Informativeness of Sleep Cycle Features in Bayesian Assessment of Newborn Electroencephalographic Maturation

Vitaly Schetinin, Livia Jakaite, and *Joachim Schult
*University of Bedfordshire, UK, *Universität Hamburg, Germany*
vitaly.schetinin@beds.ac.uk

Abstract

Clinical experts assess the newborn brain development by analyzing and interpreting maturity-related features in sleep EEGs. Typically, these features widely vary during the sleep hours, and their informativeness can be different in different sleep stages. Normally, the level of muscle and electrode artifacts during the active sleep stage is higher than that during the quiet sleep that could reduce the informativeness of features extracted from the active stage. In this paper, we use the methodology of Bayesian averaging over Decision Trees (DTs) to assess the newborn brain maturity and explore the informativeness of EEG features extracted from different sleep stages. This methodology has been shown providing the most accurate inference and estimates of uncertainty, while the use of DT models enables to find the EEG features most important for the brain maturity assessment.

1. Introduction

There are clinical evidences that the newborn brain development can be assessed from the maturity-related patterns recognizable in sleep electroencephalograms (EEG) [1], [2]. These patterns are used by clinical experts to estimate a newborn's physiological age [3], despite the fact that they widely vary during sleep hours and between patients that makes the analysis difficult and laborious. Automated analysis of the maturity-related patterns has been shown promising to assist experts in clinical interpretation [2], [4].

Typically, sleep EEG is recorded during a few hours and comprises one or more cycles of the active sleep (AS) and quiet sleep (QS) stages. These cycles or stages are recognizable in sleep EEG, and the information about them is useful for clinical interpretation of maturity-related patterns [5].

From developmental physiology it is known that sleep stages are recognizable in newborn EEGs since approximately 30 weeks post-conception. At this age, the QS is recognized as a pattern with high voltage bursts of delta, theta and alpha activity interrupted by periods with very low voltage. In contrast, the AS pattern is recognized as a longer period of uninterrupted medium-voltage theta and delta activity. The cyclic variations in the voltage and frequency corresponding to the sleep stages become more distinguishable with the brain maturation. For full-term newborns, the AS pattern is characterized by low to moderate voltage activity in theta, alpha and beta bands, whereas the QS pattern is often characterized by a high voltage delta activity [2], [5], [6].

It has been shown that the QS and AS patterns were significantly different in terms of voltage, powers in the delta and theta bands, as well as in terms of the number and length of pseudo-stationary segments [7]. In [8], the dimensional complexity of neonatal EEGs has been explored and shown to be significantly higher during the AS. In [9], the sleep stages have been differentiated by using 88 statistical features representing the voltage, frequency and cepstral coefficients. A technique for segmentation of newborn EEGs into pseudo-stationary intervals to be then clustered by the mean frequency and voltage has been proposed in [10]. As a result, the EEG intervals from the QS and AS stages were assigned to different clusters. In [11], a difference in the maturity-related patterns in the QS and AS stages has been found; particularly, the powers in theta and beta bands were most informative during the AS, whereas the alpha band was the most important feature during the QS. The most informative maturity-related feature found in [12] was the minimal voltage during the QS.

In this paper, we aim to explore the informativeness of EEG features extracted from the QS and AS stages for the brain maturity assessment. In general, feature importance is defined by the assessment method, and therefore our research is conditioned on a chosen approach. Our approach to the assessment is based on Bayesian averaging over classification models, in particularly, over decision trees (DTs) [13], [14]. Recently, this methodology has been shown promising for the assessment of brain maturity, while providing the information on EEG feature importance [15].

The next sections of the paper are as follows. Section 2 introduces the methodology of BMA over DTs. Section 3 describes the methods used for artifact removal and sleep stage segmentation. Section 4 describes the experiments, and Section 5 concludes the paper.

2. Bayesian model averaging over decision trees

Let us define parameters $\boldsymbol{Q}$ of a DT to be trained on the data $\boldsymbol{D}$ represented by a *m*-dimensional input $\boldsymbol{x}$. Given parameters $\boldsymbol{Q}_i$, we can then use the Bayesian theorem to count the class posterior distribution $p(y|\boldsymbol{Q}, \boldsymbol{D})$, where $y = \{1, C\}$ an outcome of the DT for a given input $\boldsymbol{x}$, and C is the number of classes. Having a conditional distribution $p(\boldsymbol{Q}|\boldsymbol{D})$, we can draw N samples $\boldsymbol{Q}_i$ from it and then write:

$$p(y|\boldsymbol{x}, \boldsymbol{D}) \simeq \sum_{i=1}^{N} p(y|\boldsymbol{x}, \boldsymbol{Q}_i, \boldsymbol{D}) p(\boldsymbol{Q}_i|\boldsymbol{D}) = \frac{1}{N} \sum_{i=1}^{N} p(y|\boldsymbol{x}, \boldsymbol{Q}_i, \boldsymbol{D})$$

In practice, the above approximation is efficiently made with Markov Chain Monte Carlo (MCMC) simulation technique. The main idea of this technique is to draw the samples $\boldsymbol{Q}_i$ from a Markov Chain, that is a DT model, when the distribution $p(\boldsymbol{Q}|\boldsymbol{D})$ is stationary, and samples are drawn from areas of interest with a high density of the probability (see e.g. [13], [14]).

The parameters of a DT are defined by the position, attribute and rule given for each splitting node. Having priors, users can specify a maximal number of splitting nodes, an expected DT configuration, as well as the preferences of using splitting attributes. The use of priors enables to explore DTs which split the data in many ways, taking into account that the areas of interest have to be sampled proportionally to the numbers of DT configurations.

In practice, when DTs are grown from given data, and their dimensionality is variable, the MCMC simulation has to be extended with Reversible Jump (see e.g. [13], [14]). Within this extension, we can change the dimensionality of a DT by using the *birth* and *death* moves.

The birth move randomly splits the data points falling in one of the DT terminal nodes by adding a new splitting node with an attribute and a rule drawn from the given prior distributions. This move increases the dimensionality of a DT. In the reverse, the death move decreases the dimensionality and randomly picks a splitting node with two terminal nodes to make this node a terminal.

The other moves are normally assumed as changing parameters of a DT: we can change a splitting attribute as well as a splitting rule. The *change-split* move picks a random splitting node and then changes its attribute along with its splitting rule drawn from the given prior distributions. The *change-rule* move picks a random splitting node whose rule is replaced by a new one drawn from a given prior distribution.

The birth and death moves change the DT configuration that enables the MCMC algorithm to explore possible ways of splitting data. The change-split and change-rule enable the MCMC algorithm to explore a DT within its current dimensionality. The change-split move enables to make "large" jumps which increase the chance of sampling the posterior distribution from different areas of interests, whilst the change-rule move makes "local" jumps enabling to explore an area of interest in detail.

The MCMC algorithm starts exploring a DT consisting of one splitting node with parameters (an attribute and a rule) which were randomly assigned within the predefined prior distributions. Making the birth and death moves, the algorithm lets a DT grow. While the DT grows, its likelihood tends to increase. and therefore during this phase the Markov Chain is unstable. This period named *burn-in* should be preset sufficiently long in order for the Markov Chain to become enough stable. When it is achieved, the phase named *post burn-in* is initiated during which the desired DT parameters are collected for counting the posterior distribution.

3. Artifact removal and segmentation of the quiet and active sleep stages

This section describes the methods used in our experiments for removal of artifacts from sleep EEG and segmentation of the quiet and active sleep stages.

3.1. Artifact removal

Typically, newborn sleep EEG recorded via the standard C3T3 and C4T4 electrodes are weak signals ranging between -127 and +127 μV, and their average amplitude is typically around 50 μV. During sleep hours EEGs are contaminated by noise and artifacts, so that there is the need of cleaning EEG data. Before processing, EEGs are normally rectified to make all amplitudes positive.

The variability of an EEG recorded during sleep hours of a newborn is quite high and can additionally affect the accuracy of recognition of age-related patterns in EEG. We found that the Mean-to-Deviation Ratio (MDR), defined as μ/σ, was around 1.0, where μ and σ are the mean and standard deviation of rectified EEG amplitudes.

The common artifacts, such as muscle, cardiac, eye blinking, breathing, and electrode movement, can

be labeled by an expert and then removed from EEG data. In our EEG data, the rate of labeled artifacts was widely ranged from 0.01 to 0.5, and the average rate of artifacts was around 0.1.

The EEGs were recorded in a number of clinics, and artifacts were labeled by different EEG experts. Consequently, we could not expect that the EEG artifacts were labeled consistently and so decided to remove from the EEG data only amplitude artifacts. We defined these artifacts as EEG samples of abnormally high amplitudes. Therefore, such artifacts can be automatically detected by using the standard method of adaptive thresholding (see e.g. [16]).

The idea of this method is based on the observation that the probability of abnormal EEG samples is distinctly smaller than that of normal samples. For stationary signals, the abnormality of samples can be adequately estimated in terms of their amplitudes. However, EEG are non-stationary signals, and abnormalities should be estimated within a window sliding over the EEG recording. The standard deviation over samples in a window has been shown providing more accurate estimates of the abnormality than the mean over sample amplitudes as its value is more sensitive to the non-stationarity of EEG [16].

In our implementation, a window of length W is moved over an EEG of length N, and the deviation d_i over samples in the window is counted for its central sample $i=\frac{W}{2}+1,\ldots,N-\frac{W}{2}$. The probability distribution over d_i is estimated in order to find the most frequent value d, as well as the maximal deviation d_{max}. Consequently, we expect that the normal EEG samples appear most frequently with the deviation d, and the abnormal samples appear with a higher deviation. This allows us to count probability q_i that the ith sample with deviation d_i is an artifact:

$$q_i=\frac{(d_i-d)}{d_{max}}, d\le d_i\le d_{max}.$$

Given an acceptable probability of artifacts in the window, q_0, we can then label a sample as an artifact if its deviation exceeds the threshold d_0:

$$d_0=d+(d_{max}-d)q_0.$$

The above technique is based on finding a reasonable trade-off between the accuracy of artifact detection and the amount of normal EEG samples being removed. In our experiments, we found that such a trade-off is achieved with a sliding window of 10-s duration and $q_0=0.225$.

Figure 1 shows an example of removing artifacts from a sleep EEG. The upper plot shows the raw EEG which was massively contaminated by artifacts visible as samples of a high amplitude in the interval between 120 and 200 min. The second plot shows the labels of these artifacts, whose rate was 0.126. The third plot shows the clean EEG. We see that the MDR of the raw EEG was 0.99, and for the clean EEG it increased to 1.11 due to removal of the amplitude artifacts.

3.2. Segmentation of the quiet and active sleep stages

The background information about sleep patterns in EEG given in the Introduction inspired us to segment EEG into quiet (QS) and active (AS) sleep stages. Similarly to the above technique of detecting EEG artifacts, these sleep stages can be segmented by adaptive thresholding. Within this technique, duration of the window and threshold value are made adjustable in order to achieve the best accuracy of segmentation on EEG data labeled by an expert. Besides, the information about a minimal duration of QS stage and a maximal duration of breaks, which can happen during QS, are used to improve the segmentation accuracy.

In our experiments, a threshold was adapted to an EEG recording as follows. First, we counted the deviation of samples, $d_k, k=1,\ldots,K$, in a window sliding over an EEG recording, where $K=Ent\left(\frac{N-W}{L}\right)$ is the number of windows counted in an EEG, and L is the window shift step. Then we counted a probability dis-

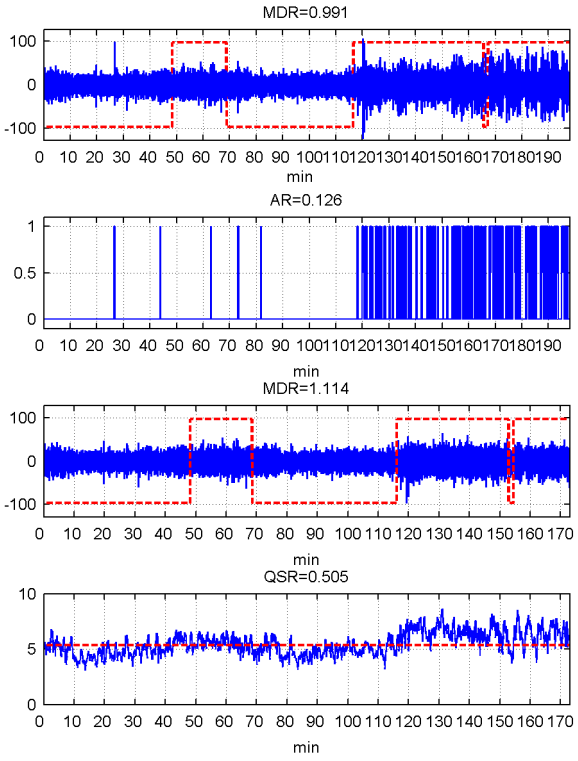

Figure 1. Segmentation of QS stages in EEG

tribution histogram, $p_i, i=1,\ldots,M$, over the deviation values v_i, where M is the number of bins in the histogram. Namely, the probabilities p_i are the portions of samples whose deviation values are between v_i and v_{i+1}. Thus, given a probability P_0, we can find a bin M_0:

$$\sum_{i=1}^{M_0} p_i \simeq P_0, M_0 \leq M \quad ,$$

and then choose a corresponding deviation interval, whose center is defined as the desired threshold. In this context, probability P_0 is normally associated with a prior on the frequency of an event detected by the segmentation technique.

In our experiments on EEG with the labeled QS stages, we have set $P_0=0.5$ to reflect the fact that the QS intervals were, on average, of a half of an EEG recording. The best accuracy of segmentation was achieved when the maximal duration of QS break intervals was set equal to one min. The minimal duration of QS intervals was set to seven min to enable segmenting QS intervals fragmented at the beginning and at the end of an EEG recording.

The above technique was first tested on a simulated EEG which has been modeled by smoothing white noise with a 300-sample window and weights equal to 1/300. Then we applied a 100-sample window with weights of 1/100 to simulate the two QS stages, the first between 30 and 60 min and the second between 120 and 150 min, so that the QS rate was 0.34. Finally, we set the amplitudes of the QS stages to be two times higher than those of AS stages.

The upper plot in Figure 2 shows the simulated EEG as the solid line. The dash line in this plot shows the detected QS stages which, as we observe, accurately match the given model. The second plot shows the amplitude artifacts which were detected during the QS stages. The third plot shows the EEG cleaned from the artifacts. The lower plot illustrates the process of thresholding detection of QS stages in the clean EEG. The dashed line in the third plot shows the result of QS detection. As the segmentation is made on the clean EEG, finally we extend the duration of sleep stages to the artifacts removed from the raw EEG. The resultant QS labels are shown in the upper plot as the dashed line.

The QS segmentation of a real EEG is illustrated in Figure 1. The upper plot shows that the QS stages have been properly segmented despite the high level of artifacts shown to be detected in the second plot. The window duration W and shift L were set to 30 s and 1.5 s, respectively.

In the next section, we describe the experiments with EEG features extracted from the labeled QS and AS intervals.

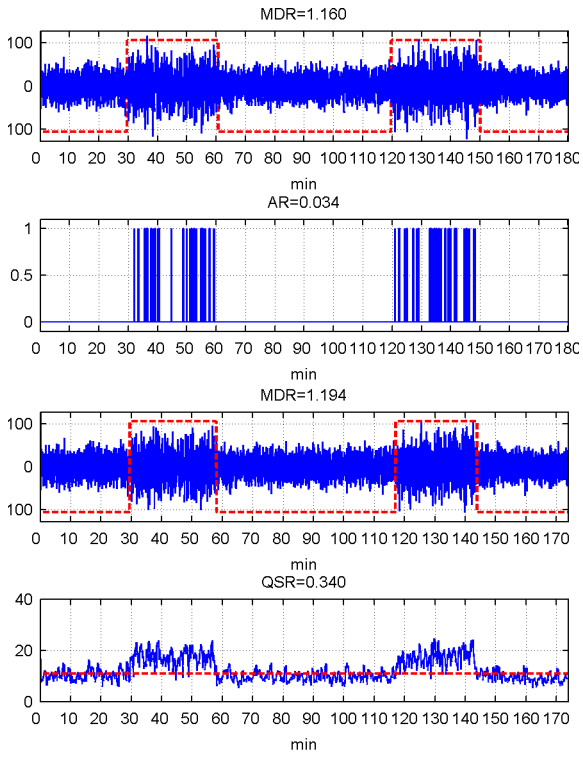

Figure 2. Segmentation of a simulated EEG

4. Experiments

In our experiments we used the methodology of BMA over DTs, described in Section 2, to assess the maturity of newborns of 36 and 41 weeks post-conception from sleep EEG recorded via the standard two-electrode system: the first age group included 107 and the second 104 recordings. The artifacts were detected and removed, and the sleep stages in each EEG were segmented into QS and AS intervals as described in Section 3.

Table 1 shows the rates of artifacts removed from the QS and AS intervals as well as from the whole EEG in each age group. The artifact rates are shown with the mean and 2σ intervals counted over the EEG recordings. We can observe that the rate of artifacts detected in the QS intervals is more than twice higher than that in the AS intervals. The second plot in Figure 1, which shows the labeled artifacts in the raw EEG, confirms that the artifact rate in the QS intervals is higher than that in the AS intervals. At the same time, the artifact rates in EEGs of 36 and 41 weeks seem quite similar to each other.

The rate of QS intervals in these experiments was, on average, 0.5 for both age groups. Within our segmentation technique described in Section 3.2, this rate includes EEG samples which have been detected as artifacts.

Table 1. Artifact rates in the sleep stages for both age groups

Sleep stage	36 weeks	41 weeks
QS	0.12±0.18	0.13±0.22
AS	0.05±0.07	0.05±0.08
QS and AS	0.09±0.12	0.09±0.14

After cleaning and segmentation, each EEG has been represented with the spectral powers in the following six standard bands: (1) *Subdelta*, (2) *Delta*, (3) *Theta*, (4) *Alpha*, (5) *Beta*, and (6) *Beta2*. These powers have been computed with a fast Fourier transform over 6-s epochs, which then were averaged within each band in order to represent an EEG by a six-element vector.

In the first two experiments with the Bayesian classification, we used the features extracted from the QS and AS intervals. The third experiment has been run with the features extracted from the whole, unsegmented EEG in order to compare the informativeness of features within the BMA methodology.

All the experiments have been run with the identical settings for MCMC simulation. The length of the burn-in phase and the numbers of samples to be collected during the post burn-in phase were set to 100,000 and 10,000, respectively. The probabilities of moves for birth, death, change-split, and change-rule have been set to 0.15, 0.15, 0.1, and 0.6, respectively. The proposals for making the change-rule moves have been drawn from a normal distribution with the zero mean and the standard deviation 1.0. The minimal size of data splits has been set to 2.

Table 2 shows the accuracy of the BMA obtained with the above settings within a 5-fold cross validation for the features extracted from the QS and AS stages as well as from the whole EEG. The performance is shown by the mean accuracy within 2σ standard deviation intervals. The entropy E is counted over an ensemble of K models, collected during the post burn-in phase, in order to estimate the uncertainty of the ensemble: $E=\sum_{i=1}^{K} p_i \log_2(p_i)$, where p_i are the class posterior probabilities of the DT models.

From Table 2, we see that the best accuracy, 85.75%, was obtained with EEG features extracted from the QS stage. These features provide a distinctly higher performance than those extracted from the AS stage.

Figure 3 shows the informativeness of the EEG features extracted from the QS and AS intervals. These features are represented by the six standard spectral power bands. Their informativeness is compared on the QS and AS intervals of EEG in terms of the posterior probabilities counted as the frequencies

Table 2. Performances of Bayesian age classification

Sleep stage	Performance, %	Entropy
QS	**85.75±11.27**	20.91±5.64
AS	76.73±29.34	26.31±3.74
QS and AS	81.98±13.76	22.26±4.08

of using these bands in DT models collected during the post burn-in phase.

From Figure 3, we observe that the two bands, (3) *Theta* and (4) *Alpha*, are most important for the age classification. We can observe that the informativeness of the sleep stages seems different for bands (2) *Delta* and (5) *Beta*, and similar for the other bands.

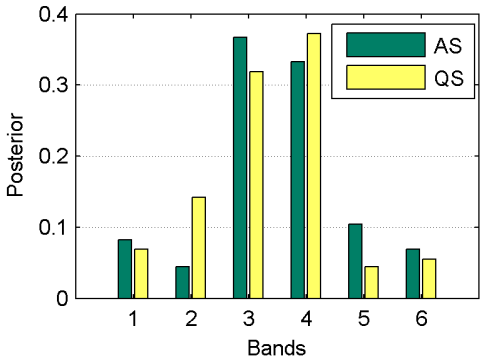

Figure 3. Posterior probabilities of EEG features for the AS and QS intervals

5. Discussion and conclusions

In this paper, we explored the informativeness of EEG features used for the newborn brain maturity assessment within the methodology of Bayesian averaging over decision tree (DT) models. This methodology has been shown providing the most accurate estimates of class posterior distribution, while the use of DT models has been shown capable of finding features making valuable contribution to the outcome.

Based on clinical observations that the EEG features in various sleep stages are different, we assumed that there exist EEG intervals which provide the most informative features. We further assumed that such features can be extracted from intervals of quiet sleep.

For testing this assumption, the EEGs were automatically segmented into the quiet and active sleep intervals. Before the segmentation, the EEGs were automatically cleaned from the amplitude artifacts. Both the segmentation and artifact detection have been made with the standard adaptive thresholding techniques. For each sleep stage, the segmented EEG intervals have been split into epochs and represented by the standard spectral bands. Finally, each band has

been averaged over the epochs to represent the EEG intervals by a vector entry.

In our experiments, we used the EEG data recorded from 211 newborns at age 36 and 41 weeks post-conception. We found that the EEG features extracted from the quiet sleep intervals have provided more accurate age classification in the comparison with the features extracted from the active sleep intervals: the mean accuracies of age classification were 85.75% and 76.3%, respectively. Besides, the features extracted from intervals of quiet sleep have provided a better performance in term of the ensemble entropy.

The above allows us to conclude that intervals of the quiet sleep in EEG are more informative for the newborn brain maturity assessment within the methodology of BMA over DT models. Obviously, this result is conditioned on the methods chosen in our research for segmenting EEG into sleep intervals, extracting features from the segmented EEG intervals as well as for classification of age-related patterns. These issues will be further explored to extend the results.

Acknowledgments. The research is funded by the Leverhulme Trust, UK. The authors are grateful to the University of Jena, Germany, for granting the clinical EEG data for the research, and to the anonymous reviewers for their useful comments on the paper.

6. References

[1] K. Holthausen et al. "Clinical relevance of age-dependent EEG signatures in the detection of neonates at high risk for apnea," *Neuroscience Letters Volume*, vol. 268, iss. 3, pp. 123–126, 199.

[2] R. Cooper, C. Binnie, and J. C. Schaw, "Clinical neurophysiology: EEG, pediatric neurophysiology, special techniques and applications," C. Binnie, R. Cooper, F. Mauguiere, J. W. Osselton, P. F. Prior, and B. M. Tedman, Eds. Elsevier Science, 2003.

[3] M. S. Scher, "Neurophysiological assessment of brain function and maturation: A measure of brain adaptation in high risk infants," *Pediatric Neurology*, vol. 16, iss. 3, 1997.

[4] M. S. Scher et al., "Automated state analyses: proposed applications to neonatal neurointensive care," *Journal of Clinical Neurophysiology*, vol. 22, iss. 4, pp. 256–270, 2005.

[5] G. B. Boylan, D. M. Murray, and J. M. Rennie, "Neonatal cerebral investigation," , J. M. Rennie, C. F. Hagmann, and N. J. Robertson, Eds. Cambridge University Press, 2008, pp. 83–91.

[6] E. Niedermeyer and L. F. H. da Silva, "Electroencephalography: basic principles, clinical applications, and related fields, 5th ed.," Lippincott Williams & Wilkins, 2005.

[7] K. Paul et al. "Comparison of quantitative EEG characteristics of quiet and active sleep in newborns," *Sleep Medicine*, vol. 4, iss. 6, pp. 543–552, 2003.

[8] S. Janjarasjitt, M. S. Scher, and K. Loparo, "Nonlinear dynamical analysis of the neonatal EEG time series: The relationship between neurodevelopment and complexity," *Clinical Neurophysiology*, vol. 119, iss. 8, p. 822–836, 2008.

[9] J. Lofhede and et al. "Automatic classification of background EEG activity in healthy and sick neonates," *Journal of Neural Engineering*, vol. 7, iss. 1, 2010.

[10] V. Krajca and et al. "Modeling the microstructure of neonatal EEG sleep stages by temporal profiles," in *Proc. IFMBE Proceedings of the 13th International Conference on Biomedical Engineering (CBME2008)*, 2009, pp. 133–137.

[11] M. S. Scher et al. "Maturational trends of EEG-sleep measures in the healthy preterm neonate." *Pediatric Neurology*, vol. 12, iss. 4, pp. 314–322, 1995.

[12] D. A. Viniker, D. E. Maynard, and D. F. Scott, "Cerebral function monitor studies in neonates," *Clinical Electroencephalography*, vol. 15, iss. 4, pp. 185–92, 1984.

[13] H. Chipman, E. George, and R. McCullock, "Bayesian CART model search," *Journal of American Statistics*, vol. 93, pp. 935–960, 1998.

[14] D. Denison et al. *Bayesian Methods for Nonlinear Classification and Regression*. 2002, Wiley.

[15] L. Jakaite et al. "Bayesian decision trees for EEG assessment of newborn brain maturity," in *Proc. UKCI 2010, the 10th Annual Workshop on Computational Intelligence*, 2010.

[16] H. Nolan, R. Whelan, and R. B. Reilly, "FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection," *Journal of Neuroscience Methods*, vol. 192, iss. 1, pp. 152–62, 2010.