An automated bedside measure for monitoring neonatal cortical activity: a supervised deep learning-based electroencephalogram classifier with external cohort validation

Moghadam, Saeed Montazeri

2022-12

Moghadam , S M , Airaksinen , M , Nevalainen , P , Marchi , V , Hellström-Westas , L , Stevenson , N J & Vanhatalo , S 2022 , ' An automated bedside measure for monitoring neonatal cortical activity: a supervised deep learning-based electroencephalogram classifier with external cohort validation ' , The Lancet Digital Health , vol. 4 , no. 12 , pp. e884-e892 . https://doi.org/10.1016/S

http://hdl.handle.net/10138/355183
https://doi.org/10.1016/S2589-7500

An automated bedside measure for monitoring neonatal cortical activity: a supervised deep learning-based electroencephalogram classifier with external cohort validation

Saeed Montazeri Moghadam, Manu Airaksinen, Päivi Nevalainen, Viviana Marchi, Lena Hellström-Westas, Nathan J Stevenson, Sampsa Vanhatalo

Summary

Background Electroencephalogram (EEG) monitoring is recommended as routine in newborn neurocritical care to facilitate early therapeutic decisions and outcome predictions. EEG's larger-scale implementation is, however, hindered by the shortage of expertise needed for the interpretation of spontaneous cortical activity, the EEG background. We developed an automated algorithm that transforms EEG recordings to quantified interpretations of EEG background and provides simple intuitive visualisations in patient monitors.

Methods In this method-development and proof-of-concept study, we collected visually classified EEGs from infants recovering from birth asphyxia or stroke. We used unsupervised learning methods to explore latent EEG characteristics, which guided the supervised training of a deep learning-based classifier. We assessed the classifier performance using cross-validation and an external validation dataset. We constructed a novel measure of cortical function, brain state of the newborn (BSN), from the novel EEG background classifier and a previously published sleep-state classifier. We estimated clinical utility of the BSN by identification of two key items in newborn brain monitoring, the onset of continuous cortical activity and sleep-wake cycling, compared with the visual interpretation of the raw EEG signal and the amplitude-integrated (aEEG) trend.

Findings We collected 2561 h of EEG from 39 infants (gestational age 35·0–42·1 weeks; postnatal age 0–7 days). The external validation dataset included 105 h of EEG from 31 full-term infants. The overall accuracy of the EEG background classifier was 92% in the whole cohort (95% CI 91–96; range 85–100 for individual infants). BSN trend values were closely related to the onset of continuous EEG activity or sleep-wake cycling, and BSN levels showed robust difference between aEEG categories. The temporal evolution of the BSN trends showed early diverging trajectories in infants with severely abnormal outcomes.

Interpretation The BSN trend can be implemented in bedside patient monitors as an EEG interpretation that is intuitive, transparent, and clinically explainable. A quantitative trend measure of brain function might harmonise practices across medical centres, enable wider use of brain monitoring in neurocritical care, and might facilitate clinical intervention trials.

Funding European Training Networks Funding Scheme, the Academy of Finland, Finnish Pediatric Foundation (Lastentautiensäätiö), Aivosäätiö, Sigrid Juselius Foundation, HUS Children's Hospital, HUS Diagnostic Center, National Health and Medical Research Council of Australia.

Lancet Digit Health 2022; 4: e884–92

BABA Center, Pediatric Research Center, Department of Clinical Neurophysiology, Children's Hospital, HUS imaging, HUS Diagnostic Center, University of Helsinki and Helsinki University Hospital, Helsinki, Finland (S M Moghadam MSc, M Airaksinen PhD, P Nevalainen MD, Prof S Vanhatalo MD); **Department of Physiology, University of Helsinki, Helsinki, Finland** (S M Moghadam, M Airaksinen, Prof S Vanhatalo); **Department of Developmental Neuroscience, Stella Maris Scientific Institute, IRCCS Fondazione Stella Maris Foundation, Pisa, Italy** (V Marchi MD); **Department of Women's and Children's Health, Uppsala University, Sweden** (Prof L Hellström-Westas MD); **Brain Modelling Group, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia** (N J Stevenson PhD)

Correspondence to:
Mr Saeed Montazeri Moghadam, Department of Physiology, University of Helsinki, 00290 Helsinki, Finland
saeed.montazeri@helsinki.fi

Introduction

Long-term electroencephalography (EEG) is routinely used for brain monitoring in neonatal intensive care units (NICUs).[1,2] EEG is the only method capable of accurate bedside detection of neonatal seizures and monitoring cerebral recovery from brain injury, such as hypoxic-ischaemic encephalopathy.[3] Monitoring of brain state is typically done from assessing hourly evolution of spontaneous cortical activity (known as EEG background). In infants at high risk, EEG monitoring can facilitate early therapeutic decisions, such as starting hypothermia treatment, and provides early predictions of long-term neurodevelopmental outcomes.

EEG monitoring is limited by insufficient 24-h expert interpretation of the EEG signals.[2,4] As an intermediate solution to facilitate bedside review, clinicians have often used time-compressed displays, such as amplitude-integrated EEG (aEEG[3,5,6]); however, aEEG can be sensitive to confounders[7] and still requires substantial training for a reliable interpretation. A particular challenge in acute bedside care is to objectively assess recovery of spontaneous brain activity (EEG background

Research in context

Evidence before this study

Recovery of spontaneous cortical activity during the first hours after brain injury, such as birth asphyxia, is the most accurate predictor of clinical outcome. Brain recovery after an insult in newborn infants at high risk is typically monitored by assessing electroencephalography (EEG) background activity, also reflected in amplitude-integrated EEG (aEEG). The continuous assessment of brain activity facilitates therapeutic decisions and provides early predictions of long-term neurodevelopmental outcomes. There are worldwide recommendations within the neurocritical care concept to do routine long-term brain monitoring with scalp-recorded EEG, which requires an EEG review available 24 h/day for the treating clinician. Lack of such bedside expertise, or a remotely available review service, has become a key bottleneck in meeting the brain-monitoring needs defined by the international guidelines. Several machine learning-based automated algorithms have been developed to provide clinician-like discrete classifications of the EEG background activity. It is hoped that automated algorithms might be eventually implemented in bedside EEG monitors to serve as clinical decision support systems. However, such development is halted by the gap in translation from the technically appealing machine learning solutions to practically appealing bedside solutions.

Added value of this study

Here, we present an end-to-end solution, from the raw EEG signal to an automated EEG interpretation visualised in the bedside monitor. The work combines a series of novelties: first, we used unsupervised learning for exploring latent EEG characteristics to guide in refining the EEG categories to such that genuinely exist in the EEG signals. Second, we trained a deep learning-based EEG background classifier that performs at an accuracy similar to that of the human inter-rater agreement. Third, we constructed a monitor-compatible visual trend, brain state of the newborn (BSN). BSN offers a holistic and intuitively interpretable index of neonatal cortical function between 0 and 100, including an estimate of its confidence for a feedback of BSN quality over time. Fourth, the algorithm performance was validated with an external dataset. Fifth, a series of proof-of-concept assessments were provided to show how well the BSN allows interpretation of the two key features of cerebral recovery, the onset of continuous cortical activity and the emergence of sleep-wake cycling.

Implications of all the available evidence

The BSN can be directly implemented into any medical patient monitor. BSN offers an intuitive, transparent, explainable, and quantified interpretation of cerebral recovery at high temporal resolution. An algorithmic EEG interpretation helps remove a key bottleneck in neurocritical care by providing EEG review anytime and everywhere and offers a way to harmonise clinical practices by removing the unavoidable ambiguity related to human EEG interpretations. Moreover, an objective and quantified bedside assessment of the evolving cortical function can facilitate clinical research and therapeutic trials by offering an important high precision benchmark for early outcomes.

activity),[8] which contains significant prognostic information about both acute and long-term recovery.[6,8–12] Most importantly, a bedside clinician needs to define the latency in hours from a brain injury to the recovery of continuous cortical activity or the emergence of sleep-wake cycling.[6,12] The background activity in the EEG signal or aEEG trend is classified visually using discrete categories. However, several EEG and aEEG background classifications are in use[5,8] and are characterised by an inherent ambiguity reflected by the substantial levels of inter-human disagreement.[13–15] Therefore, scaling up human resources for EEG interpretation cannot offer a sound solution for bedside brain monitoring.

An alternative solution for bedside EEG review is an automated algorithm, a clinical decision support system[16] that could provide EEG interpretation in bedside EEG monitors. Several computational classification algorithms have been developed towards this aim.[17,18] The algorithms generally perform well compared with clinician experts,[15] with deep learning-based classifiers performing somewhat better than the classifiers based on heuristic feature engineering.[17]

See Online for appendix

Bringing automated analysis algorithms to bedside implementations has been impeded by two factors. First, existing published algorithms are trained using different EEG classification systems, thus impeding their mutual comparison or ability to generalise across EEG datasets from different medical centres and recording systems.[19] Second, solutions that bring clinical value from such technical advance, such as an intuitive and transparent visualisation of classifier outputs in bedside EEG monitors, are scarce. We aimed to bridge these gaps by developing an end-to-end solution that transforms EEG recording data to a measure of brain state of the newborn (BSN)—a patient-monitor-compatible trend display of EEG background activity, which generalises across datasets and offers an intuitive, transparent, and clinically meaningful interpretation with human-level accuracy.

Methods

Study design

The EEG classifier algorithm was trained using a long-term EEG monitoring dataset with 2561 h of EEG from 39 newborn infants (figure 1). The EEG was scored by experts (appendix p 5) using background categories that correlate with both cerebral recovery from injury and clinical outcomes.[6,12] Inter-rater agreements were compared with classifier performance to assess human equivalence in accuracy.[20] We used an unsupervised

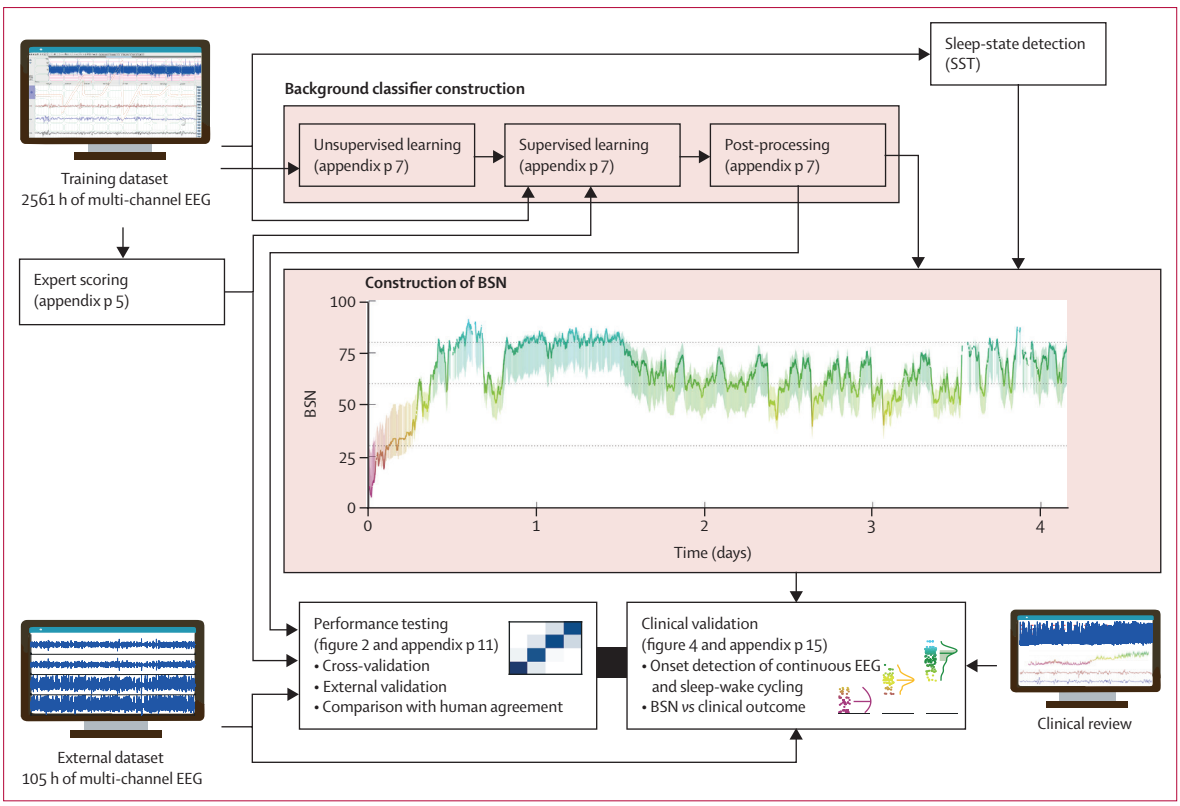


Figure 1: **Study design**
Arrows indicate the directions of data or information flow between study components. The middle graph depicts an example of BSN trend over 4 days in an infant, showing a rapid first day recovery from inactive (near zero) to a range of continuous brain activity, and the gradual emergence of rhythmic sleep-state fluctuations towards the end of the second day. BSN=brain state of the newborn. EEG=electroencephalogram. SST=sleep-state trend.

learning method, contrast-predictive coding,[21] to analyse class separation of latent signal characteristics in different time windows, which helped to guide the refining of the EEG categories and advised the subsequent classifier design. We trained a supervised learning method that was based on convolutional neural networks, a commonly used deep learning architecture, to recognise EEG background activity in 1-min epochs. We then combined output of the EEG background classifier with sleep-state trend, a deep learning-based trend display of the prediction of active versus quiet sleep states.[22] These together yielded a novel measure, BSN. BSN is a trend value that ranges from 0 to 100, containing intuitive and transparent information about both cerebral recovery and sleep-wake cycling. Finally, we did small-scale proof-of-concept experiments to assess how well the BSN trend allows recognition of the emergence of continuous EEG activity or sleep-wake cycling, the two key indices in NICU EEG monitoring.[6,8,12]

EEG datasets and scoring

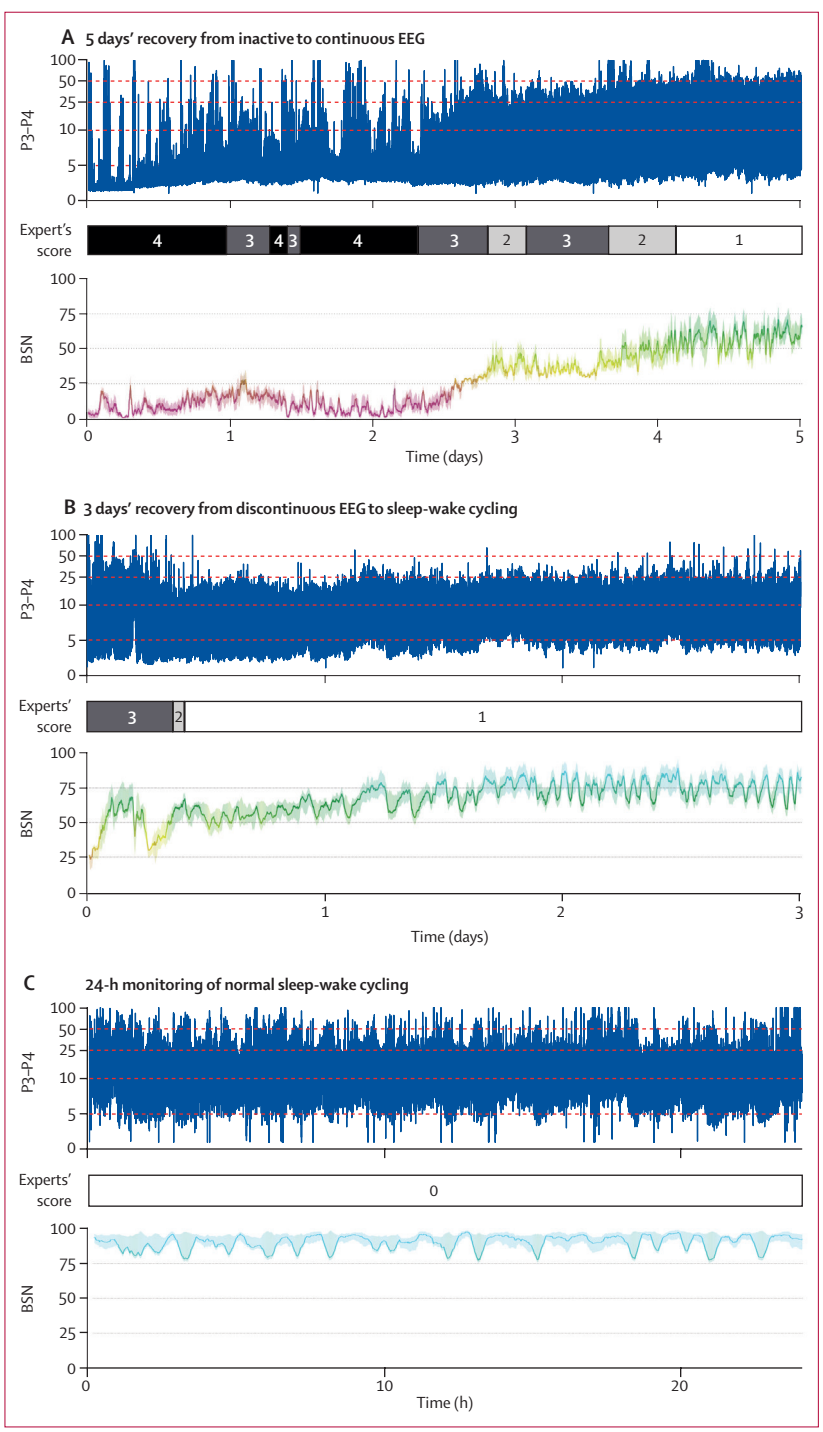
We collected the EEG recordings for the classifier training at a tertiary-level NICU at Helsinki University Hospital, Helsinki, Finland, using a four-electrode recording configuration, which is common in neonatal EEG monitoring (appendix p 4). This dataset was collated from previously published clinical cohorts[23,24] and jointly represents EEG background activity from the most severe state, an inactive EEG, to clinically normal cerebral activity, a fully continuous EEG.

We assessed the generalisation of the classifier using an external, publicly available dataset of 105 multi-channel EEG files (each 1-h duration, recorded in NICU from 31 newborn infants with hypoxic-ischaemic encephalopathy at Cork University Maternity Hospital, Cork, Ireland.[25] Processing of EEG signals was the same as described for the training data.

The EEG recording system was the same for all data (NicoletOne [Cardinal Healthcare/Natus, WI, USA]). The training dataset included frontal and parietal electrodes (F3, F4, P3, and P4) and the external validation dataset included frontal and central electrodes (F3, F4, C3, and C4); this minor difference in electrodes was considered to be beneficial for testing robustness of the algorithms.

To train a classifier for discrete EEG background scores, we used consensus expert scores from a four-category scoring system that ranges from an inactive EEG to a recovered, continuous EEG with normal sleep-wake cycling.[8] EEG scoring was done by board-certified experts PN, VM and SV, and a background

category was assigned to each hour of EEG recording. Individual expert scores were used to compare classifier performance with human inter-rater agreement. We refined the annotations as advised by the unsupervised learning method analysing score separation (appendix p 5). We also scored the external validation dataset using a four-category scoring system,[11] representing clinically similar categories with somewhat different descriptions (appendix p 5).



Construction of the EEG background classifier

To assess the separation between EEG background scores at the desired time resolution, we first used an unsupervised learning method, contrast-predictive coding[21] (appendix p 7). This analysis confirmed, as expected, less separation between categories representing continuous EEG activity with versus without sleep-wake cycling when analysing any time window between 1 and 10 min (appendix pp 9–10); therefore, these categories were combined. We then trained the classifier with 1-min window length to improve temporal resolution of the final solution.

Next, we did supervised training with consensus scores using a deep learning-based classifier (appendix p 7) that was based on convolutional neural networks. The final network architecture was adapted from a previous study[22] and optimised via an iterative process that emphasised accepting a single-channel EEG as input. Thus, the output of the classifier is the probability of EEG classes computed from 1-min epochs of single-channel EEG data.

Construction of the BSN index

Visualising the algorithmic output in bedside monitors is essential for its clinical value, and trends have long been used to display vital signs or EEG characteristics, such as aEEG,[3,5,6] seizure detection,[26,27] and sleep stages.[22] We constructed an intuitive BSN trend that takes probabilities of all background grades for each minute of EEG recording and combines these probabilities with a sleep-state classifier. A pilot visualisation of this kind was well received by clinicians.[15] The result is BSN, which is a continuous scale for EEG background activity. In the BSN scale, 0 corresponds to an inactive EEG, and the range up to 100 represents a gradual improvement through burst suppression, various degrees of declining discontinuity, until a fully continuous EEG that is typically seen in active sleep or

***Figure 2*: Comparison of BSN output with aEEG trends and expert scores in three typical cases of clinical monitoring**

(A) 5 days of EEG monitoring in an infant recovering from birth asphyxia, showing initially inactive EEG and a gradual emergence of continuity towards the end of the third day. (B) 3 days of EEG monitoring in an infant recovering from birth asphyxia, showing initially discontinuous EEG with emerging sleep-wake cycling during the second day. (C) 24 h of EEG monitoring in an infant during recovery from stroke. The BSN trend indicates that the EEG background activity is continuous with sleep-wake cycling, which was also confirmed in the normal EEG. In all three examples, the aEEG trend is depicted for the biparietal (P3–P4) derivation. The corresponding BSN trend is depicted with a solid line and the shadow around the BSN line shows the classifier uncertainty to provide the clinician with an index of classifier quality. The uncertainty is quantified by the distribution of the probability outputs of the classifier. The experts' scores of raw EEG signals are shown for comparison. The conventional aEEG views with 4 h per window of these example recordings are shown in the appendix (pp 18–32). aEEG=amplitude-integrated EEG. BSN=brain state of the newborn. EEG=electroencephalography.

wake states (appendix p 13). Notably, this combination of classifiers also allows monitoring of sleep-state fluctuations.

Moreover, BSN can be visualised as a continuous signal trend accompanied with shading around the trend to indicate the classifier's confidence on the given BSN value. The shading depicts the centres of classifier output probability mass left above and below the BSN trend (figures 1, 2; appendix pp 13–14). Visualising confidence aims to facilitate transparency and quality assessment, which might be essential for the bedside clinician using a clinical decision support system.[16]

Testing classifier performance

Classifier performance was estimated first by leave-one-subject-out cross-validation within the full training dataset (2561 h; 987–7800 epochs per infant) against consensus expert scorings using five performance metrics: confusion matrices, accuracy, unweighted F1 score (arithmetic mean of all per-class F1 scores), precision (positive predictive value) and Cohen's linearly weighted κ. We assessed generalisation with the same metrics computed from the external validation dataset. These measures reflect group-level results; however, clinical utility of the solution depends on reliability at the individual level, therefore we also computed performance metrics for each individual (appendix p 11). Additionally, we assessed the equivalence of the classifier to a human expert by comparing the agreement between classifier and human experts with the inter-rater agreement between two human experts.

Clinical proof-of-concept validation

A key advance in the BSN trend is the continuous value that allows temporally accurate tracking of cerebral recovery as well as fluctuation of vigilance states. Clinical validation of BSN is, however, challenged by a lack of ground truth with comparable fidelity because clinical conventions have been based on coarse, discrete EEG categories. Hence, there is no ground truth measure available for benchmarking BSN in cerebral function assessment. As an indirect way to provide proof-of-concept clinical assessment for the potential information value of the BSN trend, we did four small-scale experiments. Clinical experts annotated EEG recordings (PN [E1] and SV [E3]) or aEEG trends (LH-W [E4]), and BSN signals (E1 and E3). Two experts (E1 and E3) have over 15 years' experience in clinical reviewing of neonatal EEG records together with aEEG review, and one expert (E4) is a pioneer in developing existing aEEG paradigms. The filenames were randomised and the experts annotated them independently.

For the first experiment, we studied the range of BSN values at the time when human experts indicated an onset of continuous EEG or emergence of sleep-wake cycling. In the second experiment, we compared BSN levels with the well established key aEEG categories: inactive, burst suppression, and continuous normal voltage.[5,6] The aEEG categories were identified from ten randomly selected timepoints in each infant and were compared with the mean BSN levels during that hour. In the third experiment, we compared how well the BSN versus EEG can be used for identifying the postnatal age at which the cortical (EEG) activity becomes continuous activity or when sleep-wake cycling emerges. E1 and E3 defined these timepoints from the randomised BSN trends without knowing the corresponding EEG. Fourth, the full time courses of BSN trends were compared with the clinical outcome available in 25 infants in the training dataset.

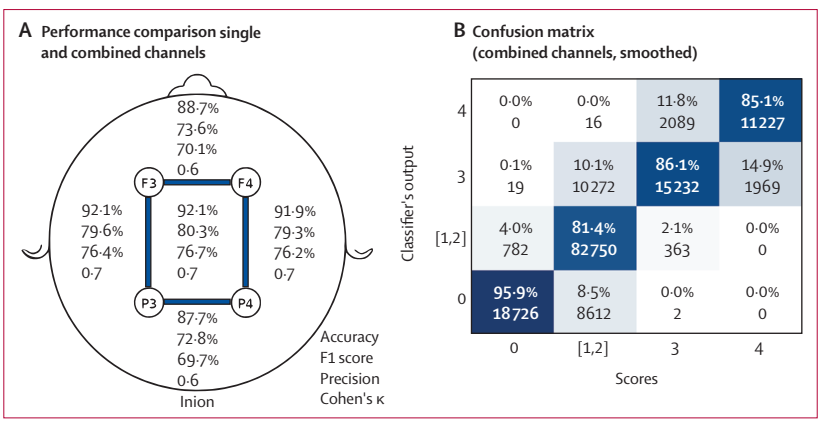


Figure 3: **Performance of the EEG background classifier**
(A) Performance comparison between single EEG derivations (blue lines) and the output after post-processing (results from the combined channels shown in the middle). (B) Confusion matrix of the EEG classifier on the y-axis and the expert scores on the x-axis. The percentages (and corresponding colours) denote the recall value of each category. The integer values denote the number of 1-min segments. EEG=electroencephalogram.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, writing of the report, or the decision to submit for publication.

Results

Comparison of the EEG background classifier outputs with the human experts' consensus score shows a high agreement, with overall accuracy ranging from 87·7% to 92·1% between EEG derivations (figure 3A; appendix pp 11–12). Combining classifier results from multiple EEG channels led to a minor but expected improvement in the classifier performance (figure 3B; appendix pp 11–12), and the accuracy was similar across the range of gestational ages (35·0–42·1 weeks; appendix p 11).

Closer inspection of the confusion matrix showed only a low (2–15%) confusion between neighbouring categories, and a negligible confusion (<1%) with categories further away from the target (figure 3B). Classifier accuracy was robust to individual variations, with all infants in our training dataset showing a clinically useful level of accuracy greater than 85% (appendix p 11).

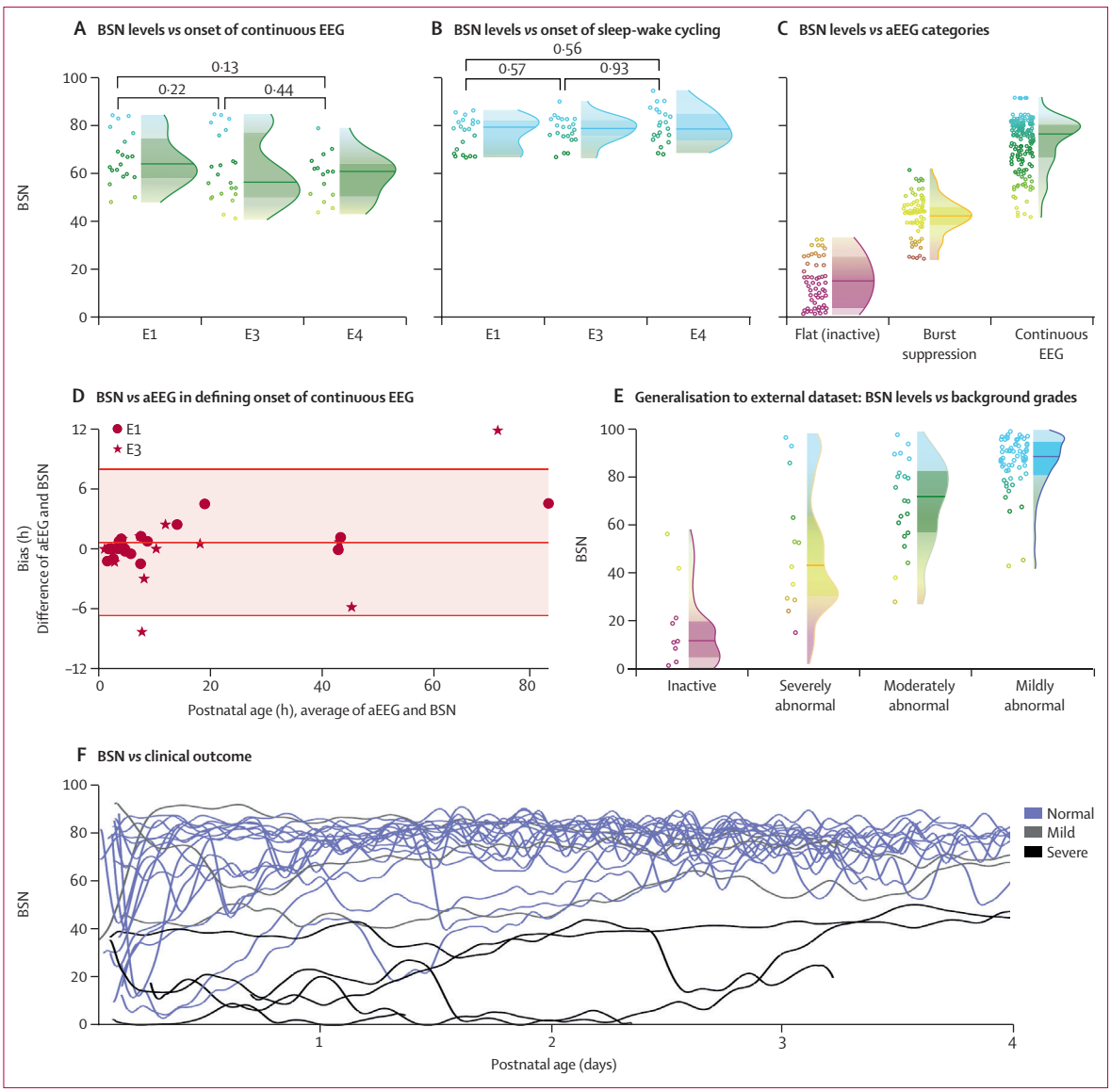


Figure 4: **Clinical proof-of-concept validation experiments**
(A) Comparison of BSN levels at the time when human experts defined the onset of continuous cortical activity. (B) Comparison of BSN levels at the time when human experts defined the onset of sleep-wake cycling. (C) Comparison of BSN levels during three different aEEG categories. (D) Bland-Altman analysis of the differences in time when human experts defined continuous EEG onset on the basis of the EEG signal versus BSN trend. Only those points are shown where EEG recovered from discontinuous to continuous during the inspection time (E1 n=22 and E3 n=23). The grey zone depicts limits of agreement (–6·6 to 7·8 h). (E) Comparison of the BSN values and EEG categories in the external validation dataset. Note the clear separation between EEG categories, whereas the two middle scores also show substantial variation across the BSN scale. The dots represent average BSN values of the given EEG hour, while the histograms show the distribution of BSN values for all minutes within these hours. All minute-wise BSN values are shown in the appendix (p 15). The BSN values are depicted with corresponding colours to allow an easier comparison between figures. (F) Full time courses of the BSN trends in individual infants during the first 4 postnatal days. The trends are colour coded according to clinical outcomes and smoothed using a median filter with 1-h window size. Note the clear separation between infants with severe versus other outcomes during the second day of life. aEEG=amplitude-integrated EEG. BSN=brain state of the newborn. E1–4=expert 1–4. EEG=electroencephalography.

We also compared the classifier's performance with human-to-human inter-rater agreement (appendix p 10) in distinguishing between three EEG background scores. The κ level between two human experts (κ=0·83 for E1 *vs* E2) was similar to the κ levels between classifier and human experts (κ=0·73 [minute level] and κ=0·79 [hour level] for E1 *vs* classifier; κ=0·77 [minute level] and κ=0·84 [hour level] for E2 *vs* classifier; appendix p 10). This finding suggests performance equivalent to near-human level—ie, a human expert could be replaced by the classifier without major loss of agreement.

We tested classifier generalisation using an external dataset from a different centre using a slightly different

recording setting and scoring system. Despite these differences, the overall accuracy in classification was 88·0% (weighted Cohen's κ=0·72; range 0·61–1·00 for individual infants) with little confusion between categories, suggesting a good generalisation (appendix p 12).

The overall cerebral recovery from inactive to fully continuous activity can be readily observed over multiple days with no previous training in BSN interpretation (figure 2A). Likewise, a gradual emergence of sleep-wake cycling over 3 days can be easily observed in an infant with gradual recovery from birth asphyxia (figure 2B). Finally, an example from an infant monitored for stroke (figure 2C) shows how the overall background stayed clearly higher than in the infants with hypoxic-ischaemic encephalopathy (figure 2B), and also exhibits clear rhythmicity in the sleep states.

In the first proof-of-concept experiment, we assessed the use of BSN trend in monitoring key indicators of brain recovery from injury (figures 4A, B). The onset of continuous EEG or emergence of sleep-wake cycling showed a high inter-rater agreement between experts (appendix p 16).

In the second experiment, we compared BSN values with well established aEEG background categories[28] to search for an easy transfer from an aEEG-based review to a BSN-based review. We found a clear separation in BSN levels according to aEEG categories: an inactive aEEG corresponded to a BSN of 0–33 (median 14·8; 95% CI 12·8–18·4), burst suppression corresponded to a BSN of 27–61 (41·8; 38·4–43·4); and a continuous normal voltage aEEG corresponded to a BSN of 61–91 (75·8; 70·4–73·6). Notably, these ranges were bounded by the scores 1–2, 3, and 4 that were only available for this experiment. Higher values with a BSN of greater than 90 were often observed in the more normal appearing EEG (figure 2C).

In the third experiment, we compared the BSN with the aEEG for identifying the postnatal age when background activity becomes continuous or shows sleep-wake cycling. The mean difference between the BSN and the EEG in defining onset of continuous activity was 1 h (of 27 neonates, E1 did not find onset of continuous EEG in five neonates and E3 did not find onset of continuous EEG in four neonates; appendix p 4). We found no systematic bias (mean bias 0·6 h [95% CI –0·5 to 1·7]), suggesting that BSN reliably indicated the onset of continuous EEG activity (figure 4D). The emergence of sleep-wake cycling, however, was more challenging to detect (appendix p 15). We found only a minimal systematic bias between aEEG or EEG versus BSN readings (mean bias 5·5 h [1·5 to 9·4). However, limits of agreement in the Bland-Altman analysis were wider (appendix p 16) because of the inherent ambiguity of sleep-wake cycling onset detection, seen as larger variance between experts in both reading the EEG or aEEG and the BSN signals (appendix p 15).

In the external dataset, BSN values were robustly different between EEG background categories (figure 4E). When the BSN values were computed for every minute of EEG and compared to the hourly EEG category, we found substantial variation in the BSN levels, reflecting the well known temporal dynamics in cortical activity (appendix p 15).

Finally, we compared the continuous BSN time courses of individual infants during the first 4 postnatal days with their later clinical outcomes. A clear BSN trend separation between infants with severe (cerebral palsy or death) versus other outcomes became apparent during the first 2 days of life and remained throughout the observation period (figure 4F).

Discussion

We showed that an automated review of neonatal EEG monitoring is possible using a visual display of an algorithmic EEG interpretation, such as BSN. The present findings show that BSN allows detection of the onset of continuous EEG activity and sleep-wake cycling, the two key parameters in newborn neuromonitoring.[6,12] We also showed that the results generalise to an external validation dataset, and the evolution of an individual's BSN levels correlates with clinical outcomes.

Previous studies have presented several algorithmic solutions for classifying discrete EEG classes using machine learning-based methods.[17,18,29] The classification performance of these solutions is generally similar to that of our present work; however, a direct comparison is impeded by several differences that revolve around the clinical benchmark used for both the training and the testing phase. First, studies from different centres use somewhat different EEG scoring systems and little or no information is available on these systems' comparability or ambiguity, which could be measured by inter-rater agreement.[20] Second, the EEG classifiers typically combine information from multiple EEG channels,[17,18] an approach which is often chosen to improve classifier performance; however, classifiers based on multi-channel data have only limited clinical utility because the newborn EEG is typically recorded with only a few channels, many of which might need rejection because of various artifacts.[5,6] Third, using even 1-h long epochs for background classifications is common to improve classifier stability at the cost of ignoring the well known and clinically meaningful temporal dynamics in the brain states of a critically ill patient.[18] Here, we present an EEG background classifier that shows an accuracy similar to that of previous studies and human experts, and we show that this accuracy is achievable even from a few minutes of individual EEG signals.

In clinical practice, clinical observations such as EEG assessment are often categorised into discrete classes. Such an approach is understandable because of the perceptual limits of a human observer; however, states in brain activity, or the EEG signal thereof, represent a full continuum without discrete switches between categories, which is clear to the bedside clinicians who observe

gradual brain recoveries through the whole spectrum of states.[6,8–12] Use of discrete categories leads to substantial ambiguity at category boundaries, which is repeatedly shown as substantial inter-rater disagreement in EEG classifications.[13–15] Classifying different kinds of discontinuity is a prime example of the challenge: depending on the context, the graphologically similar EEG waveforms might reflect a wide range of conditions from a severe cerebral compromise in burst suppression to an ongoing cerebral recovery after injury, or a physiological discontinuity related to prematurity or quiet sleep. In the aEEG practice, some of this ambiguity has been overcome by recognising a discontinuous category with and without burst suppression.[28] Here, we solved this multidimensional challenge by visualising the output of the multi-class EEG classifier using a continuous BSN measure. The BSN intentionally avoids a discrete categorisation, yet our proof-of-concept validation studies suggest that the BSN values are correlated with the conventional EEG categories to support transparency and explainability.[30]

For an easier bedside implementation, the BSN trend was designed to resemble the common trends in the vital sign monitors, or the widely known aEEG trend in existing EEG monitors. Although BSN and aEEG trends look somewhat similar, essential differences exist. First, aEEG is a straightforward measure of signal amplitude, but BSN is an interpretation of many EEG signal characteristics that might readily escape visual recognition of EEG waveforms.[15] Therefore, BSN might represent far more clinical information content while being less sensitive to commonplace amplitude-based artifacts. Second, the BSN trend has an estimate of confidence; intuitively, the BSN trend indicates to the bedside clinician the level of trust at each point in time, providing essential information about monitoring quality. Third, whereas aEEG still requires training and remains a subjective assessment of discrete aEEG categories with substantial ambiguity,[5,6,14] the BSN trend provides an objective measure with high temporal resolution to support quantitative comparison over time and across centres.

Our work has some limitations. The background classifier could be trained with much larger datasets to possibly improve classifier performance. Larger datasets from more international study centres are needed to fully validate the accuracy of the classifier solution in different user scenarios. However, the BSN trend combines information from the EEG background classifier and the sleep-state classifier; thus, ultimate clinical validation of BSN cannot be inferred directly from metrics of classifier performance. Our clinical validation experiments can only be considered as proof of principle because of the small sample size and exploratory nature. Prospective studies with external datasets are needed to validate the true clinical feasibility and utility of the present solution. Also, our background classifier was not trained on preterm infants, or to recognise focal details, such as hemispheric asymmetry or abnormal waveforms which require conventional multi-channel recordings.[8,9,11,12] Future studies among clinical bedside users are needed to optimise visualisation of the BSN, and to assess the need for presenting sleep state as a complementary trend (appendix p 14). Finally, this work does not assess how BSN performs in the presence of seizures, which might cover a substantial proportion of newborn EEG records in infants treated in NICUs.[26,27] In neonates with suspected seizures, we propose using an automated seizure detector before computing BSN, a solution that is already implemented in our openly available cloud computation server (appendix p 17).

BSN is designed to give a numerical measure with a quantitative meaning without an explicit categorical interpretation of the underlying EEG signal. Studies have emphasised and cautioned about explainability of machine learning solutions as a shortcut to clinical validation.[30] We suggest that a genuine clinical utility of BSN should not be validated by direct comparison with existing discrete EEG interpretations; the utility and clinical deployment of a BSN-based EEG interpretation needs rigorous and thorough validation in various real-world scenarios, including prospective validation studies that include BSN-based treatment decisions.[27] This is possible through the open access to BSN trend via computational cloud service (appendix p 17), which is hoped to expedite clinical validation work and enable direct comparison with other solutions. Taken together, BSN overcomes many of the key bottlenecks in the routine and clinical trial use of neonatal EEG. When combined with automated seizure detection[26,27] BSN could become a key component of future clinical decision support systems in neonatal neurocritical care.

Contributors

SMM and SV conceived the experimental design. PN, VM, and SV did the visual annotation. SM developed and implemented the contrast-predictive coding and the convolutional neural networks classifier designs and analysed the EEG data. PN, SV, and LH-W did the visual annotation of aEEG and BSN trends for the clinical proof-of-concept validation. NJS and MA validated the statistical results. SM and SV prepared the manuscript and figures. All authors carefully reviewed, commented on, and approved the final version. SMM and SV had final responsibility for the decision to submit for publication. PN, VM and SV had access to all the data, and LH-W had access to the trend data.

Declaration of interests

We declare no competing interests.

Data sharing

Additional information related to this study is available on request to the corresponding author. Data used to train and validate the background classifier will be made available to interested research partners on reasonable request to SV or PN; the prerequisite for data sharing is a data transfer agreement, approved by the legal departments of the requesting researcher and by all legal departments of the institutions that provided data for the study, and ethics clearance. The external validation dataset that we used in this study is publicly available. We provide the BSN solution available via our cloud service (https://babacloud.fi/). The cloud interface needs credentials that are available at request from the corresponding author. The system will not store the EEG files, and the user is encouraged to use only pseudonymised files for maximal data protection.

Acknowledgments

This work was supported by the EU under the H2020-EU.1.3.1 Program (European Training Networks Funding Scheme; INFANS Project 813483) Research and Innovation Programme (Marie Skłodowska-Curie grant agreement) and the Academy of Finland (335788, 332017, and 343498).

References

1 Abend NS, Mani R, Tschuda TN, et al. EEG monitoring during therapeutic hypothermia in neonates, children, and adults. *Am J Electroneurodiagn Technol* 2011; **51:** 141–64.

2 Boylan G, Burgoyne L, Moore C, O'Flaherty B, Rennie J. An international survey of EEG use in the neonatal intensive care unit. *Acta Paediatr* 2010; **99:** 1150–55.

3 de Vries LS, Hellström-Westas L. Role of cerebral function monitoring in the newborn. *Arch Dis Child Fetal Neonatal Ed* 2005; **90:** F201–07.

4 Tsoi K, Yam KKM, Cheung HM, et al. Improving consistency and accuracy of neonatal amplitude-integrated electroencephalography. *Am J Perinatol* 2021; published online Dec 7. https://doi.org/10.1055/a-1677-9872.

5 Dilena R, Raviglione F, Cantalupo G, et al. Consensus protocol for EEG and amplitude-integrated EEG assessment and monitoring in neonates. *Clin Neurophysiol* 2021; **132:** 886–903.

6 Chalak L, Hellstrom-Westas L, Bonifacio S, et al. Bedside and laboratory neuromonitoring in neonatal encephalopathy. *Semin Fetal Neonatal Med* 2021; **26:** 101273.

7 Deshpande P, McNamara PJ, Hahn C, Shah PS, Guerguerian AM. A practical approach toward interpretation of amplitude integrated electroencephalography in preterm infants. *Eur J Pediatr* 2022; **181:** 2187–200.

8 Walsh BH, Murray DM, Boylan GB. The use of conventional EEG for the assessment of hypoxic ischaemic encephalopathy in the newborn: a review. *Clin Neurophysiol* 2011; **122:** 1284–94.

9 Watanabe K, Hayakawa F, Okumura A. Neonatal EEG: a powerful tool in the assessment of brain damage in preterm infants. *Brain Dev* 1999; **21:** 361–72.

10 Menache CC, Bourgeois BF, Volpe JJ. Prognostic value of neonatal discontinuous EEG. *Pediatr Neurol* 2002; **27:** 93–101.

11 Murray DM, Boylan GB, Ryan CA, Connolly S. Early EEG findings in hypoxic-ischemic encephalopathy predict outcomes at 2 years. *Pediatrics* 2009; **124:** e459–67.

12 Shellhaas RA, Chang T, Tsuchida T, et al. The American Clinical Neurophysiology Society's guideline on continuous electroencephalography monitoring in neonates. *J Clin Neurophysiol* 2011; **28:** 611–17.

13 Wusthoff CJ, Sullivan J, Glass HC, et al. Interrater agreement in the interpretation of neonatal electroencephalography in hypoxic-ischemic encephalopathy. *Epilepsia* 2017; **58:** 429–35.

14 Bourgoin P, Barrault V, Loron G, et al. Interrater agreement between critical care providers for background classification and seizure detection after implementation of amplitude-integrated electroencephalography in neonates, infants, and children. *J Clin Neurophysiol* 2020; **37:** 259–62.

15 Moghadam SM, Pinchefsky E, Tse I, et al. Building an open source classifier for the neonatal EEG background: a systematic feature-based approach from expert scoring to clinical visualization. *Front Hum Neurosci* 2021; **15:** 675154.

16 Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit Med* 2020; **3:** 17.

17 Raurale SA, Boylan GB, Mathieson SR, Marnane WP, Lightbody G, O'Toole JM. Grading hypoxic-ischemic encephalopathy in neonatal EEG with convolutional neural networks and quadratic time-frequency distributions. *J Neural Eng* 2021; **18:** 046007.

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

19 Stevenson NJ, Oberdorfer L, Tataranno ML, et al. Automated cot-side tracking of functional brain age in preterm infants. *Ann Clin Transl Neurol* 2020; **7:** 891–902.

20 Airaksinen M, Räsänen O, Ilén E, et al. Automatic posture and movement tracking of infants with wearable movement sensors. *Sci Rep* 2020; **10:** 169.

21 Banville H, Chehab O, Hyvärinen A, Engemann DA, Gramfort A. Uncovering the structure of clinical EEG signals with self-supervised learning. *J Neural Eng* 2021; **18:** 046020.

22 Montazeri Moghadam S, Nevalainen P, Stevenson NJ, Vanhatalo S. Sleep state trend (SST), a bedside measure of neonatal sleep state fluctuations based on single EEG channels. *Clin Neurophysiol* 2022; **143:** 75–83.

23 Nevalainen P, Metsäranta M, Marchi V, Toiviainen-Salo S, Vanhatalo S, Lauronen L. Towards multimodal brain monitoring in asphyxiated newborns with amplitude-integrated EEG and simultaneous somatosensory evoked potentials. *Early Hum Dev* 2021; **153:** 105287.

24 Nevalainen P, Metsäranta M, Toiviainen-Salo S, Lönnqvist T, Vanhatalo S, Lauronen L. Bedside neurophysiological tests can identify neonates with stroke leading to cerebral palsy. *Clin Neurophysiol* 2019; **130:** 759–66.

25 O'Toole JM, Mathieson SR, Magarelli F, Marnane WP, Lightbody G, Boylan GB. Neonatal EEG graded for severity of background abnormalities. Zenodo 2022; published online May 27. https://doi.org/10.5281/zenodo.6587973.

26 Tapani KT, Nevalainen P, Vanhatalo S, Stevenson NJ. Validating an SVM-based neonatal seizure detection algorithm for generalizability, non-inferiority and clinical efficacy. *Comput Biol Med* 2022; **145:** 105399.

27 Pavel AM, Rennie JM, de Vries LS, et al. A machine-learning algorithm for neonatal seizure recognition: a multicentre, randomised, controlled trial. *Lancet Child Adolesc Health* 2020; **4:** 740–49.

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

29 Ahmed R, Temko A, Marnane W, Lightbody G, Boylan G. Grading hypoxic-ischemic encephalopathy severity in neonatal EEG using GMM supervectors and the support vector machine. *Clin Neurophysiol* 2016; **127:** 297–309.

30 Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health* 2021; **3:** e745–50.