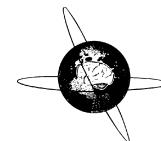Contents lists available at ScienceDirect

Clinical Neurophysiology

journal homepage: www.elsevier.com/locate/clinph

Automated classification of neonatal sleep states using EEG

Ninah Koolen [a], Lisa Oberdorfer [b], Zsofia Rona [b], Vito Giordano [b], Tobias Werther [b], Katrin Klebermass-Schrehof [b], Nathan Stevenson [a], Sampsa Vanhatalo [a,*]

[a] BABA Center, Department of Children's Clinical Neurophysiology, Children's Hospital, HUS Medical Imaging Center, Helsinki University Central Hospital and University of Helsinki, Finland
[b] Medical University Vienna, Department of Pediatrics, Division of Neonatology, Pediatric Intensive Care and Neuropediatrics, Vienna, Austria

See Editorial, pages 1039–1040

HIGHLIGHTS

- Neonatal EEG and its sleep state characteristics change across development.
- The combination of EEG features allows sleep state classification across preterm development.
- The classifier allows the generation of a sleep state probability index for long term monitoring.

ABSTRACT

Objective: To develop a method for automated neonatal sleep state classification based on EEG that can be applied over a wide range of age.
Methods: We collected 231 EEG recordings from 67 infants between 24 and 45 weeks of postmenstrual age. Ten minute epochs of 8 channel polysomnography ($N = 323$) from active and quiet sleep were used as a training dataset. We extracted a set of 57 EEG features from the time, frequency, and spatial domains. A greedy algorithm was used to define a reduced feature set to be used in a support vector machine classifier.
Results: Performance tests showed that our algorithm was able to classify quiet and active sleep epochs with 85% accuracy, 83% sensitivity, and 87% specificity. The performance was not substantially lowered by reducing the epoch length or EEG channel number. The classifier output was used to construct a novel trend, the sleep state probability index, that improves the visualisation of brain state fluctuations.
Conclusions: A robust EEG-based sleep state classifier was developed. It performs consistently well across a large span of postmenstrual ages.
Significance: This method enables the visualisation of sleep state in preterm infants which can assist clinical management in the neonatal intensive care unit.

© 2017 International Federation of Clinical Neurophysiology. Published by Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Despite recent advances in perinatal care, the incidence of compromised outcome in preterm infants remains a significant challenge (Aarnoudse-Moens et al., 2009; Johnson and Marlow,

2016). Prior studies have shown that a considerable proportion of developmental adversity arises from a prolonged stay in the neonatal intensive care unit (NICU). Progress in neurological care, therefore, holds significant promise for improving neurodevelopmental outcomes. A key challenge in early neurological care is the assessment or monitoring of brain function in infants who have limited ability to express their neurological function. Long term brain monitoring with scalp EEG has been widely adopted as a means to guide neurological care in the NICU (Bonifacio et al., 2011; Glass et al., 2011). A large array of studies using visual (Périvier et al., 2016; Song et al., 2015; Wikström et al., 2012) or computational analysis (Iyer et al., 2015b; Schumacher et al.,

2013) show the relationships between bedside recorded EEG activity and clinical course or outcome.

Methodologically, it has become clear that many computational methods could improve the assessment of EEG signals, and many of them could also be applied in bedside monitors (Iyer et al., 2015a, 2015b; O'Toole et al., 2016; Stevenson et al., 2012, 2013). However, such analysis tools would benefit from the automated segmentation of EEG into epochs based on vigilance states. This is because computational features can be sensitive to the large differences in EEG characteristics that is commonly seen between vigilance states (Palmu et al., 2013; Paul et al., 2003; Piryatinska et al., 2009; Räsänen et al., 2013). Sleep state classifiers have been developed for term infants (Paul et al., 2003; Piryatinska et al., 2009), but there is no classifier available that would be able to automatically identify sleep states across the whole neonatal age range from early prematurity to term age.

Clinically, an ability to express alternating vigilance states, i.e. sleep wake cycling, is considered a global surrogate marker for neurological wellbeing (El-Dib et al., 2014; Kidokoro et al., 2012; Klebermass et al., 2011; Natalucci et al., 2013; Osredkar et al., 2005; Thoresen et al., 2010). Literature on the early ontogenesis of sleep states is, nevertheless, controversial. Clinical neurophysiologists have established visually clear EEG patterns in active and quiet sleep state from 30 weeks postmenstrual age (André et al., 2010). Yet, clearly fluctuating patterns in the amplitude integrated EEG (aEEG) trend have been reported, and their time scales are compatible with genuine vigilance state cycling (Klebermass et al., 2011; Kuhle et al., 2001; Natalucci et al., 2013; Olischar et al., 2007; Palmu et al., 2013; Scher et al., 2005a; Reynolds et al., 2014; Stevenson et al., 2014; Karlsson et al., 2011; Lehtonen and Martin, 2004). The fluctuation in brain states, as observed in the EEG trends, has been taken as proxy of sleep wake cycling (Kidokoro et al., 2012; Thoresen, 2010). The interpretation of these trends, such as the aEEG, is highly subjective, requires specialised training, and it is sensitive to artefacts that readily corrupt EEG monitoring in the NICU.

There appears to be two complementary needs in the neonatal EEG interpretation. First, current analytical methods require the ability to segment the EEG signal according to sleep states. Second, bedside clinicians need an easier and more robust way to follow vigilance state fluctuations over extended times during the course of long term brain monitoring.

This study has set out to develop a sleep state classifier that provides a solution for both of these needs. We reasoned that a large cohort of preterm EEG records covering a wide range of postmenstrual ages would enable the development of a sufficient training dataset for a machine learning approach. We hypothesized that the number of computational features needed for the differentiation of sleep states is limited, and relatively stable across individuals and age ranges.

Finally, we wanted to test the possibility that this automated classifier could be used to implement a sleep state probability index for the visualisation of sleep state fluctuations over time.

2. Methods

An overview of the methods is presented in Fig. 1.

2.1. Data acquisition

All preterm infants (gestational age 23.1–28.0 weeks, median 25 weeks) born between January 2011 and July 2015 at the Medical University Vienna, Austria, were prospectively recruited into the study. Exclusion criteria consisted of congenital/chromosomal anomalies and severe perinatal asphyxia. All patients had regular

cerebral ultrasounds once a week until 34 weeks of postmenstrual age. The study was approved by the local ethics committee (Medical University Vienna) and the study protocol was registered (EK Nr 67/2008). Parental consent was required for inclusion.

A neonatal EEG polysomnography (hereafter referred to as PSG) was performed every second week for at least 3 h. The recordings were sampled at 256 Hz with Micromed EEG amplifier (System Plus Evolution Amplifier SAM 25 FO Mod. FC_1, Micromed S.p.A. Mogliano Veneto, Treviso, Italy) with eight EEG channels (FP1, FP2, C3, C4, T3, T4, O1, O2, GRASS gold electrodes 10 mm), which were placed according to the 10–20 international standard system and reference electrode G2. Polygraphic channels included electrocardiogram (ECG; Skintact pre-wired ECG electrodes, Philips Neo-LEad 3- Lead ECG Electrodes), respiratory rate sensor (Micromed respiratory belt) and video monitoring.

2.2. Pre-processing and epoch selection

All EEG recordings were visually inspected as well as the aEEG trend from the C3-C4 derivation to select epochs of quiet sleep (QS) and active sleep (AS). The putative quiet sleep and active sleep epochs were identified based on EEG characteristics (continuity, lower and upper aEEG border, bandwidth span and cycling in the aEEG) coupled to the polygraphic information (heart rate, respiratory rate and motility and eye movement). The most characteristic 10-min epochs for both AS and QS in the given infant were selected from every EEG recording (Table 1).

The 10-min AS and QS epochs were then visually checked and excluded if contaminated by persistent high amplitude artefact (>200 μV) on at least one EEG channel, long periods of low voltage (<20 μV), rhythmic artefacts, and movement artefacts. This resulted in the inclusion of 323 epochs out of the original 460 epochs (181 from QS and 142 from AS; 67 patients) recorded at postmenstrual age (PMA) 24.8–45.4 weeks (median 31.4 weeks).

In the present work, we rely on the evidence that at least two brain activity states are genuinely present in the early preterm infants (see also Introduction). We also make a pragmatic decision to search for the most characteristic epochs for each state in each infant.

2.3. Computation of EEG features

In search of a feature combination sufficient for classification of sleep states, we extracted 55 features that can be categorised into three types (Fig. 1A): (i) frequency-domain, (ii) time-domain, and (iii) features that capture spatial connectivity. In addition, we also included gestational and postmenstrual age. An overview of the feature set is given in Table 2; further details can be found in Appendix A.

The frequency content of the EEG has been shown to differentiate sleep states in term neonates (Paul et al., 2003; Piryatinska et al., 2009; Scher et al., 1996). We have, therefore, used frequency band power as a feature.

Summary measures of the amplitude over time have also been shown to differ between sleep states (Klebermass et al., 2011; Wikström et al., 2012). Here, we have used the EEG trend called range EEG (rEEG), the line length (LL) and multifractal detrended fluctuation analysis (MFDFA; Ihlen, 2012; Kantelhardt et al., 2002). MDFA was chosen, because it was recently shown to distinguish between levels of EEG continuity in pathological situations (Matic et al., 2015).

In contrast to the above features, which were estimated on each EEG channel and then averaged across channels, we also included features that reflect spatial coordination, or synchrony, between EEG channels. To this end, we computed the activation synchrony index (ASI) (Räsänen et al., 2013) in monopolar and bipolar deriva-
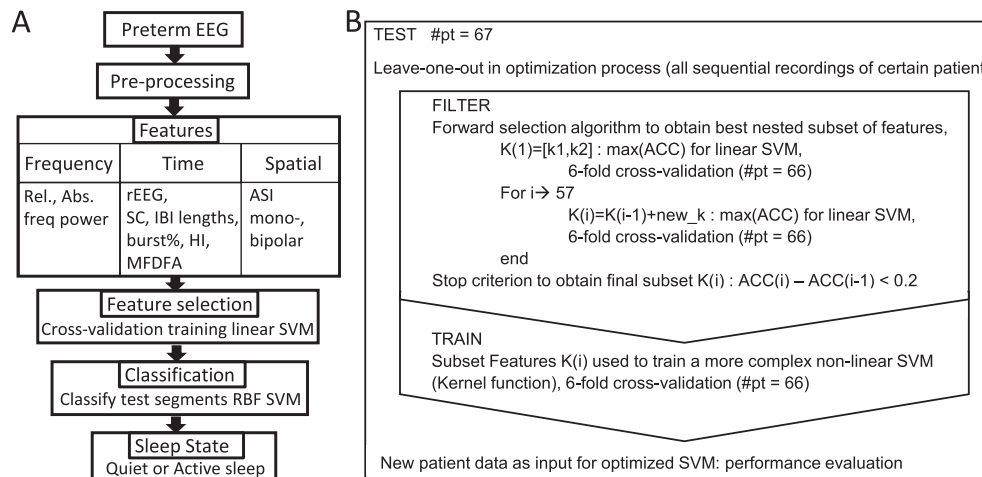
Fig. 1. (A) Schematic overview of the steps in the presented method for sleep state classification. (B) Detailed steps of the algorithm: feature selection by training linear SVMs and classification sleep epochs by optimized RBF SVMs.

Table 1
Characteristics of the data set. Data are presented as median (interquartile range) or as a count.

Basic characteristics	
Patients (N)	67
Gender (M/F)	33/34
PSGs per patient	3 (3–4)
Number of all PSG studies	231
Birth weight (g)	708 (606–915)
Gestational age (wks)	25 (24.5–27.2)
Postmenstrual age (wks)	31.4 (28.8–34.8)
IVH any	14
IVH severe (III-III+)	5

Notes: M – male, F – female, PSG – polysomnography, IVH – intracranial haemorrhage.

tions (Koolen et al., 2014b, 2016). Monopolar ASI derivations were averaged over brain areas.

In addition to these features, we have used a burst or spontaneous activity transient (SAT) detector based on the nonlinear energy operator (NLEO) to estimate the SAT percentage. This SAT percentage signal has been shown to correlate with the sleep-wake cycling pattern (Palmu et al., 2013; Stevenson et al., 2014). It is not used as a feature for the support vector machine (SVM) classifier but rather as a trend for qualitative comparison to the sleep state probability trend derived from the SVM output.

2.4. Feature selection and classification

The high number of features extracted from the EEG data challenges the classification algorithms by increasing training time and reducing classification accuracy due to overfitting. The high dimensionality was reduced using feature selection. The optimal feature subset K consists of k features from the original set of 57 features (55 EEG features, plus gestational and postmenstrual age, Table 2). Using different subsets of the features as an input for the SVM, we trained and optimized the classification of QS and AS epochs. The optimization criterion to be maximized during training was the classification accuracy, which is the ability to classify both QS and AS epochs correctly.

Many features are known to differ between sleep states. However, an optimal classifier needs to strike a balance between minimizing the number of features and maximizing the classification

Table 2
Overview of the 57 feature set, used as an input for feature selection by training SVMs.

Feature	Feature description	Single features for classification	References
1, 2	Age	Gestational age, postmenstrual age	
	Frequency-domain		
3 → 7	Relative power	delta1 (0.5-1 Hz), delta2 (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta(13-20 Hz)	
8 → 12	Absolute power	delta1 (0.5-1 Hz), delta2 (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta(13-20 Hz)	
	Time-domain		
13	EEG discontinuity	mean of the Suppression Curve	Dereymaeker et al. (2016)
14 → 17	EEG discontinuity	IBI perc95, IBI maximum, IBI median, burst percentage (after LL transform)	Koolen et al. (2014a)
18 → 24	rEEG measures	mean, median, UM (95th percentile), LM (5th percentile), STD, IQR, skewness	Navakatikyan et al. (2016); O'Reilly et al. (2012)
25	Histogram index	combined LL histogram features	Koolen et al. (2015)
26 → 35	LL histogram features	centers histogram bin (−2.7, −2.6, −2.5, −2.4, −2.3, −2.2, −2.1, −2, −1.9, −1.8)	Koolen et al. (2015)
36 → 43	LL histogram features	Mean, median, STD, IQR, skewness, kurtosis, 5th and 95th perc.	
44 → 47	MFDFA (3–8 Hz)	Mean_Dq, width_Dq, mean_hq, width_hq	Ihlen (2012), Kantelhardt et al. (2002)
	Spatial connectivity		
48 → 53	ASI bipolar derivations	Fp_1C_3/Fp_2C_4, Fp_1T_3/Fp_2T_4, Fp_1O_1/Fp_2O_2, C_3T_3/C_4T_4, C_3O_1/C_4O_2, T_3O_1/T_4O_2	Koolen et al. (2014b)
54 → 57	ASI monopolar derivations	Median GS, median IHS, median IntraH_R, median IntraH_L	Koolen et al. (2016)

Notes: IBI – Inter-burst Interval, LL – Line Length, UM – Upper Margin, LM – Lower Margin, MFDFA – Multifractal Detrended Fluctuation Analysis, ASI – Activation Synchrony Index, GS – Global Synchrony, IHS – Interhemispheric Synchrony, IntraH_R – Intrahemispheric Synchrony Right, IntraH_L – Intrahemispheric Synchrony Left.

accuracy. To this end, the features should be relevant but not excessively redundant (Guyon and Elisseeff, 2003). To build a good classifier, the usefulness of the *combined feature subset* depends on the performance improvement. We have opted for the forward selection method (Guyon and Elisseeff, 2003). This greedy algorithm initially searches for a combination of two features which have the best training accuracy. First, all pairwise combinations of 57 features are tested. Next, one extra feature k is included in the subset, for which the prediction accuracy of the classifier is maximized. This process is iterated 56 times to obtain the best order to include the features in the subset.

The detailed steps of the algorithm are shown in Fig. 1B. In each test loop, all EEG recordings of one patient are excluded (leave-one-out) to train and optimize the SVM classifier. Within the embedded training loop, feature selection is performed with a linear SVM applying 6-fold cross validation. The use of a linear SVM in feature selection decreases the computational requirements of training with negligible reduction in feature selection performance (Hsu et al., 2003). Feature selection is stopped, resulting in n_k features, by applying a stopping criterion: an increase in training accuracy smaller than 0.2%. Once the feature selection is performed, the parameters of a SVM with a radial basis function (RBF) kernel are chosen (Bi et al., 2003; Guyon and Elisseeff, 2003). For the RBF kernel SVM, two parameters are selected (RBF sigma σ and Box_Constraint C) using a fixed grid-search with 6-fold cross validation. In the final step, the features extracted from the left out infant's EEGs are used as an input for the trained RBF SVM for performance evaluation.

2.5. Classifier performance

The overall performance of the sleep state classifier is expressed using sensitivity and specificity. The sensitivity is the ability of the SVM to correctly classify the QS epochs, whereas the specificity is the ability to correctly classify AS epochs (1).

$$
\begin{aligned}
Sens &= \frac{TP}{TP + FN} \\
Spec &= \frac{TN}{TN + FP}
\end{aligned}
\tag{1}
$$

TP is the number of correctly classified QS epochs and TN the number of correctly classified AS epochs. The accuracy is defined as the number of correctly classified epochs (TP + TN) over the total number of epochs. We also used the area under the receiver operator characteristic (AUC) to express classifiers' performance.

2.6. Effects of channel number and epoch length

Our original data included 8 channels of EEG and relatively long, good quality data epochs. However, such recordings are not always available in the NICU where recordings may be with a reduced number of electrodes or interrupted by artefacts and care procedures. Hence, we felt that it would be clinically useful to evaluate the performance of sleep state classification with less recording channels and reduced epoch length. To this end, we assessed the performance of our classifier with 5 or 2.5-min EEG epochs extracted from the preselected 10-min epochs. We also computed classifier performance for a four electrode montage (Fp1, Fp2, C3 and C4), the current practice in preterm brain monitoring. Limiting the number of recording electrodes excluded the calculation of several spatial features, resulting in a modified feature set of 52 features.

2.7. Construction of sleep state probability (SSP) trend index

We constructed a probability trend based on the SVM output using consecutive 10-min EEG epochs with 50% overlap. The trend indicates the likelihood of active sleep. Recently, a comparable trend index was introduced for displaying seizure likelihood as a display solution for an automated bedside seizure detector (Mathieson et al., 2016). The algorithm computes the posterior probabilities using Platt's method (Platt, 1999). This proof of concept trend was compared, qualitatively, to other trend displays of neonatal vigilance state: (i) aEEG trend, (ii) suppression curve (Dereymaeker et al., 2016), (iii) burst percentage curve derived from the LL transform (Koolen et al., 2014a) and (iv) SAT% computed from the NLEO transform (Palmu et al., 2010). In the latter, active sleep is considered to be represented by a higher SAT percentage (Palmu et al., 2013), which is reflected in increasing discontinuity in the context of the suppression curve (Dereymaeker et al., 2016).

3. Results

3.1. Classification performance

A trained SVM, with a subset of the 57 features' set, was obtained for each patient. The classifier accuracy was found to increase by including up to 5–8 features in the subset for all iterations of the cross-validation (up to 88%; Fig. 2A). However, a clear ceiling effect and an increasing inter-subject variability was observed after adding 10–40 features. Indeed, an overfitting became apparent and the classifier performance tended to deteriorate when more than 40 features were used.

A median of 6.6 features were included resulting in a training accuracy of 88.44% (86.99–89.75%). Comparison of the distribution of individual features between infants showed that seven features were most often selected (% of infants): feature #3 relative delta 1 power (100%), feature #21 LM rEEG (100%), feature #52 bipolar ASI C3O1-C4O2 (97%), feature #48 bipolar ASI Fp1C3-Fp2C4 (84%), feature #19 median rEEG (42%), feature #16 median IBI length (42%) and feature #37 HI median (40%) (Fig. 2B). The discrimination of sleep states using the two most commonly selected features is shown in Fig. 2C. An example of the optimal feature values for QS and AS sleep is shown in Fig. 3.

The performance of the sleep state detection method is shown in Table 3 for 323 epochs of 67 infants. The EEG manifestation of sleep states is known to evolve with age, with a transition between tracé discontinue to tracé alternant at approximately 30 weeks PMA (André et al., 2010). Hence, we assessed the classifier performance separately for the early premature (PMA < 32 weeks) and less premature (PMA > 32wks) infants (Fig. 2D). For patients < 32 weeks PMA, the AUC was 0.834 with 95% confidence interval of 0.774–0.896 ($p < 0.001$). For older infants, the AUC was slightly higher (AUC = 0.879 with 95% confidence interval of 0.815–0.943; $p < 0.001$). However, there was no statistically significant difference between the age groups (Mann–Whitney, $Z = -0.718$, $p = 0.472$).

3.2. Effects of reducing epoch length or EEG channels

Classification based on shorter EEG epochs resulted in a small reduction in classification accuracy (Table 3). Conversely, the optimal set of features was relatively unaffected (Supplementary Material S.1, Supplementary Fig. S1). A decrease in the number of EEG channels from 8 channels to 4 channels (Fp1, Fp2, C3 and C4) resulted in a 2.2% decrease in classification accuracy (Table 3).
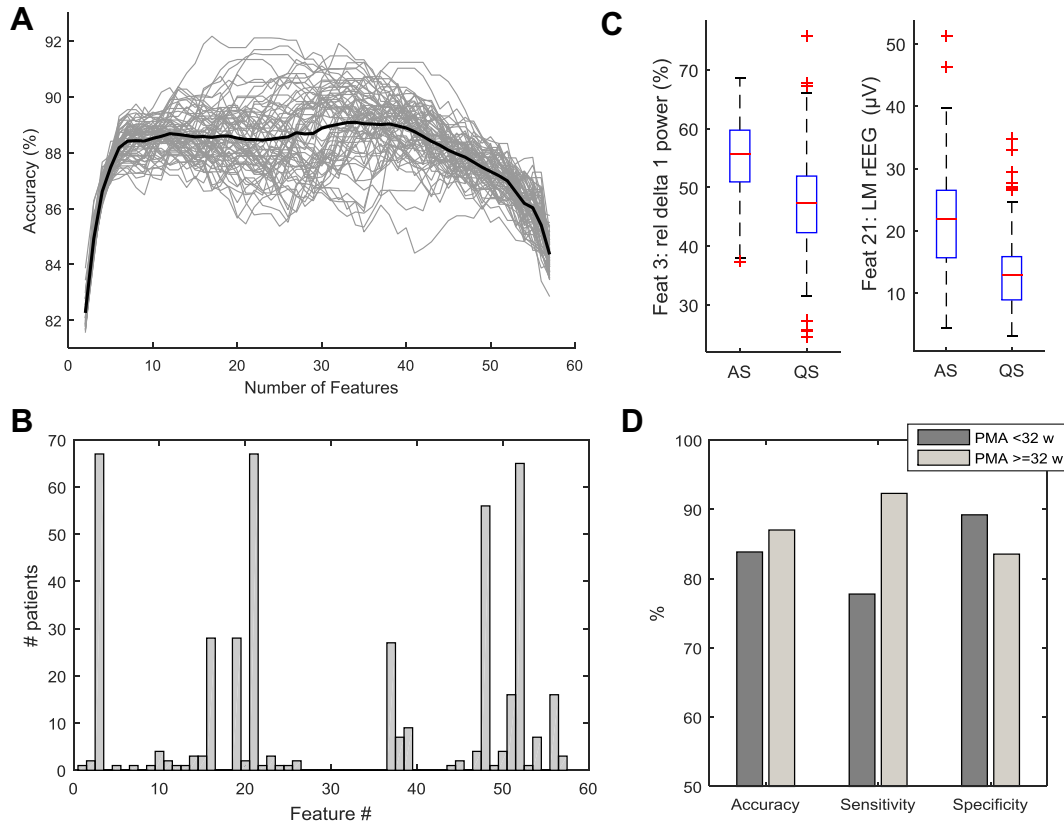
Fig. 2. Sleep state classifier performance. (A) Increasing the number of features in the reduced feature set induces a ceiling effect on the classifier's training performance. (B) Distribution of individual features used in the reduced feature set for classification. (C) Discrimination between sleep states for the primary selected features. (D) Classifiers' performance for early premature (PMA < 32 weeks) and less premature (>32 weeks) infants.
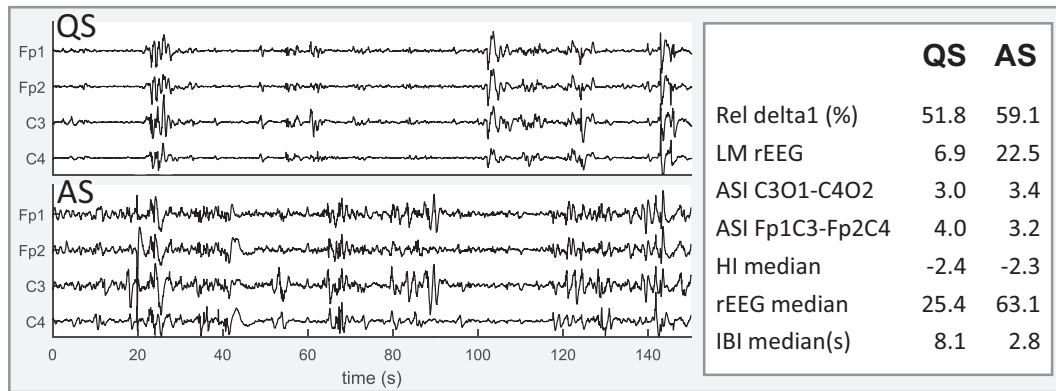

Fig. 3. Example of 2.5-min QS and AS epochs, with the extracted primary feature values.

Table 3
Classification performance with different epoch lengths and reduced number of EEG channels.

Epoch length	Number of channels	Accuracy (%)	Sensitivity (%)	Specificity (%)
10 min	8	85.14	83.10	86.74
5 min	8	84.72	80.69	87.89
2.5 min	8	82.19	79.26	84.54
10 min	4	82.97	81.69	83.98

3.3. Sleep state probability (SSP) trend

We finally wanted to test whether the classifier could also be used to generate a trend display for sleep state fluctuations (i.e. sleep wake cycling). To this end, the full EEG recording was given as an input to the RBF SVM classifier, which calculated sleep state probability (SSP) for every overlapping (50%) 10-min window. Fig. 4 shows comparative examples of SSP and other trend displays for very early and moderately preterm infants (PMA 27 and 34 weeks, respectively). These ages were selected to represent,
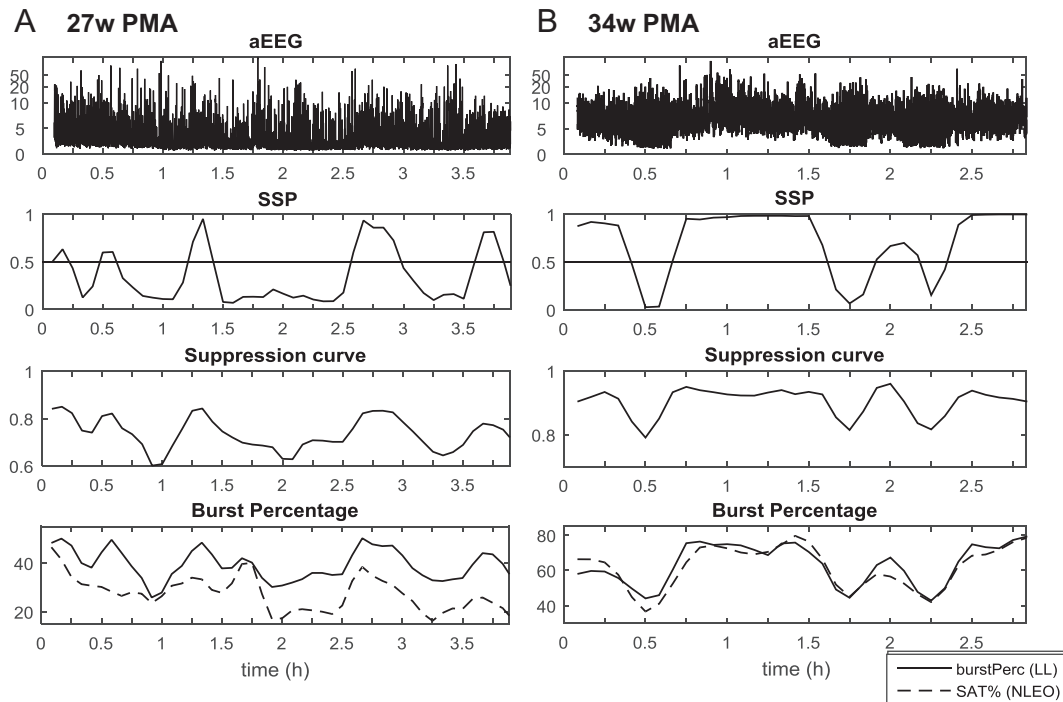
Fig. 4. Sleep state probability (SSP) trend during multiple hours EEG recording. (A) SSP trend of a very early preterm infant (27 weeks PMA) and (B) SSP trend of a more mature preterm infant (34 weeks PMA). The SSP trend is compared to the conventional aEEG, the suppression curve and the burst percentage (using the LL and NLEO transform). Note how the fluctuation is hardly visible in (lower boundary of) the aEEG trend of the younger preterm infant, but it becomes more clear with the previously published detection-based trends (Suppression curve and Burst Percentage), and very clear when using our novel SSP trend.

respectively, the earlier developmental phase with a consistently discontinuous EEG tracé and the later developmental phase characterized by discontinuity during only quiet sleep (André et al., 2010). As previously reported (Burdjalov et al., 2003), some fluctuation can be recognized with careful inspection in both examples, though it becomes clearer in the older age group. The SSP trend shows a robust alternation between brain states, however, the alternative displays only exhibit relative fluctuation of their baseline without indication of brain state.

There is, nevertheless, a correlation between SSP and the other trends. SSP correlates significantly with the suppression curve (very early preterm $r = 0.83$, mature preterm $r = 0.81$), the burst percentage curve (very early preterm $r = 0.83$, mature preterm $r = 0.90$) and the SAT percentage curve (very early preterm $r = 0.50$, mature preterm $r = 0.88$).

Finally, we wanted to assess the sensitivity of SSP to intermittent artefacts that may fragment the EEG to shorter than the 10 min epochs used in our initial training. Hence, we assessed the impact of window length on the quality of SSP trend. As shown in the Supplementary Fig. S2, computing the SSP trends from shorter windows preserved the overall structure, however, the trend becomes expectedly noisier and would benefit from smoothing prior to plotting on monitor displays.

4. Discussion

Our study shows that classification of sleep states is possible across a wide range of PMAs using a combination of features computed from the EEG. Prior studies have attempted to classify sleep states in narrower age ranges, focusing on either preterm (Janjarasjitt et al., 2008; Kidokoro et al., 2012; Thoresen et al., 2010; Werth et al., 2016) or term (Paul et al., 2003; Piryatinska et al., 2009; Scher et al., 2005b, 2005c) infants. The present work extends prior attempts by developing a classifier that performs

well across a wide range of preterm development, as well as by providing a proof of concept for clinical implementation, a bedside trend for visualising fluctuations in brain states.

A clear clinical strength in our current dataset is inclusion of both healthy and medically ill patients. Our work collated data from all recruited infants when AS and QS sleep states were identifiable, including babies with very poor EEG background. This practical choice was done for the purpose of training the classifier for clinical context. Future work is needed to study our classifier's performance specifically on different sub-populations, such as severely abnormal EEG records. It would be advisable to include multiple human raters to select sleep epochs. However, we would not expect this to strongly affect the results as inter-rater agreement for sleep segmentation is high (Stefanski et al., 1984; Crowell et al., 1997).

We show that the classification of EEG epochs into two sleep states can be as high as 85%, which compares well with inter-rater agreement in sleep scoring of conventional polysomnographic recordings (Crowell et al., 1997). Feature selection improves the performance of the classifier. In this case, the forward selection method shows good results, because it is optimized for the SVM classification algorithm (Bennasar et al., 2015). However, there is a risk of overfitting the data by including too many redundant features in the feature subset. Our work suggests that a relatively low number of independent features performs well. The most distinctive features were the delta 1 power and the lower margin of the rEEG signal, which are readily reconciled by considering the most salient elements of sleep EEG signals (André et al., 2010).

Although the SVM based optimization is, by design, ignorant of prior physiological knowledge, it seems that the relationship between sleep states and EEG maturation is inherently incorporated in these features: For instance, preterm development of EEG is characterized by a gradual increase in the amplitudes of

intermittent quiescence (Lombroso, 1985; Olischar et al., 2004, 2007; O'Reilly et al.,2012; Selton et al., 2000; Vecchierini et al., 2007), hence the rEEG feature (lower margin) incorporates an age component as well. Likewise, delta 1 power has been shown to decrease with age (Niemarkt et al., 2011; Okumura et al., 2003; Scher et al., 1995).

A state-specific segmentation of EEG is important for most analytic approaches, because neonatal EEG activity is different between sleep states (André et al., 2010; Koolen et al.; 2016; Palmu et al., 2013; Tokariev et al., 2015). Our work provides a method that can be readily incorporated into clinical as well as scientific software to allow automated segmentation of EEG records. As to epoch lengths, we observed slightly improving classifier performance with epochs lengths up to 10 min. Such epoch length is obviously much longer than the traditional 30 s epochs used in polysomnographic sleep studies, however most computational EEG analyses require such epochs with at least several minutes of signal (Koolen et al., 2014b, 2016; Matic et al., 2015; Palmu et al., 2013; Tokariev et al., 2015).

We wanted to test whether such classifier could produce a potentially useful trend index in the long term brain monitoring. The current clinical practice in bedside EEG monitoring is based on assessing sleep wake cycling as a fluctuating pattern in the aEEG trend (Hellström-Westas et al., 2006; Kuhle et al., 2001; Olischar et al., 2004). That approach is widely adopted, but it suffers from lack of intra- and inter-rater reliability, as well as sensitivity to artefact contamination. Inspection of the aEEG in Fig. 4 shows the alternation between sleep states, but it also demonstrates the difficulty in identifying sleep states, especially in the early preterm infant groups. As shown in Fig. 4, the baselines of already published trends of suppression curve and the burst/SAT percentage curves correlate with the sleep state classifier output. However, EEG activity during QS develops into tracé alternant patterns with less marked bursts and reliable sleep cycles cannot be visualised with previously published straightforward methods based on burst detection (Dereymaeker et al., 2016; Palmu et al., 2013). The advantage of using the SVM sleep state classifier is clear; it's ability to track the SSP over a broad range of postmenstrual age from early prematurity to term age.

In clinical practice, the EEG signal is commonly contaminated by artefacts, and it would be advantageous to have methods that are less sensitive to epoch length or the number of recording electrodes. We observed that, expectedly, longer EEG epochs give better classifier performance; however, EEG epochs as short as 2.5 or 5 min gave comparable performance for trend displays. This is clinically important since it is much easier to find high quality 2.5-min epochs than 10-min epochs. The ability to use short epochs for the classifier is most helpful for building the SSP trend display; a high quality SSP trend can be constructed automatically from the EEG epochs after automated artefact removal followed by smoothing of the trend as shown in our work. Since the standard clinical practice in brain monitoring is based on less than full array of EEG channels, it is clinically important that our data shows minimal decrease in classification accuracy with less than 8 EEG channels.

The classifier developed and tested in our present work paves the way for an automated EEG analysis which holds promise for improving brain monitoring. The artefact rejection paradigm can also be automated and incorporated into the pipeline to improve reliability. While the state-specific segmentation provides a preprocessing function for other analyses, the SSP trend is already a clinical implementable tool. Currently, this would enable the clinician to identify high-risk-patients for brain malfunction at an early state as delayed development of sleep wake cycling is correlated to impaired neurodevelopmental outcome (Natalucci et al., 2013; Klebermass et al., 2011). Recognition of early sleep states would also improve developmental care as less interruption during

quiet sleep can be a therapeutic strategy (Ludington-Hoe et al., 2006). The utility of a sleep state display will be determined in prospective clinical trials and by assessing the perceived bedside added value in the infant care.

Acknowledgements

Appendix A. Feature description, additional information on Section 2.3

Frequency-domain features

We computed 10 frequency features that represent the absolute and the relative power, estimated as the ratio of the absolute power in a specific frequency band divided by the total energy within the whole frequency range (0.5–20 Hz). We computed power for the following frequency ranges: delta1 (0.5–1 Hz), delta2 (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–20 Hz). Although the cortical rhythms of preterm infants are noticeably slower than that of the adult, the use of the aforementioned bands has utility in neonatal EEG analysis (see Tokariev et al., 2012).

Time-domain features

The novel EEG trend called range EEG (rEEG) is defined as the maximum EEG value minus the minimum EEG value in non-overlapping 2 s windows (Navakatikyan et al., 2016; O'Reilly et al., 2012). In each 10-min epoch, we computed 7 features of the rEEG transform: the mean, median, upper margin (UM), lower margin (LM), standard deviation, interquartile range and the skewness. LM and UM are the 5th and 95th percentiles of the rEEG.

In addition to direct amplitude measures, we computed a nonlinear derivation of amplitude using the line length (LL) (Koolen et al., 2014a). We included the following 24 features:

- The QS period is characterized by tracé discontinue/alternant pattern, which equals to longer inter-burst intervals (IBI). We hence detected bursts using the detection algorithm as presented in Koolen et al. (2014a), based on LL, and we extracted four features of the IBI: 95th percentile, maximum, median and burst percentage which was computed as the relative (%) amount of bursts in the EEG 10-min epochs (see also Palmu et al., 2013).
- The mean of the suppression curve; representing the level of discontinuity of the EEG signal (Dereymaeker et al., 2016).
- Histogram features; the distribution of log LLs was summarized in a histogram index (Koolen et al., 2015). This was motivated by our observation that QS includes a wide range of LL values, however AS is mainly represented by large LL values. We included 18 histogram features; the height of 10 predefined histogram bins and 8 statistical measurements: mean, median, standard deviation, interquartile range, skewness, kurtosis and the 5th and 95th percentile.

Besides the amplitude features, multifractal detrended fluctuation analysis (MFDFA; Ihlen, 2012; Kantelhardt et al., 2002) has been shown to be useful in AS vs QS distinction). We computed 4 MFDFA features: mean Dq, width Dq, mean hq, width hq (Ihlen, 2012; Kantelhardt et al., 2002).

Synchrony features

We also included features representing the spatial connectivity between EEG channels, derived from both monopolar and bipolar derivations. These synchrony features are all extracted from 2.5-min EEG epochs as it was shown to be optimal for the Activation Synchrony Index calculation (Koolen et al., 2014b, 2016), and subsequently averaged across the 10-min epochs. We selected the following 10 synchrony features:

- Global synchrony (GS): the average of 28 pairwise monopolar channel combinations to characterize global connectivity.
- Interhemispheric synchrony (IHS): the average of symmetric monopolar channel combinations between hemispheres.
- Intrahemispheric synchrony Left and Right: the average of the 6 ASI values obtained by the 6 channel monopolar combinations in each hemisphere.
- Interhemispheric synchrony from symmetric bipolar derivations; Fp_1C_3/Fp_2C_4, Fp_1T_3/Fp_2T_4, Fp_1O_1/Fp_2O_2, C_3T_3/C_4T_4, C_3O_1/C_4O_2, T_3O_1/T_4O_2. When using 4 EEG channels for sleep state classification, the only bipolar derivation used is Fp_1C_3/Fp_2C_4.

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.clinph.2017.02.025.

References

Aarnoudse-Moens CS, Weisglas-Kuperus N, van Houdoever JB, Oosterlaan J. Meta-analysis of neurobehavioral outcomes in very preterm and/or very low birth weight children. Pediatrics 2009;124(2):717–28.

André M, Lamblin M-D, d'Allest AM, Curzi-Dascalova L, Moussalli-Salefranque F, Nguyen S, et al. Electroencephalography in premature and full-term infants. Developmental features and glossary. Clin Neurophysiol 2010;40(2):59–124.

Bennasar M, Hicks Y, Setchi R. Feature selection using joint mutual information maximisation. Expert Syst Appl 2015;42:8520–32.

Bi J, Bennett K, Embrechts M, Breneman C, Song M. Dimensionality reduction via sparse support vector machines. J Mach Learn Res 2003;3:1229–43.

Bonifacio SL, Glass HC, Peloquin S, Ferriero DM. A new neurological focus in neonatal intensive care. Nat Rev Neurol 2011;7(9):485–94.

Burdjalov VF, Baumgart S, Spitzer AR. Cerebral function monitoring: a new scoring system for the evaluation of brain maturation in neonates. Pediatrics 2003;112:855–61.

Crowell DH, Brooks LJ, Colton T, Corwin MJ, Hoppenbrouwers TT, Hunt CE, et al. Infant polysomnography: reliability. Collaborative Home Infant Monitoring Evaluation (CHIME) Steering Committee. Sleep 1997;20(7):553–60.

Dereymaeker A, Koolen N, Vervisch J, Ortibus E, Jansen K, De Vos M, et al. The suppression curve as a quantitative approach for measuring brain maturation in preterm infants. Clin Neurophysiol 2016;127(8):2760–5.

El-Dib M, Massaro AN, Glass P, Aly H. Sleep wake cycling and neurodevelopmental outcome in very low birth weight infants. J Matern Fetal Neonatal Med 2014;27(9):892–7.

Glass HC, Bonifacio SL, Shimotake T, Ferriero DM. Neurocritical care for neonates. Curr Treat Options Neurol 2011;13(6):574–89.

Guyon I, Elisseeff A. An Introduction to variable and feature selection. J Mach Learn Res 2003;3:1157–82.

Hellström-Westas L, Rosén I, De Vries L, Greisen G. Amplitude integrated EEG classification and interpretation in preterm and term infants. NeoReviews 2006;7(2):76–87.

Hsu C-W, Chang C-C, Lin C-J. A practical guide to support vector classification. Department of Computer Science and Information Engineering, National Taiwan University; 2003. p. 1–16.

Ihlen EA. Introduction to multifractal detrended fluctuation analysis in Matlab. Front Physiol 2012;3:141.

Iyer KK, Roberts JA, Hellström-Westas L, Wikström S, Hansen Pupp I, Ley D, et al. Early detection of preterm intraventricular hemorrhage from clinical electroencephalography. Crit Care Med 2015a;43(10):2219–27.

Iyer KK, Roberts JA, Hellström-Westas L, Wikström S, Pupp I Hansen, Ley D, et al. Cortical burst dynamics predict clinical outcome early in extremely preterm infants. Brain 2015;138(Pt 8):2206–18.

Janjarasjitt S, Scher MS, Loparo KA. Nonlinear dynamical analysis of the neonatal EEG time series: the relationship between sleep state and complexity. Clin Neurophysiol 2008;119(8):1812–23.

Johnson S, Marlow N. Early and long-term outcome of infants born extremely preterm. Arch Dis Child 2016. http://dx.doi.org/10.1136/archdischild-2015-309581.

Kantelhardt JW, Zschiegner SA, Koscielny-Bunde E, Havlin S, Bunde A, Stanley HE. Multifractal detrended fluctuation analysis of nonstationary time series. Phys A Stat Mech Appl 2002;316:87–114.

Karlsson KA, Arnardóttir H, Robinson SR, Blumberg MS. Dynamics of sleep-wake cyclicity across the featl period in sheep (Ovis aries). Dev Psychobiol 2011;53(1):89–95.

Kidokoro H, Inder T, Okumura A, Watanabe K. What does cyclicity on amplitude-integrated EEG mean? J Perinatol 2012;32(8):565–9.

Klebermass K, Olischar M, Waldhoer T, Fuiko R, Pollak A, Weninger M. Amplitude-integrated EEG pattern predicts further outcome in preterm infants. Pediatr Res 2011;70(1):102–8.

Koolen N, Jansen K, Vervisch J, Matic V, De Vos M, Naulaers G, et al. Line length as a robust method to detect high activity events: automated burst detection in premature EEG recordings. Clin Neurophysiol 2014a;125(10):1985–94.

Koolen N, Dereymaeker A, Räsänen O, Jansen K, Vervisch J, Matic V, et al. Interhemispheric synchrony in the neonatal EEG revisited: activation synchrony index as a promising classifier. Front Hum Neurosci 2014b;8:1030.

Koolen N, Dereymaeker A, Räsänen O, Jansen K, Vervisch J, Matic V, et al. Data-driven metric representing the maturation of preterm EEG. Conf Proc IEEE Eng Med Biol Soc 2015:1492–5. http://dx.doi.org/10.1109/EMBC.2015.7318653.

Koolen N, Dereymaeker A, Räsänen O, Jansen K, Vervisch J, Matic V, et al. Early development of synchrony in cortical activations in the human. Neuroscience 2016;322:298–307.

Kuhle S, Klebermass K, Olischar M, Hulek M, Prusa AR, Kohlhauser C, et al. Sleep-wake cycles in preterm infants below 30 weeks of gestational age: preliminary results of a prospective amplitude-integrated EEG study. Wien Klin Wochenschr 2001;113:219–23.

Lehtonen L, Martin RJ. Ontogeny of sleep and awake states in relation to breathing in preterm infants. Semin Neonatol 2004;9(3):229–38.

Lombroso CT. Neonatal polygraphy in full-term and premature infants: a review of normal and abnormal findings. J Clin Neurophysiol 1985;2(2):105–55.

Ludington-Hoe SM, Johnson MW, Morgan K, Lewis T, Gutman J, Wilson PD, et al. Neurophysiologic assessment of neonatal sleep organization: preliminary results of a randomized, controlled trial of skin contact with preterm infants. Pediatrics 2006;117(5):909–23.

Mathieson SR, Stevenson NJ, Low E, Marnane WP, Rennie JM, Temko A, et al. Validation of an automated seizure detection algorithm for term neonates. Clin Neurophysiol 2016;127(1):156–68.

Matic V, Cherian P, Koolen N, Ansari A, Naulaers G, Govaert P, et al. Objective differentiation of neonatal EEG background grades using detrended fluctuation analysis. Front Hum Neurosci 2015;9:189.

Natalucci G, Rousson V, Bucher HU, Bernet V, Hagmann C, Latal B. Delayed cyclic activity development on early amplitude-integrated EEG in the preterm infant with brain lesions. Neonatology 2013;103(2):134–40.

Navakatikyan MA, O'Reilly D, Van Marter LJ. Automated measurement of interburst interval in premature neonates using range EEG. Clin Neurophysiol 2016;127(2):1233–46.

Niemarkt HJ, Jennekens W, Pasman JW, Katgert T, Van Pul C, Gavilanes AW, et al. Maturational changes in automated EEG spectral power analysis in preterm infants. Pediatr Res 2011;70(5):529–34.

Okumura A, Kubota T, Toyota N, Kidokoro H, Maruyama K, Kato T, et al. Amplitude spectral analysis of maturational changes of delta waves in preterm infants. Brain Dev 2003;25:406–10.

Olischar M, Klebermass K, Kuhle S, Hulek M, Kohlhauser C, Rücklinger E, et al. Reference values for amplitude-integrated electroencephalographic activity in preterm infants younger than 30 weeks' gestational age. Pediatrics 2004;113(1 Pt1):61–6.

Olischar M, Klebermass K, Waldhoer T, Pollak A, Weninger M. Background patterns and sleep-wake cycles on amplitude-integrated electroencephalography in preterms younger than 30 weeks gestational age with peri-/intraventricular haemorrhage. Acta Paediatr 2007;96(12):1743–50.

Osredkar D, Toet MC, van Rooij LG, van Huffelen AC, Groenendaal F, de Vries LS. Sleep-wake cycling on amplitude-integrated electroencephalography in term newborns with hypoxic-ischemic encephalopathy. Pediatrics 2005;115(2):327–32.

O'Reilly D, Navakatikyan MA, Filip M, Greene D, Van Marter LJ. Peak-to-peak amplitude in neonatal brain monitoring of premature infants. Clin Neurophysiol 2012;123:2139–53.

O'Toole JM, Boylan GB, Vanhatalo S, Stevenson NJ. Estimating functional maturity in very and extremely preterm neonates using automated analysis of the electroencephalogram. Clin Neurophysiol 2016;127(8):2910–8.

Palmu K, Kirjavainen T, Stjerna S, Salokivi T, Vanhatalo S. Sleep wake cycling in early preterm infants: comparison of polysomnographic recordings with a novel EEG-based index. Clin Neurophysiol 2013;124(9):1807–14.

Palmu K, Stevenson N, Wikström S, Hellström-Westas L, Vanhatalo S, Palva JM. Optimization of an NLEO-based algorithm for automated detection of

spontaneous activity transients in early preterm EEG. Physiol Meas 2010;31 (11):85–93.

Paul K, Krajca V, Roth Z, Melichar J, Petránek S. Comparison of quantitative EEG characteristics of quiet and active sleep in newborns. Sleep Med 2003;4 (6):543–52.

Périvier M, Rozé JC, Gascoin G, Hanf M, Branger B, Rouger V, et al. Neonatal EEG and neurodevelopmental outcome in preterm infants born before 32 weeks. Arch Dis Child Fetal Neonatal Ed 2016;101(3):F253–9.

Piryatinska A, Terdik G, Woyczynski W, Loparo K, Scher MS, Zlotnik A. Automated detection of neonatal EEG sleep stages. Comput Methods Programs Biomed 2009;95(1):31–46.

Platt J. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Advances in Large Margin Classifiers. MIT Press; 1999. p. 61–74.

Räsänen O, Metsäranta M, Vanhatalo S. Development of a novel robust measure for interhemispheric synchrony in the neonatal EEG: activation synchrony index (ASI). Neuroimage 2013;69:256–66.

Reynolds LC, Pineda RG, Mathur A, Vavasseur C, Shah DK, Liao S, et al. Cerebral maturation on amplitude-integrated electroencephalography and perinatal exposures in preterm infants. Acta Paediatr. 2014;103(3):96–100.

Selton D, André M, Hascoët JM. Normal EEG in very premature infants: reference criteria. Clin Neurophysiol 2000;111(12):2116–24.

Scher MS, Steppe DA, Banks DL, Guthrie RD, Sclabassi RJ. Maturational trends of EEG-sleep measures in the healthy preterm neonate. Pediatr Neurol 1995;12:314–22.

Scher MS, Dokianakis SG, Sun M, Steppe DA, Guthrie RD, Clabassi RJ. Computer classification of sleep in preterm and full-term neonates at similar postconceptional ages. Sleep 1996;19(1):18–25.

Scher MS, Johnson MW, Holditch-Davis D. Cyclicity of neonatal sleep behaviors at 25 to 30 weeks' postconceptional age. Pediatr Res 2005a;57:879–82.

Scher MS, Waisanen H, Loparo K, Johnson MW. Prediction of neonatal state and maturational change using dimensional analysis. J Clin Neurophysiol 2005b;22 (3):159–65.

Scher MS, Turnbull J, Loparo K, Johnson MW. Automated state analyses: proposed applications to neonatal neurointensive care. J Clin Neurophysiol 2005c;22 (4):256–70.

Schumacher EM, Larsson PG, Sinding-Larsen C, Aronsen R, Lindeman R, Skjeldal OH, et al. Automated spectral EEG analyses of premature infants during the first three days of life correlated with developmental outcomes at 24 months. Neonatology 2013;103(3):205–12.

Song J, Xu F, Wang L, Gao L, Guo J, Xia L, et al. Early amplitude-integrated electroencephalography predicts brain injury and neurological outcome in very preterm infants. Sci Rep 2015;5:13810. http://dx.doi.org/10.1038/srep13810.

Stevenson NJ, O'Toole JM, Rankine LJ, Boylan GB, Boashash B. A nonparametric feature for neonatal EEG seizure detection based on a representation of pseudo–periodicity. Med Eng Phy 2012;34:437–46.

Stevenson NJ, Korotchikova I, Temko A, Lightbody G, Marnane WP, Boylan GB. An automated system for grading EEG abnormalities in term neonates with hypoxic ischaemic encephalopathy. Ann Biomed Eng 2013;41:775–85.

Stevenson NJ, Palmu K, Wikström S, Hellström-Westas L, Vanhatalo S. Measuring brain activity cycling (BAC) in long term EEG monitoring of preterm babies. Physiol Meas 2014;35:1493–508.

Thoresen M, Hellström-Westas L, Liu X, de Vries LS. Effect of hypothermia on amplitude-integrated electroencephalogram in infants with asphyxia. Pediatrics 2010;126(1):131–9.

Tokariev A, Palmu K, Lano A, Metsäranta M, Vanhatalo S. Phase synchrony in the early preterm EEG: development of methods for estimating synchrony in both oscillations and events. Neuroimage 2012;60(2):1562–73.

Tokariev A, Videman M, Palva JM, Vanhatalo S. Functional brain connectivity develops rapidly around term age and changes between vigilance states in the human newborn. Cereb Cortex 2015. http://dx.doi.org/10.1093/cercor/bhv219.

Vecchierini M-F, André M, d'Allest AM. Normal EEG of premature infants born between 24 and 30 weeks gestational age: terminology, definitions and maturation aspects. Neurophysiol Clin 2007;37(5):311–23.

Werth J, Atallah L, Andriessen P, Long X, Zwartkruis-Pelgrim E, Aarts RM. Unobtrusive sleep state measurements in preterm infants - a review. Sleep Med Rev. 2016. http://dx.doi.org/10.1016/j.smrv.2016.03.005. pii: S1087-0792 (16)30006-5.

Wikström S, Pupp IH, Rosén I, Norman E, Fellman V, Ley D, et al. Early single-channel aEEG/EEG predicts outcome in very preterm infants. Acta Paediatr 2012;101(7):719–26.