Estimation of Quiet Sleep in Preterm and Full-Term Newborns Using Machine Learning Algorithms Based on Cardio-Respiratory and Motion Signals

Houda Jebbari, Sandie Cabon, Patrick Pladys, Guy Carrault, Fabienne Porée

Univ Rennes, INSERM, LTSI - UMR 1099, F-35000, Rennes, France

Abstract

Monitoring sleep states of newborns, especially those born prematurely, before 37 weeks of gestation, is essential for tracking their development. This study presents an automated method for estimating the state of Quiet Sleep (QS). Given that QS is characterized by regular cardiorespiratory rhythm and non-movement, this approach combines machine learning algorithms trained on cardiorespiratory features with body motion segmentation. It was evaluated on manually annotated recordings from 10 preterm and 5 full-term newborns. Each newborn was recorded for eight hours during their first week of life, and preterm newborns were recorded again at 37 weeks Post-Menstrual Age (PMA). The results achieved an average balanced accuracy of 78% and a Cohen's kappa of 0.51 across all recordings. For neonates with a PMA greater than 33 weeks, these values increased to 82% and 0.58, respectively. This approach proves effective and holds promise for continuously monitoring QS in newborns with a PMA greater than 33 weeks using non-invasive signals.

1. Introduction

Babies born before 37 weeks of Gestational Age (GA) are called premature. Due to their immature functions resulting from early birth during a critical phase of development, they necessitate admission to the Neonatal Intensive Care Unit (NICU) [1]. In the NICU, the baby's physiological signals, such as ECG, respiration, and oxygen saturation, are continuously monitored to promptly address any issues that may arise. Moreover, NICU staff engages in behavioral observations, closely analyzing facial expressions, vocalizations, and movements. This process aids in monitoring newborn development and sleep patterns. The latter aspect provides valuable insights into brain maturation and enhancing care practices [1]. However, extracting Sleep States (SS) from these observations is challenging. An alternate approach is polysomnography, offering an in-depth assessment of newborn SS. Nevertheless, this requires additional electrodes on the baby's scalp and body, causing discomfort and intrusion.

State-of-the-art techniques for extracting premature SS from physiological signals continuously monitored in the NICU include deep learning on ECG parameters [2], thresholding on respiratory signals [3], and classical automatic learning using heart rate, respiratory rate, and oxygen saturation [4]. These methods have performance limitations, especially in distinguishing certain stages.

Video processing techniques, that account for the behavioral aspects for newborn SS, have also been explored. They include facial expression recognition [5], as well as data extraction related to eye states and body movements from video, and cry sounds from audio recordings, and subsequently integrating this information [6]. However, these approaches overlook the genuine NICU conditions where the newborn's face is often not visible.

There are five sleep and wake states in newborn: Quiet Sleep (QS), active sleep, drowsiness, quiet wakefulness, and active wakefulness. Among these, QS is essential for brain development [1], and it is defined by regular cardiorespiratory rhythms and absence of motion. This study presents a method for estimating QS by combining Machine Learning (ML) algorithms trained on cardiorespiratory features with body motion segmentation obtained from video data.

2. Method

In this section, acquisition protocol and database are first described. Then, methodology is divided into three parts (see Figure 1): (i) ML models based on cardiorespiratory features for QS first estimation (QS-1), (ii) newborn motion segmentation, (iii) fusion of ML model estimation and motion segmentation for the final QS estimation (QS-2).

2.1. Clinical data

The database used in this study is part of the European Digi-NewB project database [7]. It includes infrared video recordings and physiological signals. Videos were captured at a rate of 25 frames per second using MPEG-4 encoding. Electrodes were employed to record respiration and ECG at 62.5 Hz and 500 Hz sampling frequencies, re-

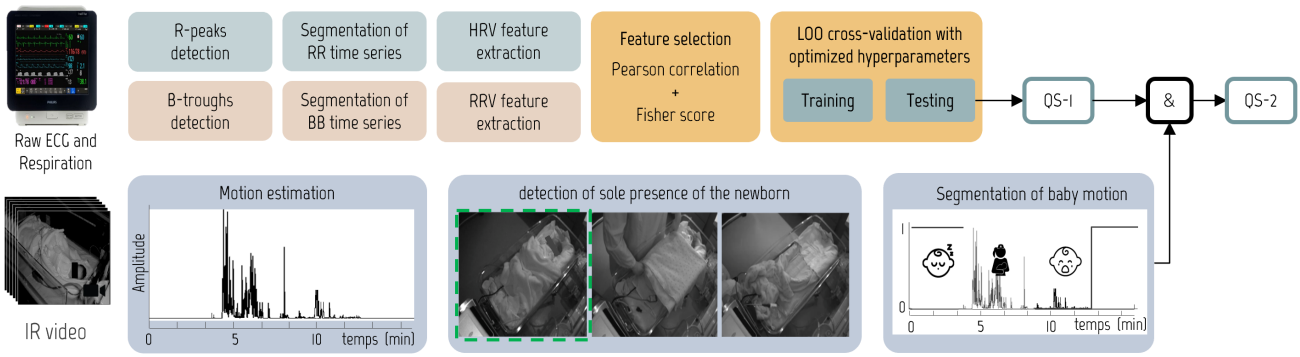

Figure 1. Overview of QS estimation method.

spectively. The study involved data from 15 newborns, all of whom underwent a medical examination to ensure the absence of any pathology. Newborns were categorized into three groups based on their GA in weeks:

- Group 1 : 5 very premature babies ($27 \leq GA < 29$) ;
- Group 2 : 5 moderately preterm ($33 \leq GA < 36$) ;
- Group 3 : 5 full-term ($GA \geq 39$).

Recordings were conducted within the first week of life for all newborns (day 1). Additionally, groups 1 and 2 were recorded again as they approached discharge from the hospital (day 2), at around 37 weeks Post-Menstrual Age (PMA), that corresponds to GA plus chronological age. Recording sessions took place at night, from 10 p.m. to 6 a.m., resulting in a total of 25 recordings and around 200 hours of data.

An expert manually annotated QS phases in each recording [1]. Periods when analysis was impossible due to the absence of the baby, the presence of an adult in the camera's field, or the absence of physiological signals, were excluded. The rest of the intervals were automatically considered as Non-QS state, encompassing the other four sleep and wake states.

2.2. ECG and respiration signal processing

2.2.1. Feature extraction

R-peaks in the ECG signal and Breathing troughs (B-troughs) in the respiration signal were detected using a modified Pan and Tompkins method with neonate-specific filter coefficients [8] and an algorithm described in [9]. A 40-second sliding window with 50% overlap was applied to segment these time series. Subsequently, Neurokit Python library was used to extract Heart Rate Variability (HRV) and Respiratory Rate Variability (RRV) features for each segment [10]. In total, 60 HRV features and 17 RRV features were extracted in both temporal and non-linear domains. Due to short duration of segments, i.e., less than one minute, frequency domain features were excluded.

2.2.2. Feature selection

Standardization: Feature values were standardized within each recording (z-scoring) to create a uniform scale and minimize individual variability.

Feature ranking: Fisher's score was calculated for every cardiorespiratory feature, by comparing the variance between classes, i.e., QS and Non-QS, to the variance within classes, providing valuable insights into how effectively each feature distinguishes between the two classes. A higher score indicates more significant discriminatory power for the feature. To eliminate non-significant features from the dataset, only those with scores higher than 0.1 were kept.

Correlation: Numerous dataset features showed significant correlations, indicating potential redundant information that could lead to overfitting. To address this, Pearson's correlation coefficient is used. When a strong correlation (greater than 0.7) was found between two features, the one with the higher Fisher score was kept.

2.2.3. Comparison of ML models

To estimate QS-1, ML models were explored using selected cardiorespiratory features and PMA. Three approaches were compared: Decision Tree (DT), Random Forest (RF), and Support Vector Machine (SVM).

Suitable hyperparameters for each model were determined through nested cross-validation (nCV), as shown in Figure 2. A Leave-One-Out Cross-Validation (LOOCV) strategy was used in the outer loop. Within the inner loop, hyperparameters for the three classifiers were fine-tuned using the grid search method. The training data was divided into four folds, grouped according to patient-level considerations and stratified to preserve the percentage of samples for each class. The inner loop returns the hyperparameters resulting in the highest averaged Balanced Accuracy (BA) on 4-folds validation set. The optimal hyperparameters were then determined by choosing the most selected in outer loop.

In the end, a regular LOOCV was executed to compare

the models' performance. In each iteration, the recording(s) of one newborn was reserved for testing, while the recordings from the remaining newborns were utilized for training. For newborns with two recordings (groups 1 and 2), each recording's data was tested separately to ensure a comprehensive evaluation.

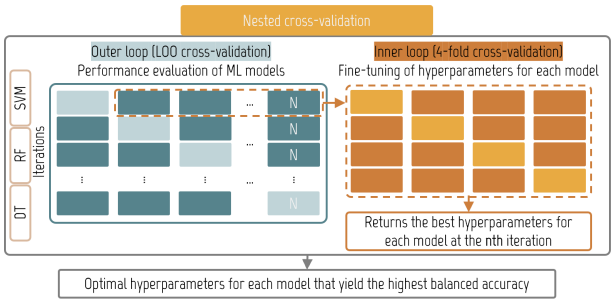

Figure 2. Diagram of the nested cross-validation, where N = 15 is the number of newborns in the dataset.

2.3. Motion processing

Extraction of newborn motion was based on a method, developed by our research team [11], involving three main steps. First, motion is estimated by using inter-image differencing. Then, periods when the baby was not in bed or when an adult was present in the field of the camera were filtered out using deep learning. Finally, a RF was trained to classify motion and non-motion intervals. This resulted in a binary signal: 0 for motion, and 1 for non-motion.

2.4. Fusion of ML model estimation and motion segmentation

The QS-1 estimation, given by ML models, was combined with the outcomes of motion segmentation. First, QS-1 vector was upsampled to 25 Hz (motion frequency). Then, a logical "AND" operation was performed between the two vectors. This fusion served to correct false positive QS-1 estimations and resulted in QS-2. Short motion periods (< 5 seconds) within estimated QS-1 were still considered part of the QS state, as they may represent startles or sighs movements [1].

3. Results

This section includes four subsections. Firstly, it explores the interpretability of the chosen compacted feature set for classification. Next, it presents the results of nCV, which involves identifying optimal hyperparameters for three ML models and selecting the best-performing one. Then, it discusses the fusion of QS-1 with motion segmentation for QS-2 estimation. Lastly, it analyzes the results of recordings grouped by their age.

3.1. Selected features and their interpretability

Feature selection identified 5 significant cardiorespiratory features for QS estimation, ranked in descending order of importance:

- *Fuzzy entropy of HRV*: Measures the complexity and unpredictability of R-R intervals. It increases during QS, indicating higher complexity and unpredictability.
- *Coefficient of variation of RRV*: Assesses Breath-to-Breath (B-B) interval variability. It decreases in the QS, indicating greater regularity and consistency in successive breaths.
- *20th percentile of HRV*: Represents a specific point within R-R intervals, below which 20% of the intervals are shorter. It is higher in QS, indicating lower heart rates during this state.
- *Median coefficient of variation of RRV*: Provides insight into relative variability of B-B intervals. It shows lower values in QS, indicating more consistent and less variable distribution centered around the median.
- *Average inspiration duration*: Reflects inspiratory phase duration. In QS, the duration is notably longer, reflecting a slower and more relaxed respiratory pattern.

3.2. Best-performing ML model

Optimal hyperparameters for the three classifiers determined through nCV are presented and highlighted in Table 1. Performances were assessed by comparing the BA and Cohen's Kappa of the average performance of regular LOOCV (see Table 2). The RF classifier outperformed the others, showing the highest average BA and Cohen's Kappa scores. Therefore, for the subsequent analysis, we will utilize the RF model to generate the results.

3.3. Quiet Sleep Estimation

The final QS estimate (QS-2) was obtained by combining QS-1 with motion segmentation. This led to a slight improvement in performance, with an average BA of 78.16±10.06% and a Cohen's kappa of 0.52±0.19. Despite the modest nature of this improvement, the integration of motion segmentation prevented the misclassification of motion intervals as QS periods.

3.4. Evaluation of results according to age

Performances were examined in terms of age of the babies (see Figure 3). For newborns in group 1 day 1, performances were relatively lower. For other groups, i.e., with a PMA greater than 33 weeks, the method yielded good results with an averaged BA of 82±6.79% and an average Cohen's kappa of 0.58±0.14. Lower performance of very

Table 1. Hyperparameters designated for grid search.

Hyperparameters	**DT**	**RF**	**SVM**
Class weight	balanced	balanced, balanced_subsample	balanced
Number of estimators	-	200; 250; 300	-
Max depth	10; 20; 30; 50	10; 20; 30; 50	-
Min samples split	2; 20; 50; 200	2; 20; 50; 200	-
Min samples leaf	0,004; 20; 50; 100	0.004; 20; 50; 100	-
Max features	'sqrt'; 0.4	'sqrt'; 0.4	-
C	-	-	0.1; 1; 10; 100; 1000
Kernel	-	-	'linear'; 'rbf'; 'poly'; 'sigmoid'

Table 2. Performances of ML models.

	DT	RF	SVM
BA	75.3±11.0	77.8±10.6	71.1±14.7
Cohen's Kappa	0.45±0.20	0.50±0.21	0.45±0.30

premature newborns in early days is likely due to the fact that, in these newborns, periods of QS are rare and shorter [1], making them more challenging to detect. Additionally, in these babies, the information present in the ECG and respiration is probably less comprehensive due to their significant immaturity.

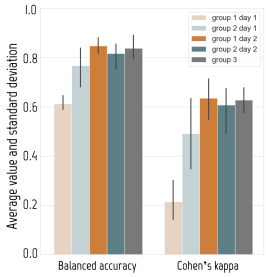

Figure 3. QS estimation performances as a function of age.

4. Conclusion

This study introduced a method to estimate QS by combining ECG and respiratory signal features with motion data. Best model for cardiorespiratory feature classification was combined with motion segmentation using an "AND" for final QS estimate. The approach's effectiveness was evaluated using a real-life clinical database with data from newborns of different GA and PMA. Results indicated age-dependent performance, with less accurate results for very premature babies, especially those monitored immediately after birth. Their significant immaturity may result in less comprehensive information available from ECG and respiration data, making QS detection more challenging. Future work involves optimizing the fusion of cardiorespiratory classification and motion data and aiming to estimate all sleep states in newborns. Upcoming plans also include a comprehensive clinical study to demonstrate the benefits of monitoring QS in NICU for tracking newborn maturation.

References

[1] Cailleau L, Weber R, Cabon S, *et al*. Quiet sleep organization of very preterm infants is correlated with postnatal maturation. Front Pediatr Sep. 2020;8.

[2] Werth J, Radha M, Andriessen P, *et al*. Deep learning approach for ECG-based automatic sleep state classification in preterm infants. Biomed Signal Process Control 2020; 56:101663.

[3] Isler J, Thai T, Myers M, Fifer W. An automated method for coding sleep states in human infants based on respiratory rate variability. Dev Psychobiol 2016;58(8):1108–1115.

[4] Sentner T, Wang X, de Groot E, *et al*. The sleep well baby project: an automated real-time sleep-wake state prediction algorithm in preterm infants. Sleep 2022;45(10):zsac143.

[5] Morita K, Shirai N, Shinkoda H, *et al*. Automatic neonatal alertness state classification based on facial expression recognition. J Adv Comput Intell Intell Inform 2022; 26(2):188–195.

[6] Cabon S, Porée F, Simon A, *et al*. Audio-and video-based estimation of the sleep stages of newborns in neonatal intensive care unit. Biomed Signal Process Control 2019; 52:362–370.

[7] Digi-NewB - GCS HUGO - CHU - Monitoring system. http://www.digi-newb.eu. 14 April 2020.

[8] Doyen M, Ge D, Beuchée A, *et al*. Robust, real-time generic detector based on a multi-feature probabilistic method. Plos one 2019;14(10):e0223785.

[9] Khreis S, Ge D, Rahman H, Carrault G. Breathing rate estimation using Kalman smoother with electrocardiogram and photoplethysmogram. IEEE Trans Biomed Eng 2019; 67(3):893–904.

[10] Makowski D, Pham T, Lau Z, *et al*. Neurokit2: A Python toolbox for neurophysiological signal processing. Behav Res Methods 2021;1–8.

[11] Cabon S, Weber R, Simon A, *et al*. Functional age estimation through neonatal motion characterization using continuous video recordings. IEEE J Biomed Health Inform 2022; 27(3):1500–1511.

Address for correspondence:

Houda Jebbari
LTSI INSERM U1099, University of Rennes
houda.jebbari@univ-rennes.fr