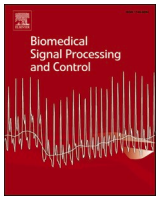
Biomedical Signal Processing and Control 108 (2025) 107904

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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



Towards unobtrusive sleep stage classification in preterm infants using machine learning

Nathalie Demme [a,c], Maha Shadaydeh [c,*], Laura Schieder [a,b], Claus Doerfel [b], Stella Jähkel [b], Knut Holthoff [a], Hans Proquitté [b], Joachim Denzler [c], Jürgen Graf [a,*]

[a] Department of Neurology, Jena University Hospital, Jena, Germany
[b] Department of Pediatric and Adolescent Medicine, Jena University Hospital, Jena, Germany
[c] Computer Vision Group, Friedrich Schiller University Jena, Jena, Germany

ARTICLE INFO

Keywords:
Automated sleep stage classification
Preterm infants
Sleep
Machine learning
Support vector machine

ABSTRACT

In the neonatal intensive care unit (NICU), preterm infants are usually unable to fulfil their sleep demands due to frequent disruptions. Real-time sleep monitoring could be an essential tool not only to shift elective care to their wake periods but also to track their developmental sleep profile as an indicator of healthy brain maturation. The current gold standard for sleep measurement, polysomnography, is invasive and labour-intensive, limiting its applicability for continuous monitoring. We propose an automatic sleep stage classification method using only the routinely available electrocardiogram (ECG) and patient movement data recorded with a piezo mat. For this study we recorded data from 28 preterm infants (13 females and 15 males) at 35.7 ± 0.5 weeks postmentrual age. We employed a support vector machine (SVM) to classify sleep stages into wakefulness (W), active sleep (AS), and quiet sleep (QS). The combined piezo + ECG model demonstrated superior accuracy (92 %) and strong agreement with expert annotations (Cohen's kappa = 0.83) compared to ECG-only or piezo-only models. This approach offers a reliable, unobtrusive solution for continuous sleep monitoring in NICUs, facilitating individualised, sleep-based medical care for preterm infants.

1. Introduction

According to the latest statistical surveys, approximately 10 % of all newborns are born prematurely (<37 weeks of gestation) [1]. Preterm birth significantly increases the risk of developing a wide range of complications. Neurological disorders include sensory perception disorders (up to 50 % of disorders), learning disabilities, depression, and attention deficit hyperactivity disorders [2–4]. Consequently, in the NICU, emphasis should be placed not only on ensuring survival but also on reducing the risk for sequelae. One reliable indicator of normal brain development in neonates is the establishment of normal sleep cycles [2]. Organized sleep patterns at the premature age may predict a better neurodevelopmental outcome [5,6]. However, in the NICU environment, frequent interventions and environmental factors can make it challenging for preterm infants to sleep adequately [7]. Therefore, it is necessary to track the preterm infant's sleep in order to: (1) adjust interventions and care to less disruptive times and (2) utilise sleep cycle information as a diagnostic indicator for normal brain development.

The clinical gold standard for the evaluation of preterm infants' sleep characteristics and associated events is polysomnography (PSG) [8]. It consists of the following biophysical parameters: electroencephalogram (EEG), electrooculogram, chin electromyogram, ECG, nasal airflow, respiratory effort, oxygen saturation, leg and arm movement, and body position. Despite the variety of biosignals measured during PSG, the classification of sleep stages remains reliant on manual assessment by clinicians [9].

In this paper, we propose a method for the automatic monitoring of preterm infants' sleep stages using as few biosignals as possible. We aim for a three-stage classification of W, AS, and QS. We utilised an SVM and trained three different classifiers. One utilises features from both the piezo mat and ECG, and two others use features from either the piezo mat or ECG exclusively. The proposed approach ensures reliable sleep stage detection, even if one signal source is unavailable. The combined piezo + ECG model, with a mean kappa of 0.83, outperformed the models using either ECG or piezo mat, which achieved kappas of 0.75 and 0.73, respectively. To our knowledge, this is the first unobtrusive

* Corresponding authors at: Am Klinikum 1, 07747 Jena, Germany (J. Graf). Ernst-Abbe-Platz 2, 07743 Jena (M. Shadaydeh).
E-mail addresses: maha.shadaydeh@uni-jena.de (M. Shadaydeh), juergen.graf@med.uni-jena.de (J. Graf).

https://doi.org/10.1016/j.bspc.2025.107904
Received 7 November 2024; Received in revised form 6 March 2025; Accepted 14 April 2025
Available online 26 April 2025

approach that achieved human expert-level performance in measuring all sleep stages in preterm infants.

1.1. Related work

Over the past decades, numerous attempts have been made to develop algorithms for automatic sleep stage classification [9]. Most of them rely on EEG as it contains important information about brain activity that defines sleep stages [10]. This approach may facilitate the work of clinicians when they analyse PSG recordings, but it is not suitable for continuous monitoring for two reasons: (i) the additional use of EEG electrodes might be harmful to the fragile skin of neonates and (ii) contact-based methods might further disrupt their precious sleep [8]. Hence, unobtrusive methods for sleep stage measurement are preferred [8]. Alternative approaches have included respiration [11], ECG [12], or a combination of vital signs to classify sleep stages in preterm infants [13–15]. For a three-stage classification of W, AS, and QS, the best performance to date was achieved by Wang et al., with a Cohen's kappa of 0.52. They used a combination of heart and respiratory rate sampled at 0.4 Hz. However, the achieved accuracy is still insufficient for clinical application [16]. Another method involved video signals to monitor sleep [17–19]. While they achieved a high accuracy in classifying sleep and wake, they did not separate AS and QS. Additionally, video-based monitoring can be compromised by low light conditions or the partial covering of the infant's body by blankets or external ventilation equipment.

In this study, we combined routinely derived ECG with a movement-sensitive sensor mat (piezo mat) for automatic sleep stage classification in preterm infants. Ranta and colleagues used a similar approach, combining a bed mattress signal and ECG [20]. Even though they achieved very high accuracy, they focussed on detecting deep sleep (N3) in 1- to 18-week-old term-born patients.

2. Methods

In this section, we first present the study population and the data acquisition procedure. Then, we describe the data annotation procedures in detail. Finally, we present our approach to automated sleep stage classification in preterm infants.

2.1. Study population

The study was designed as a single-centre observational study to develop an automated sleep stage classification based on a few vital signs. Ethical approval was given by the local Ethics Committee of the Jena University Hospital (Reg.-Nr. 2022-2831_1-MV), and parents gave their written consent. The study was conducted according to good clinical practice and the Declaration of Helsinki. The data includes 13 females and 15 males, and thus a gender disparity ratio (GDR) of 1.15. The 28 preterm infants were routinely referred to the paediatric sleep lab at a postmenstrual age of 35.7 ± 0.5 weeks (Table 1). They were born at a gestational age of 30.2 ± 2.1 weeks with a birth weight of 1411 ± 248 g. The overall heterogeneity of our patient group was high because no exclusion criteria were applied.

2.2. Data acquisition

The PSG data were collected for four hours according to the rules given by the American Association of Sleep Medicine (AASM, [10,21]). The PSG data were recorded at 200 Hz with the sleep diagnostic system ALICE 6 (Löwenstein Medical) together with the Sleepware G3 software (Philips Respironics). Electrical signals were recorded with a notch filter (50 Hz) to remove the power line noise. As a contactless movement-sensitive detector, we placed a piezo-based sensor mat (Jablotron Nanny BM-02) underneath the mattress. The signal of the piezo mat was synchronously recorded with the PSG. A parallel video recording was used to visually detect the child's movements and annotate external interventions like breastfeeding and electrode repositioning.

Table 1
Patient demographics and clinical parameters. Data are presented as mean ± SD.

Parameter	Value
Female	46.4 %
Gestational age at birth	30.2 ± 2.1 weeks
Postmenstrual age at measurement	35.7 ± 0.5 weeks
Birth weight	1,411 ± 248 g
Birth length	39.3 ± 2.9 cm
Body weight at measurement	2,148 ± 227 g
Body length at measurement	44.7 ± 1.8 cm
APGAR Score (1 min)	5
APGAR Score (5 min)	8
APGAR Score (10 min)	9
Caffeine therapy	
• Yes	57.1 %
• No	25.0 %
• Paused	17.9 %
Multiple births	32.1 %

2.3. Data annotation

Two different methods for data annotation were used in this study. The first is based on the Sleepware G3 classification using PSG and piezo mat, and the second annotation method is based on the assessments of three experts.

2.3.1. Sleepware G3

The commercial software Sleepware G3 is capable of producing an automatic sleep stage classification that we refer to as Sleepware throughout the manuscript. This automatic sleep scoring is usually insufficient for preterm infants and has to be re-evaluated by human experts.

2.3.2. Manual annotation

In clinical routine, human experts visually annotate the sleep stages based on the rules given by the AASM for infants [21,22]. Clinicians use all biophysical signals recorded with PSG to score sleep stages, including brain activity (EEG), heart rhythm (ECG), respiration, and muscle movement [8,10]. Neonatal sleep is classified in epochs of 30 s consisting of three main stages: W, AS, and QS. The W stage is characterised by irregular heart rate and respiration, open eyes and body movements, and mixed EEG patterns with movement artefacts. The AS stage includes irregular respiration, low muscle tone superimposed by twitches and phasic movements, epochs of rapid eye movements (REM), and a continuous EEG pattern of 40–80 μV in amplitude. Finally, the QS sleep stage is characterised by having regular and slow respiration, tonic motor activity and startles, no eye movement, and alternant EEG signals of high amplitude. We analysed a total of 12,957 30-s epochs from 28 preterm infants. The manual annotation of sleep stages was performed by three independent human experts (C.D., S.J., L.S.). The inter-rater variability was calculated using Fleiss' kappa to evaluate their agreement. A common label is considered as ground truth for training and testing our machine learning-derived algorithm. Therefore, a unique label for each 30-s epoch is created based on the agreement in the assessment of the sleep stage of this epoch by three experts, and two datasets are generated accordingly.

- Dataset A: This set includes only those epochs where all three experts agreed. This occurred in 68 % of the epochs (8,757 epochs, approx. 73 h). To make sure we did not introduce a bias toward any sleep stage, we evaluated the relative time in each sleep stage for all data versus Dataset A (Appendix Fig. A.1A). We noticed only a minor change in sleep stage distribution.

- Dataset B: This set is defined more liberally and includes, in addition to the epochs of Dataset A, all epochs where at least two experts agreed. This data set constitutes 98.7 % of the whole data (12,787 epochs, approx. 106 h).

We aim to test the performance of both data sets to discuss the influence of the uncertainty in the labelling on the classification performance.

2.4. Automated sleep stage classification

The process of developing an automated algorithm for classifying sleep stages is illustrated in Fig. 1. The individual steps are described in detail in the following sections.

2.4.1. Data preprocessing

To reduce noise in the ECG signals and remove baseline drift, the signals were filtered in the frequency domain, where frequencies lower than 0.025 Hz and higher than 199.975 Hz were removed. The ECG signals were corrected by mirroring the inverted R-peaks so that all R-peaks point upwards.

2.4.2. Feature extraction

This section focuses on extracting features from the piezo and ECG signals, as described in the following two subsections. The features were selected to represent the different characteristics of sleep stages as summarised in Table 2. The extracted features after standardisation serve as input for the machine learning tool as they provide more

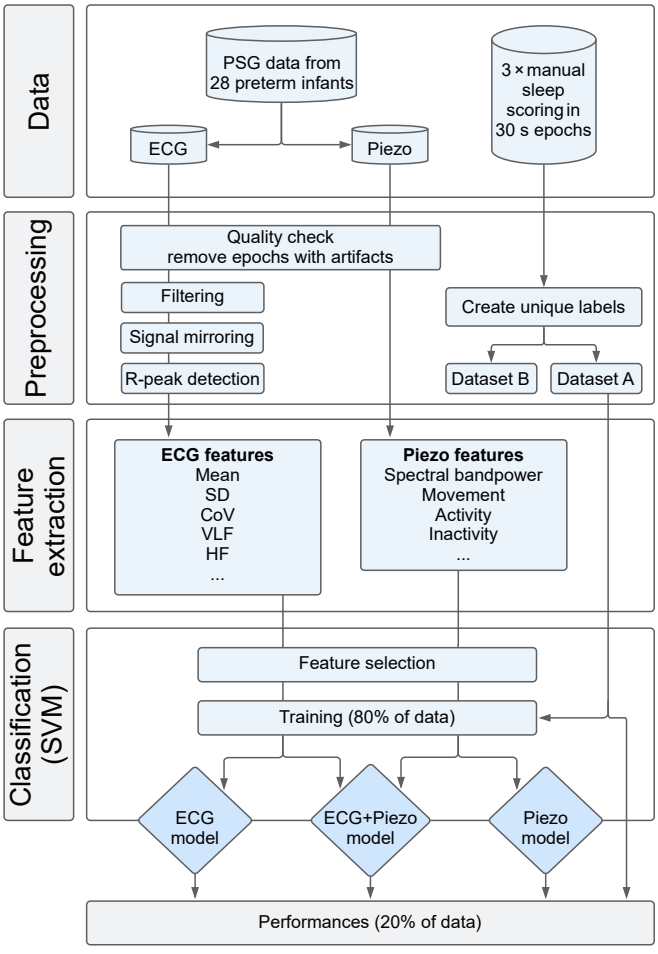


Fig. 1. Workflow for machine learning derived automatic sleep stage classification.

Table 2
Characteristics of preterm sleep stages wakefulness (W), active sleep (AS), and quiet sleep (QS), when only looking at the heart rate and movement of patients (adapted from [8,10]).

	W	AS	QS
ECG	Heart rate (HR) high, irregular	HR mostly irregular	HR mostly regular, accelerations during startles
	Heart rate variability (HRV) high	Low frequencies are dominant (0.03–0.39 Hz)	High frequencies are dominant (0.4–1 Hz)
Movement	Head and arm movements, orientation response	Wide range, small twitches, sporadic motor bursts of 5–60 s	Little or no startles, motion low

Table 3
List of extracted features that served as input for the training of an SVM.

Feature Name	Description	Used for model Piezo + ECG	Piezo	ECG
	Piezo feature			
Spectrogram	frequencies over time			
Bandpower (BP)	average power in frequency interval [2–25 Hz]			
Movement	bandpower (BP) > threshold (T) (Equation (1))			
Activity (1 min)	moving sum movement	☒	☒	
Activity (10 min)	moving sum movement	☒	☒	
Distance	maximal distance between two movements			
Inactivity (1 min)	moving sum of distance	☒	☒	
Inactivity (10 min)	moving sum of distance	☒	☒	
	ECG feature			
Mean HR (5 min)	$\frac{1}{N}\sum_{\{i=1\}}^{N} x_i$($x_i$ = the i^{th} RR interval)	☒		☒
SD HR (5 min)	$\sqrt{\left\{\frac{1}{N-1}\sum_{\{i=1\}}^{N}(x_i - mean)^2\right\}}$	☒		☒
VLF (5 min)	0.01–0.04 Hz	☒		☒
LF (5 min)	0.04–0.5 Hz			
HF (5 min)	0.5–2 Hz	☒		☒
LF/HF ratio (5 min)	LF/HF	☒		☒
TINN (5 min)	triangular interpolation of RR interval histogram	☒		☒
Mean HR (~6 s)	$\frac{1}{N}\sum_{\{i=1\}}^{N} x_i$(20 data points)	☒		☒
Mean HR (~30 s)	$\frac{1}{N}\sum_{\{i=1\}}^{N} x_i$(100 data points)	☒		☒
Median HR (~6 s)	$x\left(\frac{N+1}{2}\right)$(20 data points)	☒		☒
Median HR (~30 s)	$x\left(\frac{N+1}{2}\right)$(100 data points)	☒		☒
Diff (~3 s)	$\lvert mean_s - median_l \rvert$ (10 data points)			
SD HR (~6 s)	$\sqrt{\left\{\frac{1}{N-1}\sum_{\{i=1\}}^{N}(x_i - mean)^2\right\}}$(20 data points)	☒		☒
SD HR (~30 s)	$\sqrt{\left\{\frac{1}{N-1}\sum_{\{i=1\}}^{N}(x_i - mean)^2\right\}}$(100 data points)	☒		☒
CoV (~6 s)	SD/mean (20 data points)	☒		☒
CoV (~30 s)	SD/mean (100 data points)	☒		☒
	Time feature			
Start time [s]	starting point of each 30-s epoch	☒		☒

relevant information than the raw signals. A summary of all features is listed in Table 3.

Piezo mat features: The piezo mat records movements of the preterm infant that could arise from small or gross body movements as well as from respiration and heartbeat. To extract relevant information from the raw signal, a spectrogram was created using a 1-second sliding time window with 50 % overlap, showing the frequency composition of the signal over time. The values of the average power of the spectrogram, hereafter referred to as the bandpower (*BP*), within the frequency range of 2 to 25 Hz, were calculated for each time window and used to detect movement by applying an individual threshold *T* on *BP* for each patient. This threshold *T* was calculated for the entire recording as:

$$T = BP_{Baseline} + C \times BP_{MAD} \quad (1)$$

The individual patient parameters are the $BP_{Baseline}$ defined as the 25th percentile of *BP*, and BP_{MAD}, the median absolute deviation (MAD) of BP defined as BP_{MAD} = median(abs(*BP* – median(*BP*))). The factor *C* was experimentally set to 4.75 for all patients. These parameters were optimised through systematic variation for automatic movement detection and validated by comparing the results to visually annotated movements in a cohort of five patients. The performance was evaluated using several metrics, including the Youden index (YI), the matching index (MI) [23], and the spike time tiling coefficient (STTC) [24]. The YI was used to evaluate the sensitivity and specificity of detecting movement periods, whereas the MI and STTC measure the correlation per data point (every 0.5 s).

The activity of the preterm infant was calculated as the moving sum of the movement vector over different periods of time (long: 5, 10, 15 min; short: 0.5, 1, 1.5 min). As an indicator of inactivity, the interval between movement periods was calculated and then summed up over one and ten minutes.

ECG features: The R-peaks as an indicator of the heartbeat were detected in the filtered ECG using the *findpeaks* function in Matlab (MinPeakProminence = 60, MinPeakDistance = 0.2, MaxPeakWidth = 0.1703, WidthReference = halfprom, MinPeakHeight = 120). Next, RR intervals (distance between R-peaks) less than two-thirds the length of the surrounding intervals were identified and removed if the previous RR interval is less than one second, as they were likely false-positive R-peaks. An example of the preprocessing steps is illustrated in Appendix Fig. A.2. The resulting RR intervals were then used as input for a heart rate variability (HRV) tool for neonates [25,26]. The tool calculated features over five-minute time windows. We set the overlap to 90 %, resulting in a value for each 30-s epoch. The calculated features include time domain features such as the average heart rate and the standard deviation reflecting the total HRV. Frequency domain features of HRV include the very low-frequency bands (VLF: 0.01–0.04 Hz), low-frequency bands (LF: 0.04–0.5 Hz), high-frequency bands (HF: 0.5–2 Hz), the LF/HF ratio and the triangular interpolation of the NN interval histogram (TINN). TINN is the width of the base of the triangle that best approximates the NN interval (normal R-peaks) distribution.

These features could be affected by inaccurate R-peak detection. Therefore, periods with noise and artefacts in the filtered ECG signal were identified, as these could lead to inaccurate heartbeat detection. For this purpose, a spectrogram was calculated using a 3-second time window with a 0.5-second overlap between consecutive windows. If the bandpower within the 1–30 Hz range exceeded a certain threshold (Equation (1), with $C = 15$), no features were calculated for those periods (Appendix Fig. A.2). During this process, 2.12 % of data were discarded. R-peak detection was performed in the remaining "clean" time segments. Based on the RR intervals, the following features were calculated: Moving average, moving median, the difference between mean and median, moving standard deviation, and coefficient of variation. These features were determined over two-time windows: a short window with 20 data points and a long window with 100 data points.

In addition, a time feature was included to indicate the temporal sequence of epochs, allowing neighbouring epochs to be classified based on their likelihood of having similar sleep stages. All extracted features are summarised in Table 3.

2.4.3. Sleep stage classification using support vector machine

A supervised learning approach using an SVM was employed to classify the sleep stages of preterm infants. The extracted standardised features and the manually annotated sleep stages served as inputs for training three different models: One utilising features from both the piezo mat and ECG, and two others using features from either the piezo mat or ECG exclusively. This approach would ensure reliable sleep stage detection even if one signal source is unavailable.

A multiclass classification model was implemented using the error-correcting output codes (*fitcecoc*, Matlab) approach, which combines the results of multiple binary classifiers to make multiclass decisions. To improve model stability and enhance overall performance, features were standardised. A fivefold cross-validation was conducted to evaluate model performance. The average performance results are shown in Table A.1 in the Appendix. To optimise model performance, various kernel functions—linear, Gaussian, radial basis function (RBF), and polynomial—were tested to determine the best fit. RBF performed best and was finally selected. During each cross-validation run, hyperparameters such as coding parameter, kernel scale, and box constraint were fine-tuned to achieve the best performance. Additionally, various features and feature combinations were iteratively added to the SVM. Features that did not contribute to improved model performance were removed to minimise overfitting and to reduce the number of parameters to be estimated, ultimately retaining only the most relevant features. Table 3 summarises all extracted features, highlighting the ones finally used for the models.

To evaluate the performance of the proposed method, we compared the predicted sleep stages to the PSG-derived and manually annotated sleep stages that served as groundtruth. As performance measures, we used accuracy, sensitivity, specificity, and Cohen's kappa, which are calculated from the confusion matrix.

2.4.4. Feature importance analysis

The impact of each of the used features on the SVM Model's classification output was assessed using the SHapley Additive exPlanations (SHAP) method [27]. SHAP assigns each feature an importance value for a particular prediction. The Shapley value of a feature for a certain epoch explains the deviation of the SVM model's prediction from the average prediction, due to this feature. The SHAP analysis was applied using 'shapley' Matlab function to the three final models, allowing for a comprehensive assessment of each feature importance across all epochs of dataset A and for each of the three sleep stages W, AS, and QS. We calculated the average absolute SHAP value for each feature across all 30-s epochs.

2.5. Statistics

The relative amounts of time spent in W, AS, and QS represent compositional data, as the total for each patient sums to 100 %. To account for the interdependencies of these proportions, we applied a centred log-ratio transformation, following Aitchison [28], prior to conducting a two-way ANOVA. The log-transformed data were tested for normality using the D'Agostino-Pearson test ($p > 0.17$) and for homogeneity of variances with Levene's test ($p = 0.19$). Statistical power was assessed assuming an alpha level of 0.1. As the differences in sleep bout lengths (comparing manual vs. automatic sleep stage classification) did not meet normality assumptions, we employed the Wilcoxon signed-rank test to assess these differences. All statistical analyses were performed using OriginPro 2019 (see Section 3.3).

3. Results

In this section, we present the results of the proposed methods for the

classification of sleep stages in preterm infants. We start by evaluating the performance of the movement detection procedure based on the piezo mat signal alone, as discussed in Section 2.4.2, and then we present the performance of our approach using both piezo mat and ECG signals.

3.1. Piezo mat for movement detection

To unobtrusively measure patient movement, we placed a piezo mat beneath the mattress (Fig. 2A). While this ballistographic approach has been suggested previously [29], we conducted our own performance evaluation of the proposed piezo mat. A synchronised video recording served as ground truth for visually detectable patient movements.

The automatic extraction of movement periods is detailed in Section 2.4.2. The best parameter configuration yielded an average correlation index of 0.78 ± 0.01 (n = 5) between automatically and visually detected movements (Fig. 2B).

We assume that the piezo mat has a higher sensitivity to brief twitches of skeletal muscles, which may have escaped the human eye. This is actually an important feature because the twitches are related to AS.

As expected, during resting periods, the piezo mat provides a clear signal corresponding to patient respiration (Fig. 2C). However, this signal was not utilised as a feature for sleep stage classification due to its intermittent availability.

3.2. Automatic sleep stage classification

We trained an SVM to automatically detect the sleep stages of preterm infants and compared the results to PSG-derived annotations. Following the clinical gold standard, three human experts manually scored the sleep stages of 28 patients. The statistical agreement among them, calculated using Fleiss' kappa, was 0.7 ± 0.014 (n = 28), indicating a substantial but imperfect interrater agreement. Most inconsistencies were observed between active sleep and wakefulness, which was the case in about 20 % of all 30-s epochs (Appendix Fig. A.1B). A total number of 25 features were extracted from ECG and piezo data sources, although not all of them were finally used (Table 3). Feature importance results are summarized in Appendix Fig. A.3. In the piezo model, short-term activity was identified as the most influential feature, highlighting its relevance for model predictions. For the ECG model, the median RR-interval, TINN, and start time emerged as the most important features. In the combined piezo + ECG model the piezo-derived activity features became the most relevant ones, followed by ECG-derived features such as median RR-intervals and TINN. Notably, the inclusion of piezo data in the combined model reduced the importance of the start time feature, which played a more prominent role in the standalone ECG model.

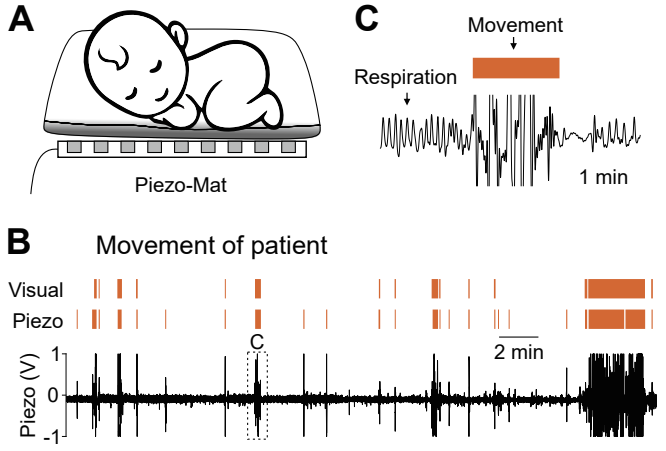


Fig. 2. Contactless measurement of movement in preterm infants. (A) A piezo-based sensor mat is placed under the mattress. (B) Raw signal from the piezo mat and the extracted movement compared to visually observed movements from the camera image. The average correlation is 0.78. Please note that brief movements cannot be seen in the video making the piezo mat generally more sensitive to movement. (C) Time-magnified segment (1 min) from B. During resting periods, there is a clear respiration signal. Movement is characterised by high-amplitude signals from the piezo mat.

Fig. 3A shows two example features over time and the corresponding hypnograms from human experts (Dataset B, see Section 2.3.2), the predicted sleep stages from our piezo + ECG model, and the exported sleep stages from Sleepware G3. This comparison illustrates a solid agreement between predicted and observed sleep stages, whereas the current Sleepware does not perform as well. Performance measures were calculated based on the confusion matrix, exemplified in Fig. 3B for the testing data (20 % of data) of one-fold. The best classification performance was achieved by our piezo + ECG model trained on dataset A with an accuracy of 92.2 ± 0.01 % (n = 5, five-fold cross-validation) and a Cohen's kappa of 0.83 ± 0.002 (Fig. 3C, Table 4). Sensitivity and specificity were 88 ± 0.01 % and 94 ± 0.01 %, respectively. We also trained models using features from either the piezo mat or the ECG alone. These models performed slightly worse than the combined model but still outperformed the automatic annotation from the Sleepware G3 which only had a kappa of 0.31 ± 0.21 (n = 28). The large SEM for all accuracy measures points to high interindividual variability of the Sleepware G3 prediction, whereas our trained models perform similarly well across all patients. We repeated this procedure on dataset B, which contains more continuous data and is therefore closer to reality. On the other hand, human experts did not fully agree in about 30 % of all epochs, introducing a level of uncertainty. As a result, the performance measures for all three models were slightly reduced compared to the models trained on Dataset A (Fig. 3D). All agreement metrics for all models, including those separated by individual sleep stages, can be found in Table A.1 in the Appendix. We would like to emphasise that predicting AS is the most challenging. The agreement for AS is consistently the poorest among all sleep stages, which is also reflected in the manual sleep scoring (Appendix Fig. A.1B). The majority of epochs discarded for dataset A were co-labeled with AS. We observed similar discrepancies between model predictions and human annotations that are evident from the confusion matrix (Fig. 3B) and the analysis of human disagreements (Appendix Fig. A.1B). AS can be confused with W and QS, but the probability of confusing QS with W is very low. This finding is also evident when comparing the accuracy ratio between Dataset A and B trained models (Appendix Table A.1). The greatest reduction in agreement for models trained on Dataset B was due to inaccuracies in the detection of W and AS (Accuracy ratio Dataset A/B for W and AS was 0.93, in comparison to QS with 0.98). This can be explained by the fact that most of the additional epochs in Dataset B contain inconsistencies between W and AS (Appendix Fig. A.1B).

3.3. Sleep characteristics

To evaluate the feasibility of transitioning sleep diagnostics to our proposed model, we assessed whether the sleep parameters derived from the model align with the current gold standard of PSG. We compared the percentage of recording time and sleep bout lengths for W, AS, and QS between observed and predicted sleep stage classifications with our piezo + ECG model. Only minor differences were observed in the average percentage of recording time for wake (observed: 37.9 ± 2.6 %, predicted: 36.3 ± 2.1 %), QS (observed: 26.5 ± 1.5 %, predicted: 24.2 ± 1.5 %) and AS (observed: 35.6 ± 1.9 %, predicted: 39.5 ± 1.5 %) (Fig. 4A). A two-way ANOVA indicated no significant interaction between classification method and sleep stage distribution ($p = 0.17$) with a reasonably high power of 0.77 ($\alpha = 0.1$) (see also Section 2.5). Notably, we observed considerable variability in sleep stage distributions across patients, with around 25 % showing a deviation of 20 %

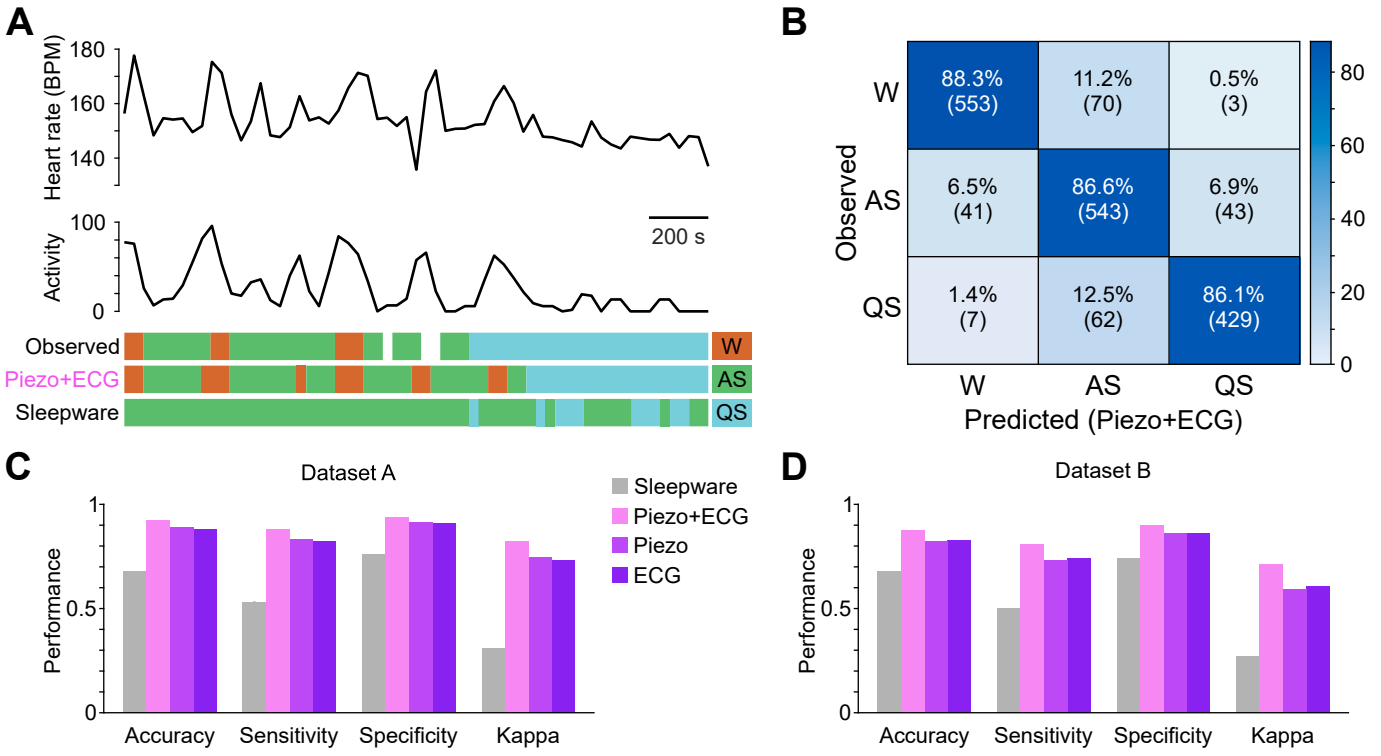


Fig. 3. Performance evaluation of sleep stage classifiers. (A) The heart rate (BPM: beats per minute) and activity (short, moving average over one minute) from a 30-minute recording are displayed as examples. Below, the manually annotated sleep stages (observed Dataset B), the predictions from the piezo + ECG model, and the classifications from the Sleepware G3 (Philips) sleep lab software are shown for comparison. (B) Confusion matrix of the piezo + ECG model trained on PSG-derived sleep data (Dataset A). The matrix shows percent agreement, with the number of testing epochs (20 % of the data) indicated in brackets. (C) All machine learning models trained on observed Dataset A sleep data outperform the current software across various agreement measures. The model combining piezo and ECG features shows a strong agreement with PSG-derived sleep stages ($\kappa > 0.8$). (D) Classifiers trained on PSG-derived sleep data (Dataset B) perform slightly worse.

Table 4
Performance measures of dataset A trained models reported as mean ± SD over five-fold cross-validation. Sleepware-generated sleep stages were compared to Dataset A of all patients (n = 28) reported as mean ± SD.

	Piezo + ECG	Piezo	ECG	Sleepware
Accuracy	0.92 (±0.01)	0.89 (±0.004)	0.88 (±0.002)	0.68 (±0.16)
Kappa	0.83 (±0.002)	0.75 (±0.01)	0.73 (±0.01)	0.31(±0.21)
Sensitivity	0.88 (±0.01)	0.83 (±0.01)	0.82 (±0.004)	0.53 (±0.38)
Specificity	0.94 (±0.01)	0.92 (±0.003)	0.91 (±0.002)	0.76 (±0.4)
Precision	0.89 (±0.01)	0.8 (±0.01)	0.82 (±0.003)	0.73 (±0.28)

between observed and predicted sleep stages. These deviations were most pronounced in patients with steeper slopes in Fig. 4A. The median sleep bout length per patient was not found to be different between observed and predicted sleep stages for W and AS (Fig. 4B, W: $p = 0.09$, AS: $p = 0.14$, $n = 28$), even though the cumulative probability distribution showed a slight left shift (Fig. 4C). For QS, however, we observed a significant reduction in median sleep bout lengths (QS: $p < 0.001$), which was also evident from the leftward shift in the cumulative probability distribution (Fig. 4C). This discrepancy may be explained by instances where long QS phases, manually classified by clinicians, were interrupted by brief W or AS bouts in the automatic classification, thereby reducing the overall median sleep bout length.

4. Discussion

The goal of the present study was to test the feasibility of an automatic sleep stage classification in preterm infants that uses as few biosignals as possible and achieves clinically acceptable accuracy. We used a combination of movement and ECG features to train three different classifiers using an SVM: One utilising features from both the piezo mat and ECG, and two others using features from either the piezo mat or ECG exclusively. The proposed approach ensures reliable sleep stage detection, even when one signal source is unavailable. Our combined piezo + ECG model achieved a kappa of 0.83, which is considered, according to Cohen's kappa measure [30,31], a strong agreement with the manually annotated sleep stages. Models based solely on piezo or ECG data also performed well, achieving kappa values of 0.75 and 0.73, respectively. These kappa values are comparable to or higher than the interrater reliability that can be achieved by manual sleep scoring according to AASM standards [16,32]. These results suggest that our models can potentially be considered clinically acceptable [33]. Surprisingly, the current software Sleepware G3 used in our paediatric sleep lab frequently produces inaccurate sleep stage outputs, often requiring manual corrections. Implementing our automatic sleep stage classification would significantly enhance clinicians' time management, allowing them to dedicate more attention to other clinical responsibilities.

Compared to previous studies, we observed notable differences in the distribution of sleep stages, particularly regarding wake times. While the literature typically reports ~ 10 % wake periods [12,13,15], our data showed a much higher percentage (38 %). Several factors could explain this discrepancy, including the infants' transport to the sleep lab, adaptation to a new environment, handling during sensor placement, and the connection of electrodes and sensors. Almost all hypnograms in our study began with extended wake periods before sleep onset (see Appendix), further supporting this observation. In contrast, the fractional amounts of AS and QS were in line with the literature, reporting ~ 60 % AS and ~ 40 % QS [8]. The analysis of sleep bout lengths provides insights into the frequency of sleep stage transitions and our models' ability to detect them. We confirmed short median wake periods of approximately 2 min [15]. However, our results differed for other

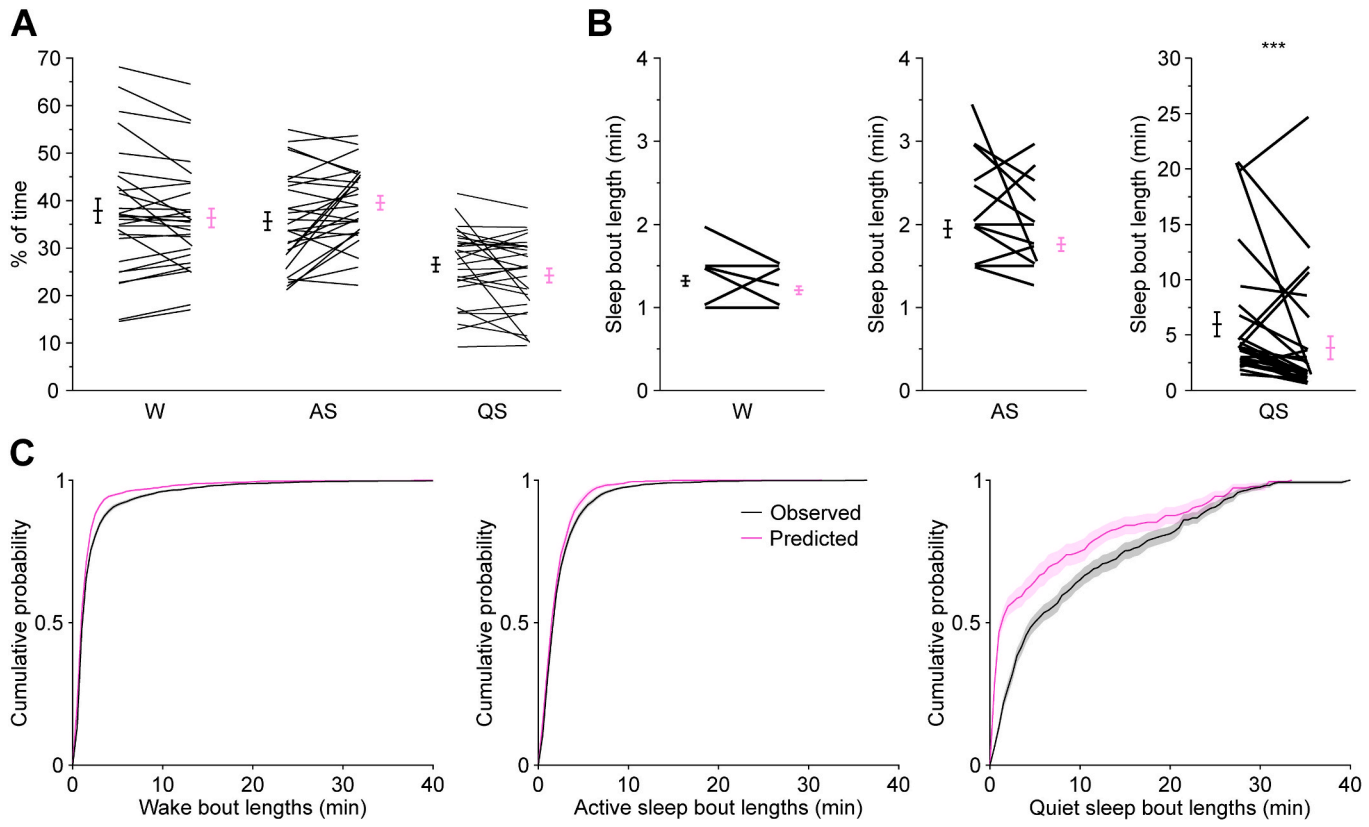


Fig. 4. Comparison of sleep characteristics between visual annotation (black) and automatic sleep stage classification (magenta, piezo + ECG model). (A) Percent of time spent in W, AS, and QS. (B) Quantification of median sleep bout length per patient for W, AS, and QS. Note the different y-scale for QS, as quiet sleep bouts are much longer. A darker color indicates that more lines are stacked on top of each other. (C) Cumulative probability distribution of epoch lengths. Distributions were averaged across patients. Data are presented as mean ± SEM. *** $p < 0.001$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

sleep stages: we observed a shorter median AS duration (2 min vs. 4 min in Wang et al.) and a longer median QS duration (6 min vs. 4 min). Notably, they also reported a tendency toward shorter sleep stage durations using their automatic LightGBM classifier [15] as we have observed for QS. In our study, long periods of QS were occasionally interrupted by brief AS or wake epochs. This raises important questions: do human experts subconsciously apply a "temporal filter" when scoring sleep stages, or could our algorithm be more precise and sensitive to detecting short stage transitions? This finding, together with the inter-rater reliability of 0.7 (Fleiss' kappa), suggests that the gold standard of manual sleep scoring may have limitations, potentially impeding the development of more accurate automatic systems by introducing inconsistencies in sleep stage classification, which may ultimately be adopted by the models themselves. It may even call for a reconsideration of existing rules for sleep stage classification.

To the best of our knowledge, our approach achieved the highest accuracy for a three-stage sleep classification (W, AS, QS) based on vital signs when compared to previous studies reporting kappa values ranging from 0.38 to 0.52 [13–15]. Zhang et al. extracted movement data as motion artefacts from the ECG [14], which may not capture brief movements like twitches. Our ECG-based model ($\kappa = 0.73$) outperformed earlier ECG algorithms ($\kappa = 0.33$) [12].

The achieved high accuracy could be attributed to different factors: (1) the integration of novel motion features of the piezo mat signal, (2) the inclusion of the temporal feature of the epochs' sequence that facilitates classifying any epoch based on its features as well as those of neighbouring epochs, (3) the careful preprocessing to reduce noise and outliers in the signals, (4) the careful selection and the reduced dimensionality of the feature space to avoid overfitting during machine learning. For comparison, Werth et al. used 47 features, which may have introduced overfitting [12].

Our detailed feature importance analysis revealed that the ranking of features in the standalone models was largely preserved in the combined piezo + ECG model (Appendix Fig. A.3). Notably, short-term activity emerged as a key predictor for AS and W which aligns with their definitions as movement-dominated states (Table 2). In contrast, long-term activity, median heart rate, and TINN (a measure HRV) were the most important features for predicting QS, consistent with its characterization by behavioural silence, low heart rate, and reduced HRV (Table 2).

Astonishingly, our models using only vital signs achieved an accuracy comparable to EEG-based algorithms [9,34–36], despite the expectation that EEG provides the most relevant features for sleep stage classification [8,10]. The limitations of EEG recordings in preterm infants are manifold. First, EEG electrodes can damage the vulnerable skin of preterm infants or cause inflammation, making them unsuitable for long-term recording [8]. Second, EEG signatures become increasingly unreliable the younger the preterm infants [37,38]. The highest performance of an EEG-based algorithm for three-stage sleep classification was reported by Fraiwan and Alkhodari, achieving an accuracy of 0.96 and a kappa of 0.91 [34]. However, their study focused on term and preterm infants at 40 weeks postconceptional age, a developmental stage where adult-like EEG patterns emerge, making their approach less applicable to younger infants. More recent studies have made advancements in minimising the number of EEG electrodes or developing algorithms tailored to younger patients [35,36]. Both studies achieved kappa values around 0.7, which is lower than the performance of our models that rely on vital signs rather than EEG.

The combination of movement and ECG-derived features for automatic sleep stage classification has previously been applied to specifically detect deep sleep (N3 stage) in infants aged 1 to 18 weeks,

achieving an accuracy of 0.97 [20]. Our combined piezo + ECG model similarly detects QS with an accuracy of 0.95 (Appendix Table A.1, which is comparable to the detection of N3 during the early postnatal weeks in term infants. However, unlike Ranta and colleagues, we performed a three-stage classification. Additionally, we evaluated our movement detection algorithm using visually annotated movements from video recordings, as we assume that the activity profile of preterm infants provides valuable information for sleep stage classification. Our algorithm demonstrates a strong correlation of 0.78 with the visual annotations. Notably, our piezo mat exhibited higher sensitivity to small, brief movements – likely related to twitches – associated with AS [39]. In contrast, prior studies using 15-second windows for movement quantification may have missed these subtle movements [29].

The unobtrusive measurement and automatic sleep stage classification enable continuous monitoring of sleep in preterm infants. On one hand, this allows for the clinical monitoring of sleep cycle development as an indicator of normal brain development [2]. On the other hand, it paves the way for individualised medical care tailored to the sleep cycles of preterm infants. The ultimate goal is to protect their precious sleep that is frequently interrupted in the NICU environment to promote their health and early brain development [8,15].

5. Conclusion

We have demonstrated that reliable sleep stage detection is feasible in preterm infants using unobtrusive vital sign measurements. Our trained SVM models, based on either or both movement and heart rate information, can compete with EEG-based algorithms. Our results suggest that algorithms based on vital signs are more suitable for preterm infants, which may be due to the fact that EEG patterns only start to emerge after 32 weeks of gestation [10]. Although the patient group used is unbiased with respect to gender (GDR ~ 1), the generalisability of our findings remains limited by the current relatively small sample size. Future work is essential to validate the applicability of our models in extremely preterm as well as term infants. Future research should also focus on exploring deep learning-based feature extraction approaches to use only previous data, thus enabling real-time applicability of the piezo + ECG model [13,15]. Beyond real-time classification, the prediction of future sleep-wake transitions is essential to facilitate better planning for caregivers and clinicians. Clustered care, tailored to the individual sleep stages of preterm infants, may offer the greatest benefit to the long-term outcomes of these vulnerable patients. At the same time, continuous sleep monitoring may improve our understanding of the relationship between sleep disturbance and brain development in preterm infants.

CRediT authorship contribution statement

Nathalie Demme: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis. **Maha Shadaydeh:** Writing – review & editing, Writing – original draft, Supervision, Software, Methodology. **Laura Schieder:** Software, Investigation, Formal analysis, Data curation. **Claus Doerfel:** Writing – review & editing, Data curation. **Stella Jähkel:** Data curation. **Knut Holthoff:** Writing – review & editing, Supervision, Resources, Funding acquisition. **Hans Proquitté:** Writing – review & editing, Supervision, Resources, Conceptualization. **Joachim Denzler:** Writing – review & editing, Supervision, Resources. **Jürgen Graf:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author used chatGPT in order to improve grammar and language style. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Funding

This project was supported by the Interdisciplinary Centre for Clinical Research (IZKF; medical scientist program MSP-18 to J.G.). K.H. received funding from the German Research Foundation (HO 2156/5–1, HO 2156/6–1). H.P. received funding from the Federal Ministry of Education & Research (BMBF FKZ: 03ZZ0482A).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bspc.2025.107904.

Data availability

The authors do not have permission to share patient data.

References

[1] E.O. Ohuma, A.-B. Moller, E. Bradley, S. Chakwera, L. Hussain-Alkhateeb, A. Lewin, Y.B. Okwaraji, W.R. Mahanani, E.W. Johansson, T. Lavin, D. E. Fernandez, G.G. Domínguez, A. de Costa, J.A. Cresswell, J. Krasevec, J.E. Lawn, H. Blencowe, J. Requejo, A.C. Moran, National, regional, and global estimates of preterm birth in 2020, with trends from 2010: a systematic analysis, Lancet 402 (2023) 1261–1271, https://doi.org/10.1016/S0140-6736(23)00878-4.

[2] L. Bennet, D.W. Walker, R.S.C. Horne, Waking up too early - the consequences of preterm birth on sleep development, J. Physiol. 596 (2018) 5687–5708, https://doi.org/10.1113/JP274950.

[3] J.P. Vogel, S. Chawanpaiboon, A.-B. Moller, K. Watananirun, M. Bonet, P. Lumbiganon, The global epidemiology of preterm birth, Best Pract. Res. Clin. Obstet. Gynaecol. 52 (2018) 3–12, https://doi.org/10.1016/j.bpobgyn.2018.04.003.

[4] M.S. Blumberg, J.C. Dooley, G. Sokoloff, The developing brain revealed during sleep, Curr. Opin. Physiol. 15 (2020) 14–22, https://doi.org/10.1016/j.cophys.2019.11.002.

[5] O. Weisman, R. Magori-Cohen, Y. Louzoun, A.I. Eidelman, R. Feldman, Sleep-wake transitions in premature neonates predict early development, Pediatrics 128 (2011) 706–714, https://doi.org/10.1542/peds.2011-0047.

[6] A.M. Morse, S.V. Kothare, Does sleep correlate with neurodevelopmental outcomes in preterm and term infants in early-preschool children? J. Clin. Sleep Med. 19 (2023) 639–640, https://doi.org/10.5664/jcsm.10522.

[7] J. Levy, F. Hassan, M.A. Plegue, M.D. Sokoloff, J.S. Kushwaha, R.D. Chervin, J.D. E. Barks, R.A. Shellhaas, Impact of hands-on care on infant sleep in the neonatal intensive care unit, Pediatr. Pulmonol. 52 (2017) 84–90, https://doi.org/10.1002/ppul.23513.

[8] J. Werth, L. Atallah, P. Andriessen, X. Long, E. Zwartkruis-Pelgrim, R.M. Aarts, Unobtrusive sleep state measurements in preterm infants - A review, Sleep Med. Rev. 32 (2017) 109–122, https://doi.org/10.1016/j.smrv.2016.03.005.

[9] S.F. Abbasi, A. Abbas, I. Ahmad, M.S. Alshehri, S. Almakdi, Y.Y. Ghadi, J. Ahmad, Automatic neonatal sleep stage classification: A comparative study, Heliyon 9 (2023) e22195, https://doi.org/10.1016/j.heliyon.2023.e22195.

[10] M.M. Grigg-Damberger, The visual scoring of sleep in infants 0 to 2 months of age, J. Clin. Sleep Med. 12 (2016) 429–445, https://doi.org/10.5664/jcsm.5600.

[11] J.R. Isler, T. Thai, M.M. Myers, W.P. Fifer, An automated method for coding sleep states in human infants based on respiratory rate variability, Dev. Psychobiol. 58 (2016) 1108–1115, https://doi.org/10.1002/dev.21482.

[12] J. Werth, M. Radha, P. Andriessen, R.M. Aarts, X. Long, Deep learning approach for ECG-based automatic sleep state classification in preterm infants, Biomed. Signal Process. Control 56 (2020) 101663, https://doi.org/10.1016/j.bspc.2019.101663.

[13] T. Sentner, X. Wang, E.R. de Groot, L. van Schaijk, M.L. Tataranno, D.C. Vijlbrief, M.J.N.L. Benders, R. Bartels, J. Dudink, The Sleep Well Baby project: an automated real-time sleep-wake state prediction algorithm in preterm infants, Sleep 45 (2022) 1–11, https://doi.org/10.1093/sleep/zsac143.

[14] D. Zhang, Z. Peng, S. Sun, C. van Pul, C. Shan, J. Dudink, P. Andriessen, R.M. Aarts, X. Long, Characterising the motion and cardiorespiratory interaction of preterm infants can improve the classification of their sleep state, Acta Paediatr. 113 (2024) 1236–1245, https://doi.org/10.1111/apa.17211.

[15] X. Wang, E.R. de Groot, M.L. Tataranno, A. van Baar, F. Lammertink, T. Alderliesten, X. Long, M.J.N.L. Benders, J. Dudink, Machine learning-derived active sleep as an early predictor of white matter development in preterm infants, J. Neurosci. 44 (2024) 1–7, https://doi.org/10.1523/JNEUROSCI.1024-23.2023.
[16] H. Danker-Hopfe, P. Anderer, J. Zeitlhofer, M. Boeck, H. Dorn, G. Gruber, E. Heller, E. Loretz, D. Moser, S. Parapatics, B. Saletu, A. Schmidt, G. Dorffner, Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard, J. Sleep Res. 18 (2009) 74–84, https://doi.org/10.1111/j.1365-2869.2008.00700.x.
[17] X. Long, J. Espina, R.A. Otte, W. Wang, R.M. Aarts, P. Andriessen, Video-based actigraphy is an effective contact-free method of assessing sleep in preterm infants, Acta Paediatr. 110 (2021) 1815–1816, https://doi.org/10.1111/apa.15740.
[18] M. Awais, X. Long, B. Yin, S. Farooq Abbasi, S. Akbarzadeh, C. Lu, X. Wang, L. Wang, J. Zhang, J. Dudink, W. Chen, A hybrid DCNN-SVM model for classifying neonatal sleep and wake states based on facial expressions in video, IEEE J. Biomed. Health Inform 25 (2021) 1441–1449, https://doi.org/10.1109/JBHI.2021.3073632.
[19] D. Huang, D. Yu, Y. Zeng, X. Song, L. Pan, J. He, L. Ren, J. Yang, H. Lu, W. Wang, Generalized camera-based infant sleep-wake monitoring in NICUs: a multi-center clinical trial, IEEE J. Biomed. Health Inform. 28 (2024) 3015–3028, https://doi.org/10.1109/JBHI.2024.3371687.
[20] J. Ranta, M. Airaksinen, T. Kirjavainen, S. Vanhatalo, N.J. Stevenson, An open source classifier for bed mattress signal in infant sleep monitoring, Front. Neurosci. 14 (2020) 602852, https://doi.org/10.3389/fnins.2020.602852.
[21] C. Iber, S. Ancoli-Israel, A.L. Chesson, Quan Stuart F. (Eds.), The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, American Academy of Sleep Medicine, Westchester, 2007.
[22] R.B. Berry, C.E. Gamaldo, S.M. Harding, R. Brooks, R.M. Lloyd, B.V. Vaughn, C. L. Marcus, AASM scoring manual version 2.2 updates: new chapters for scoring infant sleep staging and home sleep apnea testing, J. Clin. Sleep Med. 11 (2015) 1253–1254, https://doi.org/10.5664/jcsm.5176.
[23] S.A. Romano, T. Pietri, V. Pérez-Schuster, A. Jouary, M. Haudrechy, G. Sumbre, Spontaneous neuronal network dynamics reveal circuit's functional adaptations for behavior, Neuron 85 (2015) 1070–1085, https://doi.org/10.1016/j.neuron.2015.01.027.
[24] C.S. Cutts, S.J. Eglen, Detecting pairwise correlations in spike trains: an objective comparison of methods and application to the study of retinal waves, J. Neurosci. 34 (2014) 14288–14303, https://doi.org/10.1523/JNEUROSCI.2767-14.2014.
[25] A.A. Garvey, A.M. Pavel, J.M. O'Toole, B.H. Walsh, I. Korotchikova, V. Livingstone, E.M. Dempsey, D.M. Murray, G.B. Boylan, Multichannel EEG abnormalities during the first 6 hours in infants with mild hypoxic-ischaemic encephalopathy, Pediatr. Res. 90 (2021) 117–124, https://doi.org/10.1038/s41390-021-01412-x.
[26] J.M. O'Toole, Features of Heart Rate Variability for Neonates: (Matlab code), 2022. https://github.com/otoolej/hrv_features_neonates?tab=readme-ov-file#contact (accessed 22 October 2024).
[27] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Proces. Syst. (2017) 4768–4777.
[28] J. Aitchison, The statistical analysis of compositional data, J. R. Stat. Soc. 44 (1982) 139–177, https://doi.org/10.1111/j.2517-6161.1982.tb01195.x.
[29] R. Joshi, B.L. Bierling, X. Long, J. Weijers, L. Feijs, C. van Pul, P. Andriessen, A ballistographic approach for continuous and non-obtrusive monitoring of movement in neonates, IEEE J. Transl. Eng. Health Med. 6 (2018) 2700809, https://doi.org/10.1109/JTEHM.2018.2875703.
[30] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics 33 (1977) 159, https://doi.org/10.2307/2529310.
[31] M.L. McHugh, Interrater reliability: the kappa statistic, Biochem Med (2012) 276–282, https://doi.org/10.11613/BM.2012.031.
[32] Y.J. Lee, J.Y. Lee, J.H. Cho, J.H. Choi, Interrater reliability of sleep stage scoring: a meta-analysis, J. Clin. Sleep Med. 18 (2022) 193–202, https://doi.org/10.5664/jcsm.9538.
[33] H. Phan, A. Mertins, M. Baumert, Pediatric automatic sleep staging: a comparative study of state-of-the-art deep learning methods, IEEE Trans. Biomed. Eng. 69 (2022) 3612–3622, https://doi.org/10.1109/TBME.2022.3174680.
[34] L. Fraiwan, M. Alkhodari, Neonatal sleep stage identification using long short-term memory learning system, Med. Biol. Eng. Comput. 58 (2020) 1383–1391, https://doi.org/10.1007/s11517-020-02169-x.
[35] H. Zhu, L. Wang, N. Shen, Y. Wu, S. Feng, Y. Xu, C. Chen, W. Chen, MS-HNN: multi-scale hierarchical neural network with squeeze and excitation block for neonatal sleep staging using a single-channel EEG, IEEE Trans. Neural Syst. Rehabil. Eng. 31 (2023) 2195–2204, https://doi.org/10.1109/TNSRE.2023.3266876.
[36] H.A. Siddiqa, M. Irfan, S.F. Abbasi, W. Chen, Electroencephalography (EEG) based neonatal sleep staging and detection using various classification algorithms, CMC 77 (2023) 1759–1778, https://doi.org/10.32604/cmc.2023.041970.
[37] C. Dreyfus-Brisac, J.C. Larroche, Discontinuous electroencephalograms in the premature newborn and at term. Electro-anatomo-clinical correlations, Rev. Electroencephalogr. Neurophysiol. Clin. 1 (1971) 95–99, https://doi.org/10.1016/s0370-4475(71)80022-9.
[38] K. Palmu, T. Kirjavainen, S. Stjerna, T. Salokivi, S. Vanhatalo, Sleep wake cycling in early preterm infants: comparison of polysomnographic recordings with a novel EEG-based index, Clin. Neurophysiol. 124 (2013) 1807–1814, https://doi.org/10.1016/j.clinph.2013.03.010.
[39] G. Sokoloff, G. Sokoloff, R.Y. Wen, M.E. Tobias, B. McMurray, M.S. Blumberg, Spatiotemporal organization of myoclonic twitching in sleeping human infants, Dev. Psychobiol. 62 (2020) 697–710, https://doi.org/10.1002/dev.21954.