An electroencephalography-based sleep index and supervised machine learning as a suitable tool for automated sleep classification in children

van Twist, Eris; Hiemstra, Floor W.; Cramer, Arnout B.G.; Verbruggen, Sascha C.A.T.; Tax, David M.J.; Joosten, Koen; Louter, Maartje; Straver, Dirk C.G.; de Hoog, Matthijs; More Authors

DOI
10.5664/jcsm.10880

Publication date
2024

Document Version
Accepted author manuscript

Published in
Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine

Citation (APA)
van Twist, E., Hiemstra, F. W., Cramer, A. B. G., Verbruggen, S. C. A. T., Tax, D. M. J., Joosten, K., Louter, M., Straver, D. C. G., de Hoog, M., & More Authors (2024). An electroencephalography-based sleep index and supervised machine learning as a suitable tool for automated sleep classification in children. *Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine*, *20*(3), 389-397. https://doi.org/10.5664/jcsm.10880

An EEG-based sleep index and supervised machine learning as a suitable tool for automated sleep classification in children

Authors: drs. Eris van Twist[1], drs. Floor W. Hiemstra[2,3], drs. Arnout B.G. Cramer[1], dr. Sascha C.A.T. Verbruggen[1], MD PhD, dr.ir. David M.J. Tax[4], PhD, prof.dr Koen Joosten[1], MD PhD, dr. Maartje Louter[5], MD PhD, dr. Dirk C.G. Straver[5], MD PhD, prof.dr. Matthijs de Hoog[1], MD PhD, dr. Jan Willem Kuiper[1], MD PhD, dr. Rogier C.J. de Jonge[1], MD PhD.

[1] Department of Neonatal and Pediatric Intensive Care, Division of Pediatric Intensive Care, , Erasmus MC Sophia Children's Hospital, Rotterdam, The Netherlands

[2] Department of Intensive Care, Leiden University Medical Centre, Leiden, The Netherlands

[3] Laboratory for Neurophysiology, Department of Cellular and Chemical Biology, Leiden University Medical Center, Leiden, the Netherlands

[4] Pattern Recognition Laboratory, Delft University of Technology, Delft, The Netherlands

[5] Division of Clinical Neurophysiology, Department of Neurology, Erasmus MC, Rotterdam, The Netherlands

CORRESPONDING AUTHOR: Eris van Twist

Department of Neonatal and Pediatric Intensive Care, Erasmus MC Sophia Children's Hospital

Wytemaweg 80, 3015 CN Rotterdam, The Netherlands

Email: e.vantwist@erasmusmc.nl | Phone: +31631773366

INSTUTION WHERE WORK WAS PERFORMED: Erasmus MC Sophia Children's Hospital,

AUTHOR APPROVAL: All authors have read and approved the manuscript.

DECLARATIONS: All authors declare no conflicts of interest, no financial support and no off-label or investigational use.

CLINICAL TRIAL: This manuscript does not report on a clinical trial.

NUMBER OF TABLES: 4

NUMBER OF FIGURES: 1

ABSTRACT WORD COUNT: 250

BRIEF SUMMARY WORD COUNT: 115

MANUSCRIPT WORD COUNT: 4565

Brief summary

Current knowledge/study rationale: Sleep disruption is common in the paediatric intensive care unit, contributing to significant morbidity and prolonged length of stay. Currently, individual sleeping patterns cannot be optimized in critical care as objective real-time sleep classification at the bedside is not possible.

Study Impact: The present study proposes a sleep index based on the gamma to delta power ratio, derived from single-channel EEG and combined with artefact detection, smoothing, machine learning and nested-cross validation. The sleep index can perform two and three state classification at the bedside with high balanced accuracy in hospitalised, non-critically ill children from six months old to adolescence, while remaining objective, easy to interpret and generalizable to multiple EEG channels.

Abstract

Introduction Although sleep is frequently disrupted in the pediatric intensive care unit (PICU), it's currently not possible to perform real-time sleep monitoring at the bedside. In this study, spectral band powers of electroencephalography (EEG) data are used to derive a simple index for sleep classification.

Method Retrospective study at Erasmus MC Sophia Children's Hospital, using hospital-based polysomnography (PSG) recordings obtained in non-critically ill children between 2017 and 2021. Six age categories were defined: 6-12 months, 1-3 years, 3-5 years, 5-9 years, 9-13 years and 13-18 years. Candidate index measures were derived by calculating spectral band powers in different frequent frequency bands of smoothed EEG. With the best performing index, sleep classification models were developed for two, three and four states via decision tree and five-fold nested-cross validation. Model performance was assessed across age categories and EEG channels.

Results In total 90 patients with PSG were included, with a mean (standard deviation) recording length of 10.3 (1.1) hours. The best performance was obtained with the gamma to delta spectral power ratio (gamma:delta-ratio) of the F4-A1 and F3-A1 channels with smoothing. Balanced accuracy was 0.88, 0.74 and 0.57 for two, three and four state classification. Across age categories, balanced accuracy ranged between 0.83 – 0.92 and 0.72 – 0.77 for two and three state classification, respectively.

Conclusion We propose an interpretable and generalizable sleep index derived from single-channel-EEG for automated sleep monitoring at the bedside in non-critically ill children aged 6 months to 18 years, with good performance for two and three state classification.

Keywords: machine learning, sleep stage, sleep classification, pediatric intensive care unit, polysomnography

Introduction

Sleep is essential for overall health and development, and specifically for the recovery of critically ill children.[1-3] Nonetheless, sleep deprivation and sleep fragmentation are common in the paediatric intensive care unit (PICU), due to various risk factors of pharmacological, pathological and environmental nature.[4-6] Sleep deprivation is associated with significant morbidity and may prolong hospital stay.[7-9] However, sleep is not routinely monitored in the PICU.

The gold standard to monitor sleep and classify sleep stages is overnight, hospital-based Level 1 polysomnography (PSG) testing.[10] PSG consists of a multitude of measurement modalities, including electroencephalography (EEG), electromyography (EMG) and electro-oculography (EOG). These measurements capture the distinctive patterns in electrical brain activity, eye movement and muscle tone that the five sleep-wake stages exhibit.[10] Sleep classification is performed manually in accordance with the American Academy of Sleep Medicine (AASM) criteria.[11] PSG-based sleep classification is a laborious and invasive procedure, subject to interrater variation and typically done in retrospect.[10-13] As such, PSG is currently unsuitable for real-time, bedside sleep monitoring in critically ill children. Real-time sleep monitoring is desirable to optimize individual sleeping patterns, through informed decision-making and with interventions that contribute to sleep quality and quantity.

Several attempts have been made to develop an automated sleep classification algorithm, mainly in neonates and adults, but they are sparsely implemented in clinical practice.[14,15] EEG poses a suitable signal for automated sleep classification, considering its distinctive frequency waveforms and spectral band powers across different sleep stages.[16-18] Low frequency bands are common during non-REM 3 (N3) sleep, also known as slow-wave-sleep (SWS), while high frequency bands are common during wake and non-SWS (NSWS) including rapid eye movement (REM), non-REM 1 (N1) and non-REM 2 (N2) sleep.[19,20] A promising development based on EEG spectral band powers is the intensive care unit depth of sleep (IDOS) index, defined as the ratio between spectral band powers of a single EEG channel.[21] The IDOS is easy to interpret, generalizable, and has shown considerable agreement with

PSG-based sleep classification in adults.[21] However, this method has not been used and validated for critically ill children admitted to the PICU. In critically ill children, there is the additional challenge of developmental changes in both EEG and sleep patterns with increasing age, as well as the influence of morbidity and hospital admission on EEG spectra.[22]

In a first step towards sleep monitoring in the PICU, the aim of the present study was to develop a method for automated sleep classification in non-critically ill children aged 6 months and older, based on spectral band powers of a single EEG-channel. We further aimed to evaluate this method for different sleep classifications and assess its generalizability over various age categories and EEG-channels.

Methods

Study population

This retrospective study was conducted at the Erasmus MC Sophia Children's Hospital (Rotterdam, The Netherlands). PSG recordings were anonymously obtained from a database of recordings performed in non-critically ill children who underwent an overnight, hospital-based PSG for diagnostic and follow-up purposes between May 2017 to June 2021. The hospital registry of patients objecting to data usage for research purposes was consulted prior to inclusion of the patients' data. This study was approved by the internal Medical Ethics Committee (MEC) of Erasmus MC (MEC-2021-0121). To take into account the developmental changes in the sleep EEG, six age categories were defined that globally correspond to the EEG changes during development: 6-12 months, 1-3 years, 3-5 years, 5-9 years, 9-13 years, 13-18 years.[23] We did not include children under 6 months of age, as their EEG is markedly different and they do not exhibit all subtypes of NREM sleep yet.[24] For patients born preterm (<37 weeks gestational age), age was corrected until the postnatal age of two years. Fifteen recordings were randomly collected for each age category, resulting in a total of 90 recordings. PSG recordings were included if the PSG showed normal physiological sleep with presence of all sleep stages and without atypical EEG findings. Atypical EEG findings were reported in the PSG report and included epileptiform activity, polymorphic delta activity, absence of sleep spindles and K-complexes, burst suppression and isoelectric activity. PSG recordings obtained from patients with neuromuscular disease (e.g. myotonic dystrophy) and from patients who received sedative or analgesic medication were excluded. PSG recordings were also excluded if the hypnogram or PSG recording was incomplete or data quality was low due to the presence of numerous or long-lasting artefacts (> 30 min) due to e.g. movement or interference from electrical equipment.

Data acquisition

All PSG recordings were hospital-based and were performed overnight.

The PSGs were performed with a commercially available device (BrainRT, OSG, Rumst, Belgium or Morpheus, Micromed Sp.A., Treviso, Italy) using an eight-channel EEG and two-channel EOG and

EMG. EEG electrodes included the frontal (F3, F4), central (C3, C4), occipital (O1, O2) and auricular (A1, A2) electrodes and were placed according to the international 10-20 system with Ag/AgCl electrodes, sharing the same electrode as reference (Fz).[25] All possible bipolar channels (n = 28) were calculated by subtraction of the signals of each pair of EEG electrodes. The EMG electrodes were placed on the submental muscles and the EOG electrodes were placed on the right and left outer canthus (ROC and LOC) of the eye, with the ROC electrode one centimetre superior and LOC one centimetre inferior of the outer canthus. The ROC-LOC channel was derived from the two EOG signals. EEG, EOG and EMG signals were sampled at 250 Hz or 256 Hz, depending on the PSG device used. All PSG recordings were divided into 30-second epochs and visually scored by experienced PSG technicians according to the AASM criteria.[11] Scored PSG recordings were finally evaluated by an experienced clinical neurophysiologist. The raw PSG signals together with the visually scored hypnogram were manually exported from the PSG software environment BrainRT (OSG, Rumst, Belgium). Signal analysis was performed in Python (3.9.) using EEGlib (0.4), PyEDFlib (0.1.20), Skicit-learn (0.24.0), Scipy packages (1.6.1).[26-29]

Classification tasks

We evaluated classification performance for three different classification tasks: two-state, three-state and four-state classification. The term state is used to refer to the non-conventional sleep stages explained here. Two-state classification only concerned the differentiation between sleep and wake. In three-state classification, two sleep stages were distinguished: NSWS, by combining the conventional sleep labels N1, N2 and REM sleep and SWS (N3). In four-state classification, REM sleep was considered as a separate sleep stage in addition to wake, NWSW and SWS.

Preprocessing

All PSG recordings were divided into 30 second epochs. A simple artefact detection algorithm was used to identify and label epochs that contained significant artefacts in the PSG signals. Epochs with average absolute signal amplitude exceeding a predefined threshold (most often movement or 50-Hz electrical interference artefacts) or zero amplitude (impedance measurement artefact) were detected and removed

from the dataset. Next, a 16th order Butterworth band-pass filter was used for each PSG signal for additional artefact reduction by removing irrelevant frequencies. All EEG signals were filtered between 0.5-48 Hz. Spectral band powers in different frequency bands, i.e. delta (0.5 – 4 Hz), theta (4 – 8 Hz), alpha (8 – 12 Hz), beta (12 – 20 Hz) and gamma (30 – 80 Hz), were calculated according to Welch's method.[16,30] We used a discrete short-term Fourier transform on a 2-second Hann window with 50% overlap. Following filtering, we explored our data and assessed which spectral power ratios had the potential to be used in an index measure. A detailed description of this process is available in Supplemental Methods 1: Data Exploration.

Smoothing

We tested whether smoothing the indices with a moving average filter increased classification performance, since this technique allows the use of information of surrounding 30 s epochs too. The moving average was calculated using a geometric mean, as histograms of index data showed right skew (Supplemental Figure S1). Smoothing levels of 1, 2, 3, 4, 6, 8, 10, 12, 16, 20, 30 and 60 epochs were tested to assess potential positive and negative effects.

Sleep index

In order to assign sleep stages to epochs based on the value of an index measure, we used a decision tree. More information about decision trees is available in Supplemental Methods 2: Decision Tree. The decision tree was applied to data containing a single variable, the index measure and we set the maximum number of leave nodes (the final classification) equal to the number of sleep states (2, 3 or 4). We used the Gini impurity as the criterion to decide where to make the split, and balanced class weights in order to maximize balanced accuracy.

Model development

Because the brain activity, and thus EEG signal, varies within the brain, the spectral composition of the EEG signal might also vary between the various EEG channels. Therefore, classification performance of the sleep indices was evaluated for each EEG channel with nested cross validation (details described

in Supplemental Methods 3: Nested Cross Validation). Nested cross-validation was performed with five folds on every combination of EEG channel and index measure, with smoothing as a hyperparameter. With the optimal channel and index combination we determined the optimal level of smoothing. The decision tree was then fitted to the full dataset of the best performing channel index combination and the optimal smoothing level in order to create a final model for comparison with PSG-based sleep classification.

Model evaluation

Final models were visually evaluated based on their decision trees and confusion matrices, as compared to the validated hypnograms derived from PSG-based scoring. Classification performance was assessed with balanced accuracy obtained with five-fold cross validation. During this process, data was split into five sets, four for training and one for testing, stratified by participant. In each fold data was resampled to exhaust all possible combinations for training and testing. This process accounts for imbalanced data and equally considers model sensitivity and specificity. The relative measure of agreement was assessed with Cohen's Kappa. Final model performance was also evaluated on other channels and across age categories. Lastly, to evaluate the performance of sleep classification for clinically relevant parameters, we calculated several sleep quality measures for each subject and compared these to the same measures derived from manually scored PSG. The parameters are listed and defined in Supplemental Table S1. Descriptive statistics were reported as count (percentage), mean (standard deviation (SD)) or median (first quartile, third quartile (Q1, Q3)). A two-sided p-value < 0.05 was regarded as statistically significant and 95% confidence intervals (CI) were reported where applicable.

Results

Patient and data characteristics

We included 90 patients, of which 47 (52.2%) were male. Five (5.6%) were diagnosed with epilepsy, 26 (28.8%) with neurocognitive impairment and 24 (26.7%) with a hereditary syndrome. With a total of 111,076 epochs, the mean recording length was 10.3 (1.1) hours. The mean total sleep time was 8.0

(1.4) hours. A complete overview of the distribution of epochs over different sleep stages is provided in Table 1. Impedance artefacts were present in 204 (0.18%) epochs, while high amplitude artefacts were exclusively present in 117,036 (13.2%) epochs of EEG electrodes relative to the reference electrode.

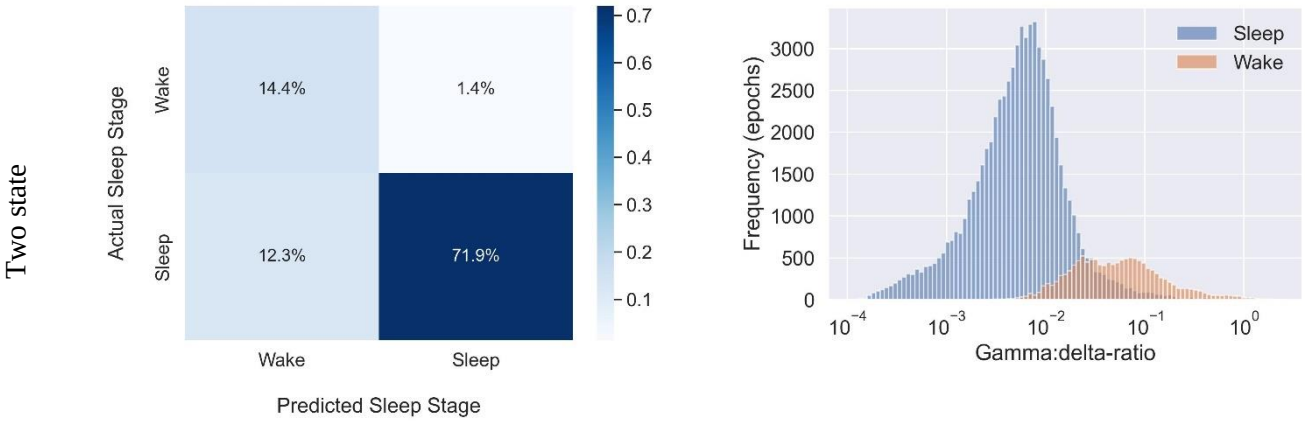
Model development

The highest rank correlations with sleep classification tasks were obtained with relative gamma, beta and delta power, as shown in Supplemental Table S2. These spectral band powers were combined into nine candidate index measures, presented in Supplemental Table S3. For two state classification, outer cross validation balanced accuracy ranged from 0.62 for relative gamma power on the O1-O2 channel to 0.88 for the gamma to delta power ratio (gamma:delta-ratio) on the F4-A1 channel. For three state classification, balanced accuracy ranged from 0.51 for relative beta power on the O1-O2 channel to 0.74 for the gamma:delta-ratio on the F3-A1 channel. For four state classification, the worst and best channel-index combinations were again relative beta power on the O1-O2 channel and the gamma:delta-ratio on the F3-A1 channel, respectively, with accuracies ranging from 0.38 to 0.57. The fifteen best channel-index combinations are shown in Table 2 for each classification task. Notice that relative gamma power, relative delta power and the gamma:delta-ratio are the only index measures featuring in this table, and that the frontal (F) channels are included in 39 of the 45 combinations. While auricular (A) and central (C) channels are also frequently present in the top fifteen, ocular (O) channels are not present at all. Next, the effect of smoothing was assessed for three state classification. The balanced accuracy for each level of smoothing is shown in Supplemental Figure S2. On average, the best performance was obtained for a smoothing window of 8 epochs, while the worst performance was obtained for a smoothing window of 60 epochs, although this was on average only 0.03 lower than for 8 epochs.

Final models

The best performances were obtained with the gamma:delta-ratio on either the F4-A1 or F3-A1 channels. Optimal smoothing was 8 epochs for two-, 10 epochs for three- and 6 epochs for four-state classification. Balanced accuracy was 0.88, 0.74 and 0.57 for two, three and four state classification, respectively. Final model results for two, three and four state classification are displayed in Table 3.

Confusion matrices comparing model-based classification with PSG-based classification and corresponding histograms of the gamma:delta-ratio for all classification tasks are provided in Figure 2. Confusion matrices for two and three state classification show no irregularities. However, the confusion matrix of the four state model showed that NSWS was not predicted for any epoch. In the corresponding decision tree, two leave nodes were classified as wake and none as NSWS (Supplemental Figure S3). The histogram of the gamma:delta-ratio for each of the four states showed great overlap between REM and NSWS (Figure 2). A post-hoc analysis was performed to resolve this issue, as described in Supplemental Results 1: Four State Model. However, this did not improve the model and we will therefore discard four state classification from further evaluation. Results of sleep quality measures for three state classification are presented in Supplemental Table S4. Both smoothed and unsmoothed models underestimated total sleep time and sleep efficiency. Differences in root mean squared error (RMSE) between smoothed and unsmoothed models were small, except for the number of awakenings and the mathematically related sleep fragmentation index, where the unsmoothed and smoothed models respectively estimated a mean of 51.5 and 9.7 awakenings per night, versus the true mean of 11.3 awakenings per night.



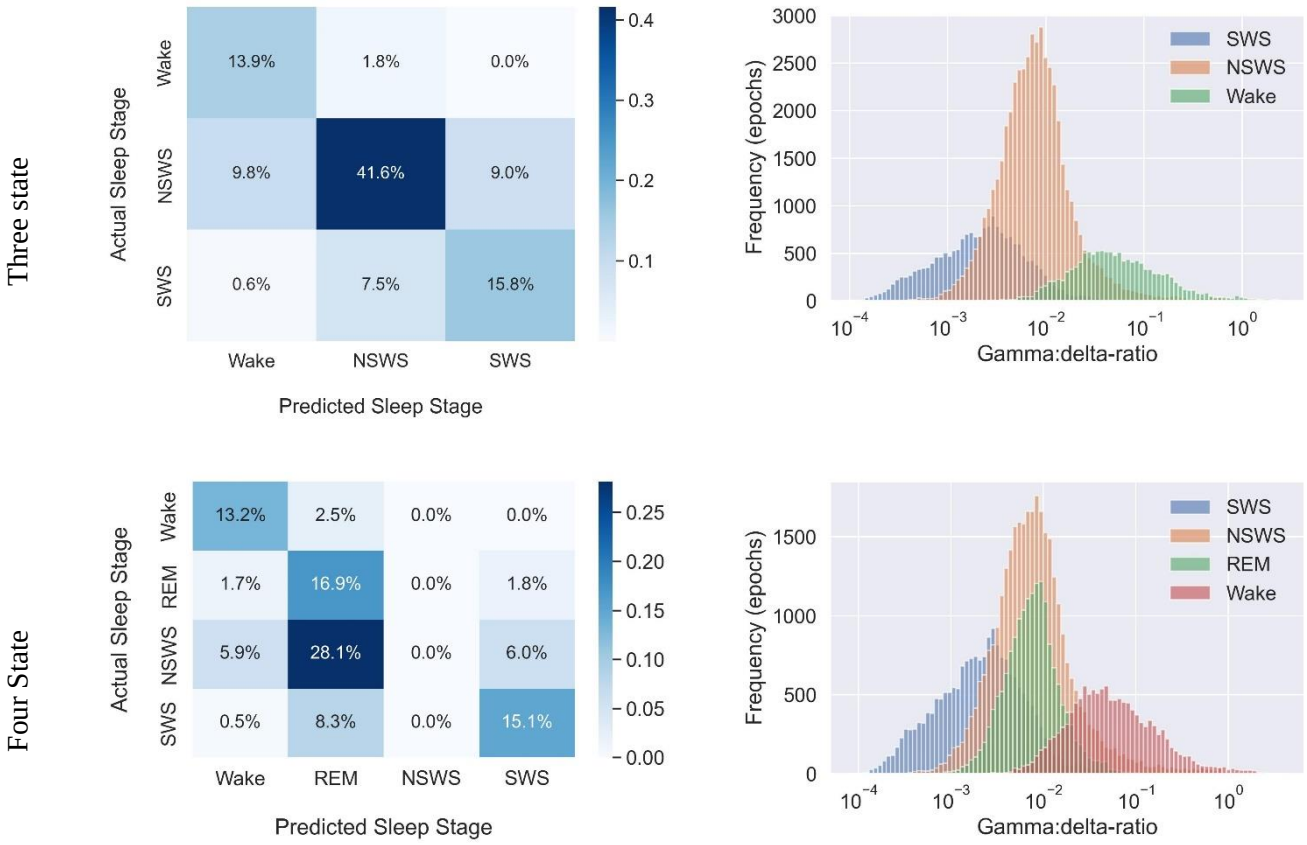


Figure 2A-F. Comparison of model-based and PSG-based sleep classification with confusion matrices and corresponding histograms of the gamma:delta-ratio of the two, three and four state models

Figures 2A, 2C and 2E (left column) contain the confusion matrices with the predicted sleep stage (model-based classification) on the x-axis and the actual sleep stage (PSG-based classification) on the y-axis for two to four states in ascending row number. Each cell shows the percentage of all epochs with the corresponding predicted and true sleep stage. Note that in four state classification no NSWS is predicted, and that actual NSWS epochs are mostly classified as REM sleep. Figures 2B, 2D and 2F (right column) contain corresponding histograms depicting the distribution of the gamma:delta-ratio over different classes. Note that in four state classification ranges of the gamma:delta-ratio for REM and NSWS largely overlap. REM = rapid eye movement, NSWS = non-slow wave sleep, SWS = slow wave sleep.

Age stratification

Across age categories, balanced accuracy ranged from 0.83 to 0.92 for two state classification and 0.72 to 0.77 for three state classification, as illustrated in Table 4. The highest classification accuracies are observed in the age categories 1- 3 years for two state and 5 – 9 years for three state, respectively.

Channel generalization

The accuracy of the final models across EEG channels is shown in Supplemental Figure S4. Balanced accuracy ranged from 0.66 to 0.88 while and from 0.53 to 0.75 for two and three state classification, respectively. Both models obtained the best performance across F-C and F-A channels.

Discussion

We aimed to develop an interpretable, single-EEG-channel sleep index for automated sleep classification in non-critically ill children aged 6 months to 18 years. We created several index measures based on individual spectral band powers that showed correlation with sleep classification. The highest classification accuracy was achieved using the ratio between the spectral gamma and delta powers of the F4-A1 and F3-A1 EEG channel, with smoothing levels of 8 and 10 epochs for two and three state classification, respectively. For four state classification the model failed to make a distinction between REM and NSWS, likely due to excessive overlap in gamma:delta-ratio between these stages. Smoothing slightly increased classification accuracy and markedly increased the accuracy of derived clinically relevant sleep parameters such as the number of awakenings. Classification accuracy was further influenced by age and by the EEG channel used to derive the sleep index.

To our knowledge this is the first study on automated sleep classification with spectral power ratios in a pediatric age-stratified cohort of non-critically ill children between 6 months and 18 years admitted to the hospital. However, in healthy adults Ganesan et al. also performed two state sleep classification using the gamma:delta-ratio, obtaining lower performance than observed here (Cohen's Kappa = 0.57, accuracy = 0.78).[31] Similarly, Reinke et al. performed depth of sleep monitoring and sleep classification using the gamma:delta-ratio in a small cohort of adult ICU patients and healthy controls, reporting both high agreement with manual classification and high inter-observer agreement (Kappa = 0.84 and 0.82 for two and three state classification, respectively).[21] The high performance may be attributable to the use of individual thresholds to determine transitions between sleep stages, however, this cannot be applied to newly admitted patients where manual classification has not been performed yet.[21] Meanwhile, the sleep index proposed in the present study is widely applicable with high balanced accuracy and moderate agreement with manual classification. The latter may be the result of imbalanced data, as Kappa tends to be low as compared to balanced accuracy.[32] However, since our objective is not to substitute a trained sleep scorer, substantial agreement comparable to the interrater agreement of manual classification (Kappa = 0.68) would already be sufficient.[12] Alternative algorithms for automated

sleep classification in non-critically ill children include pattern recognition, neural networks, long-short-term memory systems and adversarial learning.[33-37] Two algorithms may even be ready for clinical implementation.[38,39] Phan et al. applied an ensemble of deep learning methods for five state classification (Kappa = 0.85, accuracy = 0.89) in 1216 children (aged 5 – 9 years) experiencing obstructive sleep apnea.[38] In our study, we only observe similar accuracies for two state classification in the same age category. Nonetheless, they conclude that novel research should focus on interpretability, as attempted here with the sleep index. Vallet et al. developed an open source and ready-to-use algorithm for five state classification, requiring single channels of EEG, EMG and EOG, with high performance (Kappa = 0.82, accuracy = 0.84) in a cohort of 3163 healthy and non-critically ill children and adults (aged 5 – 89 years).[39] However, although several algorithms have been reported with higher performances compared to our sleep index, most are far more complex with limited interpretability and generalizability, have not been applied in the full age range from 6 months to 18 years and require more than two electrodes which hampers bedside feasibility and may even affect sleep itself. Finally, some research has explored alternative signals for automated sleep classification, such as respiratory rate and photoplethysmography, with varying performances (Kappa = 0.43 – 0.85, balanced accuracy = 0.58 – 0.90) depending on age, classification task and health status of participants.[40-42] Nonetheless, a combination of the proposed sleep index with additional physiological signals may improve accuracy.

The spectral gamma to delta power ratio was found to be the best performing index for sleep classification, regardless of classification task. Data exploration showed that correlations between absolute spectral band powers and classification task are moderate at best. This observation likely stems from the nature of the EEG as a highly dynamic non-linear signal with varying spectral characteristics throughout sleep.[16] Hence, EEG categorization is a highly complex task, prone to subjectivity. When awake, the EEG is characterized by high gamma power, albeit partly due to prominent muscle activity.[43] During sleep however, the EEG exhibits a characteristic shift in power spectrum from fast to slow wave activity, represented by the relative increase in delta power.[16,18] Te proposed sleep index captures these temporal changes in EEG associated with depth of sleep, whilst remaining objective and interpretable.

The proposed sleep index is promising for continuous two and three state sleep classification at the bedside. This allows optimization of individual sleeping patterns by aiding clinicians in informed decision making about medication administration, weaning protocols and other interventions that may disrupt sleep. To our knowledge, there are no examples of the prior in current clinical practice. Two state classification (i.e. sleep/wake monitoring) may especially be beneficial in patients where this is not visually obvious, for example in heavily sedated patients. Three state classification provides additional information with the distinction between deep sleep (SWS) and light sleep (NSWS). This may be beneficial at the bedside considering the restorative and memory consolidative function of deep sleep, which should be preserved in critical care.[18] Both the two and three state model generalized well to frontal, auricular and central electrodes. As such, there are sufficient possibilities to perform single channel EEG-measurement for two and three state classification if the optimal F3-A1 and F4-A1 channels are unavailable, e.g. due to head trauma or interference with other devices. Model performance for four state classification was insufficient, due to the inability to distinguish REM sleep from NSWS. This is due to similarity in the EEG signal, which is why in manual classification EMG and EOG are relatively more important to make this distinction.[10] While we used balanced class weights and assessed the final models with balanced accuracy, it's worth noting that the time spent in REM sleep in the included population is remarkably low compared to existing literature, possibly even insufficient for adequate training of REM detection.[44,45] Since REM sleep is important for brain health and specifically neurocognitive function, the ability to classify REM directly at the bedside is desirable in the future.

The present study has several strengths worth mentioning. Firstly, we used high quality data and applied a robust method consisting of artefact detection, smoothing and nested-cross validation. Artefacts are common in the PICU and can hamper signal quality.[46] Smoothing can provide additional artefact reduction and also allows to take surrounding epochs into consideration during classification.[47] Unsurprisingly, smoothed models obtained better performance in this study. While other studies have performed regular cross-validation, with notable high accuracies (> 0.83), our study is unique with nested cross-validation using smoothing as a hyperparameter to attain the highest achievable performance without overfitting.[33,35,39,48] However, some limitations of the present study also need to be

addressed. Firstly, we did not compare performance of our models to visual scorer agreement. Furthermore, the generalizability of our models to clinical practice is not without question, due to exclusion of PSGs with numerous or long-lasting artefacts, not taking possible comorbidity of included participants into consideration and not performing external validation.

With this study we make an essential first step in real-time bedside sleep monitoring in non-critically ill children using single-channel EEG only. Our findings implicate that spectral index measures of sleep, specifically the gamma to delta power ratio, are suitable for two and three state sleep classification in non-critically ill children across all age categories. Considering that bedside real-time sleep monitoring is currently not feasible at all, two or three state classification is already sufficient to improve individual sleeping patterns. The simplicity of the technique minimizes the effect on patient and caregivers. Future research should focus on evaluating sleep indices in an external population of non-critically ill children, and ultimately in critically ill children admitted to the PICU. The latter may pose a challenge, as research has suggested that critically ill children may experience different sleeping patterns compared to non-critically children.[49] (Cramer ABG et al, unpublished work, 2023). Interestingly, Kudchadkar et al. observed a clear distinction between healthy children and age-matched critically ill children based on nocturnal delta power alone.[20] Furthermore, research should evaluate what additional parameters that are routinely monitored in the PICU, such as vital signs, may be added to optimize sleep classification performance. With a working algorithm for sleep classification in critically ill children, the door is open to conduct large-scale research into sleep in the PICU and the effects of sleep-enhancing interventions on clinical outcome.

Conclusion

Real-time sleep monitoring and classification at the bedside is essential to optimize individual sleeping patterns in the PICU. In our study, we show that the sleep index calculated from the gamma:delta spectral power-ratio of a single EEG-channel is able to perform two and three state sleep classification in non-critically ill children aged 6 months to 18 years.

Tables

Table 1. Distribution of epochs over sleep stages of all PSG recordings

Sleep stage	Mean (SD) %[1]	Number of epochs
Wake	16.5 (11.6)	n = 17,067
REM	18.1 (6.3)	n = 18,963
N1	9.8 (5.4)	n = 10,354
N2	26.3 (10.3)	n = 27,099
N3	29.2 (11.3)	n = 30,167

[1] Calculated as percentage of all epochs. PSG = polysomnography; REM = rapid eye movement

Table 2. The 15 best performing channel-index combinations with the corresponding balanced accuracy

	Two state			Three state			Four state		
Rank	Channel	Index	Balanced accuracy	Channel	Index	balanced accuracy	Channel	Index	balanced accuracy
1	F4-A1	γ/δ	0.88 (0.75 – 1.00)	F3-A1	γ/δ	0.74 (0.62 – 0.87)	F3-A1	γ/δ	0.57 (0.45 – 0.70)
2	F3-A1	γ/δ	0.88 (0.75 – 1.00)	F4-A1	γ/δ	0.74 (0.62 – 0.87)	F4-C3	γ	0.57 (0.44 – 0.69)
3	F4-A2	δ	0.87 (0.75 – 1.00)	F4-C3	γ/δ	0.74 (0.61 – 0.86)	F4-C3	γ/δ	0.57 (0.44 – 0.69)
4	F3-A1	γ	0.87 (0.75 – 0.99)	F4-C3	γ	0.74 (0.61 – 0.86)	F4-C4	γ/δ	0.57 (0.44 – 0.69)
5	F4-A1	γ	0.87 (0.74 – 0.99)	F3-A2	γ/δ	0.73 (0.61 – 0.86)	F3-C3	γ/δ	0.57 (0.44 – 0.69)
6	F4-A2	γ	0.87 (0.74 – 0.99)	F3-A1	γ	0.73 (0.61 – 0.85)	F4-A1	γ/δ	0.57 (0.44 – 0.69)
7	C3-A1	γ/δ	0.86 (0.74 – 0.98)	F3-C4	γ/δ	0.73 (0.60 – 0.85)	F3-A2	γ/δ	0.57 (0.44 – 0.69)
8	F3-A2	γ/δ	0.86 (0.73 – 0.98)	F4-A2	γ/δ	0.73 (0.60 – 0.85)	F3-C4	γ/δ	0.56 (0.44 – 0.69)
9	C4-A2	γ/δ	0.86 (0.73 – 0.98)	F4-C4	γ/δ	0.73 (0.60 – 0.85)	F4-A2	γ/δ	0.56 (0.44 – 0.69)
10	F3-A2	γ	0.85 (0.73 – 0.98)	F4-C4	γ	0.73 (0.60 – 0.85)	F3-F4	γ	0.56 (0.44 – 0.68)
11	C4-A1	γ/δ	0.85 (0.73 – 0.97)	F3-C3	γ/δ	0.72 (0.60 – 0.84)	F3-C4	γ	0.56 (0.43 – 0.68)
12	C4-A1	γ	0.85 (0.72 – 0.97)	F4-A1	γ	0.72 (0.60 – 0.84)	F4-A1	γ	0.56 (0.43 – 0.68)
13	C3-A1	γ	0.85 (0.72 – 0.97)	F3-C3	γ	0.72 (0.60 – 0.84)	F3-C3	γ	0.56 (0.43 – 0.68)
14	C4-A2	γ	0.85 (0.72 – 0.97)	F4-A2	γ	0.72 (0.60 – 0.84)	F4-A2	γ	0.56 (0.43 – 0.68)
15	F4-C3	γ/δ	0.85 (0.72 – 0.97)	F4-A1	γ	0.72 (0.60 – 0.84)	F4-A1	γ	0.56 (0.43 – 0.68)

Balanced accuracy is the mean (95 CI%) of outer cross validation. Two state: Wake-Sleep; Three state: Wake-NSWS-SWS; Four state: Wake-REM-NSWS-SWS

F: frontal channel; A: auricular channel; C: central channel; γ = relative gamma power; γ/δ = gamma:delta-ratio; Fz = ground electrode.

Table 3. Final models

	Two state	Three state	Four state
Channel	F4-A1	F3-A1	F3-A1
Index	Gamma:delta-ratio	Gamma:delta-ratio	Gamma:delta-ratio
Smoothing	8 epochs	10 epochs	6 epochs
Cohen's Kappa	0.60 (0.59 – 0.60)	0.52 (0.51 – 0.52)	0.31 (0.31 – 0.32)
Balanced accuracy	0.88 (0.88 – 0.88)[1]	0.74 (0.74 – 0.75)	0.58 (0.57 – 0.58)

[1] 0.883 (0.881 – 0.884).

Balanced accuracy and Cohen's Kappa are the mean of 5-fold cross-validation. Two state: Wake-Sleep; Three state: Wake-NSWS-SWS; Four state: Wake-REM-NSWS-SWS.

381

Table 4. Balanced accuracy across age categories for two and three state classification

Age category	Two state	Three state
6 – 12 months	0.89 (0.89 – 0.89)	0.75 (0.74 – 0.76)
1 – 3 years	0.92 (0.92 – 0.92)	0.75 (0.74 – 0.76)
3 – 5 years	0.91 (0.90 – 0.91)	0.74 (0.73 – 0.74)
5 – 9 years	0.89 (0.89 – 0.90)	0.77 (0.76 – 0.78)
9 – 13 years	0.83 (0.82 – 0.83)	0.72 (0.71 – 0.72)
13 – 18 years	0.83 (0.71 – 0.95)	0.75 (0.75 – 0.76)

Balanced accuracy is the mean (95 CI%) of 5-fold cross-validation. Two state: Wake-Sleep; Three state: Wake-NSWS-SWS.

382

383

384

Abbreviations

A1; A2	Auricular electrodes
AASM	American Academy of Sleep Medicine
C3; C4	Central electrodes
CI	Confidence intervals
EEG	Electroencephalography
EMG	Electromyography
EOG	Electro-oculography
F3; F4	Frontal electrodes
Gamma:delta-ratio	Spectral gamma to delta power ratio
IDOS	Intensive care unit depth of sleep
LOC	Left outer canthus
MEC	Medical Ethics Committee
NSWS	Non-slow wave sleep
O1; O2	Occipital electrodes
PICU	Pediatric intensive care unit
PSG	Polysomnography
REM	Rapid eye movement
ROC	Right outer canthus
SD	Standard deviation
SWS	Slow wave sleep

Acknowledgements

First and foremost, we would like to express our sincere gratitude to all participants included in this study, as well as their parents and/or legal guardians. We further wish to thank all sleep technicians for their work. Finally, we thank Laurens Reinke, who inspired and encouraged us to conduct this study and who kindly shared his insights on sleep in the intensive care unit with us.

References

1. Schwartz JR, Roth T. Neurophysiology of sleep and wakefulness: basic science and clinical implications. *Curr Neuropharmacol*. Dec 2008;6(4):367-78. doi:10.2174/157015908787386050
2. Morrison AL. Implications of sleep deprivation for children in the pediatric intensive care unit. *Pediatr Crit Care Med*. May 2004;5(3):291-2. doi:00130478-200405000-00018 [pii] 10.1097/01.pcc.0000124012.43269.97
3. Kudchadkar SR, Aljohani OA, Punjabi NM. Sleep of critically ill children in the pediatric intensive care unit: a systematic review. *Sleep Med Rev*. Apr 2014;18(2):103-10. doi:10.1016/j.smrv.2013.02.002
4. Stremler R, Micsinszki S, Adams S, Parshuram C, Pullenayegum E, Weiss SK. Objective Sleep Characteristics and Factors Associated With Sleep Duration and Waking During Pediatric Hospitalization. *JAMA Netw Open*. Apr 1 2021;4(4):e213924. doi:2778081 [pii] zoi210143 [pii] 10.1001/jamanetworkopen.2021.3924
5. Cooper AB, Thornley KS, Young GB, Slutsky AS, Stewart TE, Hanly PJ. Sleep in critically ill patients requiring mechanical ventilation. *Chest*. Mar 2000;117(3):809-18. doi:S0012-3692(15)32743-4 [pii] 10.1378/chest.117.3.809
6. Cureton-Lane RA, Fontaine DK. Sleep in the pediatric ICU: an empirical investigation. *Am J Crit Care*. Jan 1997;6(1):56-63.
7. Weinhouse GL, Schwab RJ, Watson PL, et al. Bench-to-bedside review: delirium in ICU patients - importance of sleep deprivation. *Crit Care*. 2009;13(6):234. doi:cc8131 [pii] 10.1186/cc8131
8. Schrimpf M, Liegl G, Boeckle M, Leitner A, Geisler P, Pieh C. The effect of sleep deprivation on pain perception in healthy subjects: a meta-analysis. *Sleep Med*. Nov 2015;16(11):1313-1320. doi:S1389-9457(15)00896-5 [pii] 10.1016/j.sleep.2015.07.022
9. Ruiz FS, Andersen ML, Martins RC, Zager A, Lopes JD, Tufik S. Immune alterations after selective rapid eye movement or total sleep deprivation in healthy male volunteers. *Innate Immun*. Feb 2012;18(1):44-54. doi:1753425910385962 [pii] 10.1177/1753425910385962
10. Rundo JV, Downey R, 3rd. Polysomnography. *Handb Clin Neurol*. 2019;160:381-392. doi:B978-0-444-64032-1.00025-4 [pii] 10.1016/B978-0-444-64032-1.00025-4
11. Berry RQS, Abreu A. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, version 2.6. *American Academy of Sleep Medicine*. 2020;
12. Danker-Hopfe H, Kunz D, Gruber G, et al. Interrater reliability between scorers from eight European sleep laboratories in subjects with different sleep disorders. *J Sleep Res*. Mar 2004;13(1):63-9. doi:10.1046/j.1365-2869.2003.00375.x
13. Duce B, Rego C, Milosavljevic J, Hukins C. The AASM recommended and acceptable EEG montages are comparable for the staging of sleep and scoring of EEG arousals. *J Clin Sleep Med*. Jul 15 2014;10(7):803-9. doi:10.5664/jcsm.3880
14. Ambrogio C, Koebnick J, Quan SF, Ranieri M, Parthasarathy S. Assessment of sleep in ventilator-supported critically III patients. *Sleep*. Nov 2008;31(11):1559-68. doi:10.1093/sleep/31.11.1559
15. Ghimatgar H, Kazemi K, Helfroush MS, et al. Neonatal EEG sleep stage classification based on deep learning and HMM. *J Neural Eng*. Jun 25 2020;17(3):036031. doi:10.1088/1741-2552/ab965a

16. Nayak CS AA. *EEG Normal Waveforms.* StatPearls. Treasure Island (FL): StatPearls Publishing; 2023.
17. da Silveira TL, Kozakevicius AJ, Rodrigues CR. Single-channel EEG sleep stage classification based on a streamlined set of statistical features in wavelet domain. *Med Biol Eng Comput*. Feb 2017;55(2):343-352. doi:10.1007/s11517-016-1519-4 [pii]
10.1007/s11517-016-1519-4
18. Patel AK RV, Shumway KR et al. *Physiology, Sleep Stages*. StatPearls; 2023.
19. Mann K, Bäcker P, Röschke J. Dynamical properties of the sleep EEG in different frequency bands. *Int J Neurosci*. Dec 1993;73(3-4):161-9.
20. Kudchadkar SR, Yaster M, Punjabi AN, et al. Temporal Characteristics of the Sleep EEG Power Spectrum in Critically Ill Children. *J Clin Sleep Med*. Dec 15 2015;11(12):1449-54.
21. Reinke L, van der Hoeven JH, van Putten MJ, Dieperink W, Tulleken JE. Intensive care unit depth of sleep: proof of concept of a simple electroencephalography index in the non-sedated. *Crit Care*. Apr 09 2014;18(2):R66. doi:10.1186/cc13823
22. Carno MA, Connolly HV. Sleep and sedation in the pediatric intensive care unit. *Crit Care Nurs Clin North Am*. Sep 2005;17(3):239-44. doi:10.1016/j.ccell.2005.04.005
23. Dereymaeker A, Pillay K, Vervisch J, et al. Review of sleep-EEG in preterm and term neonates. *Early Hum Dev*. Oct 2017;113:87-103. doi:S0378-3782(17)30325-0 [pii]
10.1016/j.earlhumdev.2017.07.003
24. de Weerd AW, van den Bossche RA. The development of sleep during the first months of life. *Sleep Med Rev*. Apr 2003;7(2):179-91.
25. Klem GH, Lüders HO, Jasper HH, Elger C. The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. *Electroencephalogr Clin Neurophysiol Suppl*. 1999;52:3-6.
26. Cabañero-Gomez L, Hervas R, Gonzalez I, Rodriguez-Benitez L. eeglib: A Python module for EEG feature extraction. *SoftwareX*. 2021/07/01/ 2021;15:100745. doi:https://doi.org/10.1016/j.softx.2021.100745
27. Nahrstaedt H. PyEDFlib: Python library for EDF. Github. Accessed 30 June 2023. https://github.com/holgern/pyedflib
28. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*. 2020/03/01 2020;17(3):261-272. doi:10.1038/s41592-019-0686-2
29. Pedregosa F. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011;12(85):2825−2830.
30. Welch P. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*. 1967;15(2):70-73. doi:10.1109/tau.1967.1161901
31. Ganesan RA, Jain R. Sleep-awake Classification using EEG Band-power-ratios and Complexity Measures. 2020:1-6.
32. De Diego IM, Redondo AR, Fernández RR, Navarro J, Moguerza JM. General Performance Score for classification problems. *Applied Intelligence*. 2022/08/01 2022;52(10):12049-12063. doi:10.1007/s10489-021-03041-7
33. Zhu L, Wang C, He Z, Zhang Y. A lightweight automatic sleep staging method for children using single-channel EEG based on edge artificial intelligence. *World Wide Web*. 2022;25(5):1883-1903. doi:983 [pii]
10.1007/s11280-021-00983-3
34. Estevez PA, Held CM, Holzmann CA, et al. Polysomnographic pattern recognition for automated classification of sleep-waking states in infants. *Med Biol Eng Comput*. Jan 2002;40(1):105-13. doi:10.1007/BF02347703
35. Wang H, Lin G, Li Y, et al. Automatic Sleep Stage Classification of Children with Sleep-Disordered Breathing Using the Modularized Network. *Nat Sci Sleep*. 2021;13:2101-2112. doi:336344 [pii]
10.2147/NSS.S336344

36. Li Y, Peng C, Zhang Y, Zhang Y, Lo B. Adversarial learning for semi-supervised pediatric sleep staging with single-EEG channel. *Methods*. Aug 2022;204:84-91. doi:S1046-2023(22)00080-9 [pii]
10.1016/j.ymeth.2022.03.013
37. Held CM, Heiss JE, Estevez PA, et al. Extracting fuzzy rules from polysomnographic recordings for infant sleep classification. *IEEE Trans Biomed Eng*. Oct 2006;53(10):1954-62. doi:10.1109/TBME.2006.881798
38. Phan H, Mertins A, Baumert M. Pediatric Automatic Sleep Staging: A Comparative Study of State-of-the-Art Deep Learning Methods. *IEEE Trans Biomed Eng*. Dec 2022;69(12):3612-3622. doi:10.1109/TBME.2022.3174680
39. Vallat R, Walker MP. An open-source, high-performance tool for automated sleep staging. *Elife*. Oct 14 2021;10doi:70092 [pii]
10.7554/eLife.70092
40. Terrill PI, Wilson SJ, Suresh S, Cooper DM, Dakin C. Application of recurrence quantification analysis to automatically estimate infant sleep states using a single channel of respiratory data. *Med Biol Eng Comput*. Aug 2012;50(8):851-65. doi:10.1007/s11517-012-0918-4
41. de Goederen R, Pu S, Silos Viu M, et al. Radar-based sleep stage classification in children undergoing polysomnography: a pilot-study. *Sleep Med*. Jun 2021;82:1-8. doi:S1389-9457(21)00193-3 [pii]
10.1016/j.sleep.2021.03.022
42. Vaquerizo-Villar F, Alvarez D, Kraemer JF, et al. Automatic Sleep Staging in Children with Sleep Apnea using Photoplethysmography and Convolutional Neural Networks. *Annu Int Conf IEEE Eng Med Biol Soc*. Nov 2021;2021:216-219. doi:10.1109/EMBC46164.2021.9629995
43. Hipp JF, Siegel M. Dissociating neuronal gamma-band activity from cranial and ocular muscle activity in EEG. *Front Hum Neurosci*. 2013;7:338.
44. Daftary AS, Jalou HE, Shively L, Slaven JE, Davis SD. Polysomnography Reference Values in Healthy Newborns. *J Clin Sleep Med*. Mar 15 2019;15(3):437-443.
45. Scholle S, Beyer U, Bernhard M, et al. Normative values of polysomnographic parameters in childhood and adolescence: Quantitative sleep parameters. *Sleep Medicine*. 2011/06/01/ 2011;12(6):542-549. doi:https://doi.org/10.1016/j.sleep.2010.11.011
46. White DM, Van Cott AC. EEG artifacts in the intensive care unit setting. *Am J Electroneurodiagnostic Technol*. Mar 2010;50(1):8-25.
47. Kawala-Sterniuk A, Podpora M, Pelc M, et al. Comparison of Smoothing Filters in Analysis of EEG Data for the Medical Diagnostics Purposes. *Sensors (Basel)*. Feb 2 2020;20(3)
48. Gavin CC, Nicola LCT. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *J Mach Learn Res*. 2010;11:2079–2107.
49. Sentner T, Wang X, de Groot ER, et al. The Sleep Well Baby project: an automated real-time sleep-wake state prediction algorithm in preterm infants. *Sleep*. Jun 24 2022;doi:10.1093/sleep/zsac143