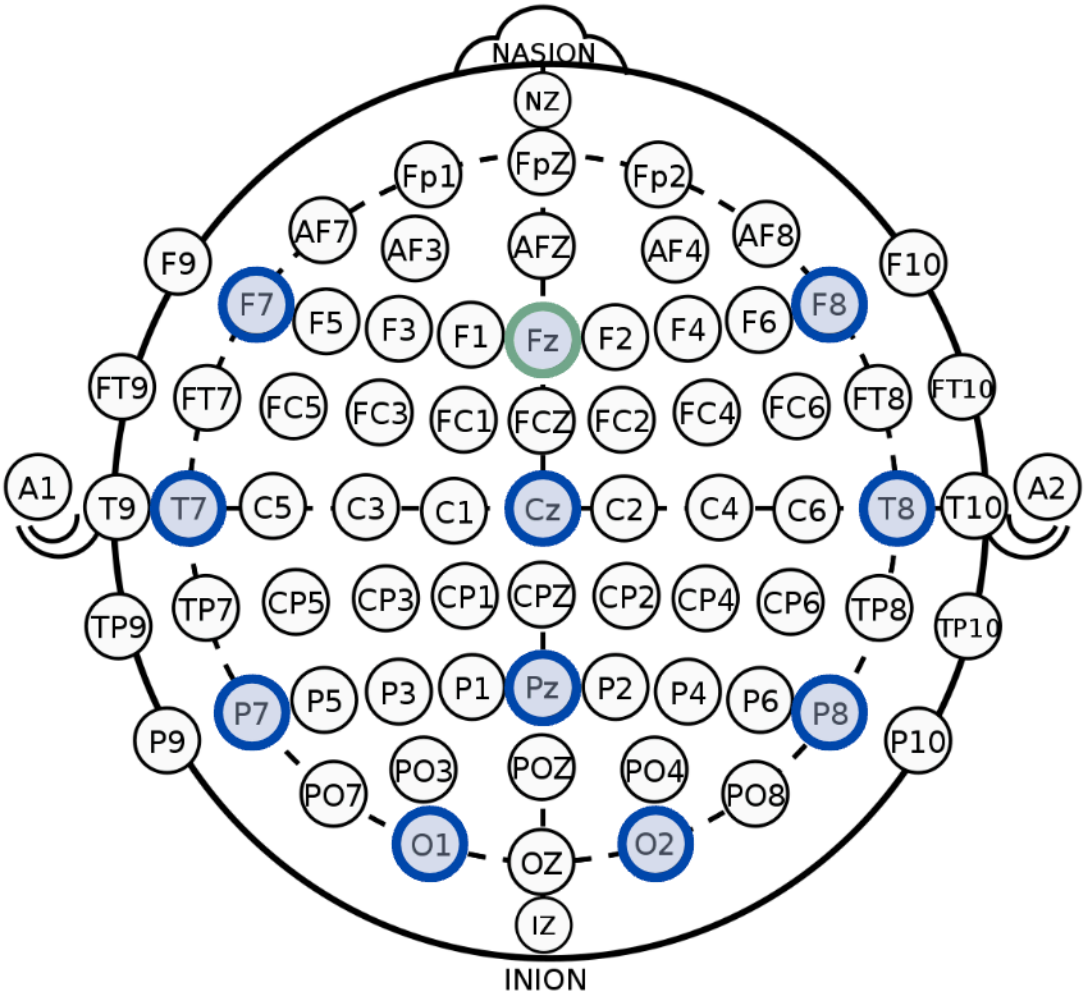
eTable 3

Final Model Parameters for Different Gestational Age Groups

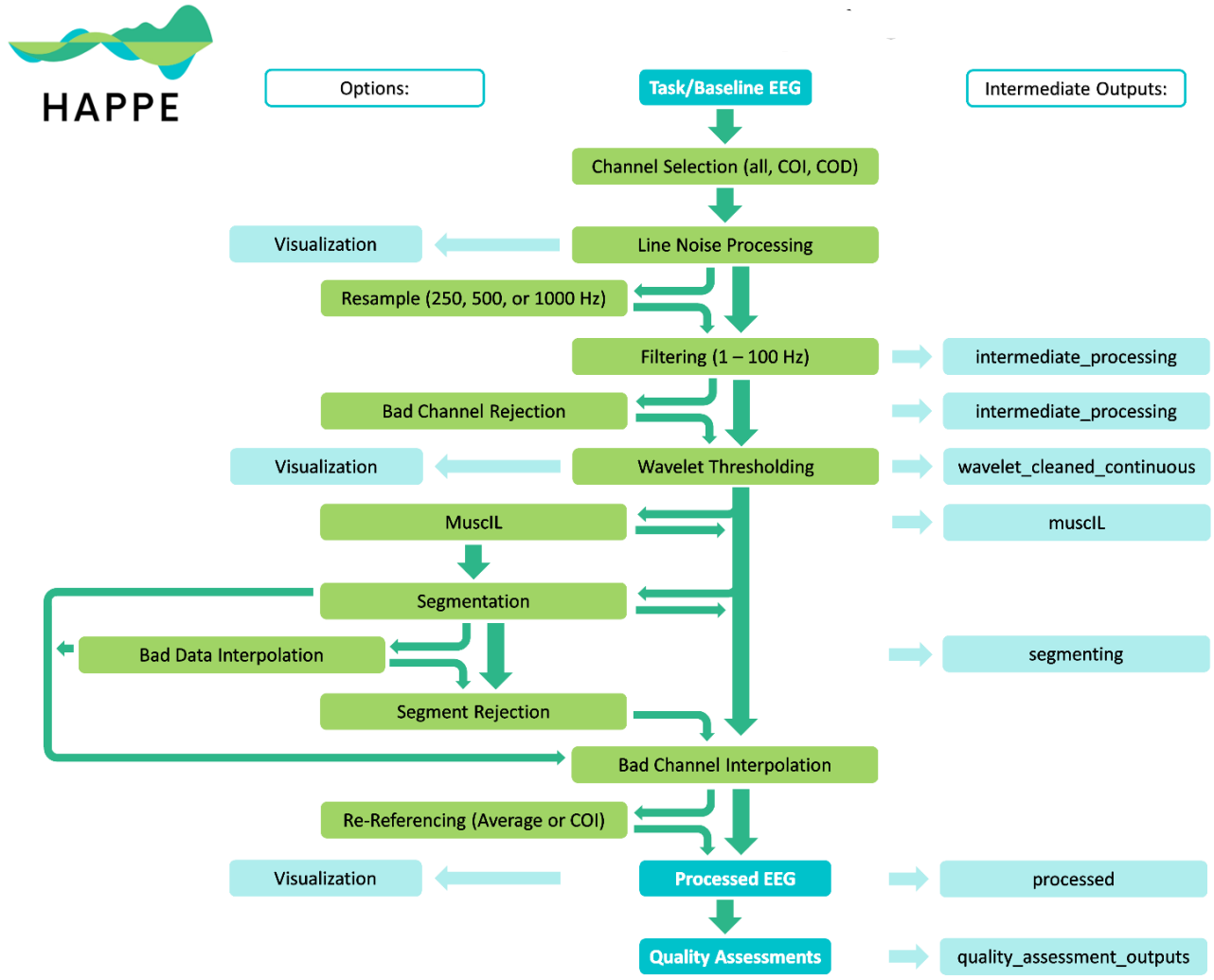
Model	Parameters	Youngest PMA	Middle PMA	Oldest PMA
XGBoost	n_estimators	500	300	500
	learning_rate	0.3	0.1	0.3
	subsample	0.6	0.6	0.8
	max_depth	3	3	3
	colsample_bynode	0.8	0.8	0.5
	min_child_weight	3	1	3
SVM	kernel	rbf	rbf	rbf
	C	1	1	1
	gamma	0.002	0.002	0.002
Random Forest	n_estimators	500	500	500
	max_depth	20	20	20
	min_samples_split	2	2	2
	min_samples_leaf	1	1	1
	max_features	sqrt	sqrt	sqrt
	bootstrap	True	True	True
Logistic Regression	max_iter	1000	1000	1000
CNN	input shape	(number of features, 1)	(number of features, 1)	(number of features, 1)
	Conv1D filters	64	64	64
	Kernel size	2	2	2
	Activation	ReLU	ReLU	ReLU
	MaxPooling1D	True	True	True
	Dense units	50	50	50
	Output units	2	2	2
	Output activation	Softmax	Softmax	Softmax

Optimizer	Adam	Adam	Adam
Loss function	Categorical Crossentropy	Categorical Crossentropy	Categorical Crossentropy
Metrics	Accuracy	Accuracy	Accuracy
Epochs	20	20	20

eFigure 1
Electrode Montage



eFigure 2
Harvard Automated Processing Pipeline for Electroencephalography HAPPILEE Preprocessing Pipeline



Chapter 3: Concluding Notes

In this thesis, an EEG-feature based machine learning algorithm was proposed for identifying the one-second epoch containing the noxious stimulus in preterm infants. The proposed models utilized features extracted from the entire EEG recording where the two one-second epochs around the lance were selected for further analysis. Common features were then combined across infants and sites to predict the final pain assessment with the sample divided into three PMA age groups. We compared the performance of five machine learning algorithms and found that different models achieved varying accuracies across the three age groups. The experimental results suggest that machine learning approaches are reliable for distinguishing EEG features associated with pain from a tissue-breaking procedure (i.e., heel lance) and non-noxious stimuli (i.e., resting in skin-to-skin) in a real-world clinical environment. These results are consistent with the reported performance in the two existing neonatal pain studies[1,2].

Our proposed methodology leverages a unique, multicentre, multinational dataset that includes EEG data from a highly challenging population in which to collect data. While there is always a potential for AI-related bias and inaccuracy, this study implemented several strategies to identify, manage, and mitigate these issues. Bias mitigation strategies included ensuring balanced training data across different PMA groups, implementing 10-fold cross-validation techniques, and performing extensive feature selection using tree-based methods like XGBoost to avoid overfitting and enhance model robustness.

Limitations and Future Directions

Given that this study focused exclusively on participants undergoing clinically-required heel lance procedures, there are inherent limitations in generalizing our findings to other pain conditions, such as chronic pain, non-skin-breaking pain or post-operative pain. Nonetheless, the preliminary classification models developed provide a foundation for assessing pain during skin-

breaking procedures using EEG signals. Future work will include more instances of non-pain-related distress epochs, such as those time-stamped when the infant was crying, receiving CPAP adjustments, or being handled. This will help to enhance the strength, utility, comprehensiveness and applicability of our models to a broader range of pain conditions.

Moreover, our dataset initially contained a large number of features, necessitating the use of tree-based feature selection to reduce their number. While effective, this method is computationally intensive and demands significant time and resources to train the models and determine feature importance. Tree-based methods, such as those employed in algorithms like Random Forest and XGBoost, capture interactions between features, thereby enhancing model performance. However, this complexity also poses a challenge in understanding and interpreting the individual importance of each feature. Clinical utility of the features detected by the machine learning algorithm are low.

Finally, considering the improved performance of multimodal models in non-machine learning and machine learning contexts, our future work will incorporate assessments including heart rate, respiration rate, oxygen saturation, and facial behavioral coding via the Neonatal Facial Coding System (NFCS) and continue to build a larger dataset

Summary

In summary, we demonstrate that the use of machine learning in the assessment of preterm infant pain makes it possible to discriminate preterm infants' cortical responses to noxious and innocuous stimuli using EEG. The accuracy of classifying epochs between a noxious and non-noxious stimulus was used as a marker to evaluate the feasibility of employing machine learning in preterm infant pain assessment. By leveraging advanced machine learning techniques and ensuring a more automated and robust feature extraction process, we address

several limitations found in existing studies. The proposed approach has significant clinical implications and can greatly enhance the current assessment practice, which is discontinuous, highly dependent on the observer's experience and subjectivity, and often limited by the lack of medical resources. Our multinational research initiative aims to revolutionize infant pain assessment by constructing an autonomous infant pain assessment tool specifically designed to identify pain responses that may not be easily observed through behavioral indicators in NICU infants. This is a step towards autonomous, real-time pain assessments in NICUs, improving the accuracy and consistency of pain management for the most vulnerable patients.

References

1. van der Vaart M, Duff E, Raafat N, Rogers R, Hartley C, Slater R. Multimodal pain assessment improves discrimination between noxious and non-noxious stimuli in infants. *Paediatr Neonatal Pain*. 2019;1(1):21-30. doi:10.1002/pne2.12007
2. van der Vaart M, Hartley C, Baxter L, et al. Premature infants display discriminable behavioral, physiological, and brain responses to noxious and non noxious stimuli. *Cerebral Cortex*. 2022;32(17):3799-3815. doi:10.1093/cercor/bhab449