STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	M. Lavanga *et al.*, "A perinatal stress calculator for the neonatal intensive care unit: an unobtrusive approach," *Physiol. Meas.*, vol. 41, no. 7, p. 075012, Jun. 2020.
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	Published
Journal homepage	https://iopscience.iop.org/journal/0967-3334
Author contact	your email mlavanga@esat.kuleuven.be your phone number +32 16 37 38 28
Abstract	Objective: Early experience of pain and stress in the neonatal intensive care unit is known to have an effect on the neurodevelopment of the infant. However, an automated method to quantify the procedural pain or perinatal stress in premature patients does not exist. Approach: In the current study, EEG and ECG data were collected for more than 3 hours from 136 patients in order to quantify stress exposure. Specifically, features extracted from the EEG and heart-rate variability in both quiet and non-quiet sleep segments were used to develop a subspace linear-discriminant analysis stress classifier. Main results: The main novelty of the study lies in the absence of intrusive methods or pain elicitation protocols to develop the stress classifier. Three main findings can be reported. First, we developed different stress classifiers for the different age groups and stress intensities, obtaining an area under the curve in the range [0.78-0.93] for non-quiet sleep and [0.77-0.96] for quiet sleep. Second, a dysmature EEG was found in patients under stress. Third, an enhanced cortical connectivity and increased brain–heart communication was correlated with a higher stress load, while the autonomic activity did not seem to be associated to stress exposure. Significance: The results shed a light on the pain and stress processing in preterm neonates, suggesting that software tools to investigate dysmature EEG might be helpful to assess stress load in premature patients. These results could be the foundation to assess the impact of stress on infants' development and to tune preventive care.

IR

(article begins on next page)

A perinatal stress calculator for the neonatal intensive care unit: an unobtrusive approach

M Lavanga[1], L Smets[1], B Bollen[2], K Jansen[2], E Ortibus[2], S Van Huffel[1], G Naulaers[2], A Caicedo[3]

[1] Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Kasteelpark Arenberg 10, box 2446, 3001, Leuven, Belgium

[2]Department of Development and Regeneration, Neonatal Intensive Care Unit, UZ Leuven, UZ Herestraat 49, box 7003 21, 3000 Leuven

[3]Department of Applied Mathematics and Computer Science, School of Engineering, Science and Technology, Universidad del Rosario, Bogotá, Colombia

May 27, 2020

Abstract. *Objective*: Early experience of pain and stress in the neonatal intensive care unit is known to have effect on the neurodevelopment of the infant. However, an automated method to quantify the procedural pain or perinatal stress in premature patients does not exist. *Approach*: In the current study, EEG and ECG data were collected for more than 3 hours from 136 patients in order to quantify stress exposure. Specifically, features extracted from the EEG and heart-rate variability in both quiet and non-quiet sleep segments were used to develop a subspace linear-discriminant analysis stress classifier. *Main Results*: The main novelty of study lies on the absence of intrusive methods or pain elicitation protocols to develop the stress classifier. Three main findings can be reported. First, we developed different stress classifiers for the different age groups and stress intensities, obtaining an area under the curve in the range [0.78-0.93] for non-quiet sleep and [0.77-0.96] for quiet sleep. Second, a dysmature EEG was found in patients under stress. Third, an enhanced cortical connectivity and increased brain-heart communication was correlated with a higher stress load, while the autonomic activity did not seem to be associated to stress exposure. *Significance*: Those results shed a light on the pain and stress processing in preterm neonates, suggesting that software tools to investigate dysmature EEG might be helpful to assess stress load in premature patients. Those results could be the foundation to assess the impact of stress on infants' development and tune preventive care.

Submitted to: *Physiol. Meas.*

1. Introduction

Early experience of pain and stress in premature infants has been under greater scrutiny by the clinical investigators due to the long-term effects on development [1]. Concerns about the impact of infant pain on neurodevelopment have been raised in the 80s [2].

During their stay in the neonatal intensive care unit (NICU), infants undergo different procedures and care which can generate a cascade of behavioral, physiological and hormonal responses [1]. This accumulation of painful and/or stressful procedures can lead to higher stress, which is defined as perinatal stress. Although procedural pain is not necessarily associated to background stress, Ranger [3] pointed out that pain and stress cannot be discriminated in clinical practice and Jones [4] has shown that there is an increase in pain reactivity in case of high exposure of background stress. Due to the possible developmental implications and the worse outcomes of premature infants, clinicians are interested to assess the effect of early-life stress during the patient's stay in the NICU and later on in their life. It is important to remind that stress is a broader concept than pain only. However, the pain-related stress is normally investigated since mainly validated pain scales are used as measurement. Therefore, stress or pain experience will be both used as synonym of *perinatal stress*.

The growing interest in perinatal stress has been shown in a variety of recent studies. Brummelte showed that perinatal stress can affect the fractal anisotropy of the subcortical white matter during infant's neurodevelopment [5]. Cong has shown how kangaroo care (KC) can alleviate the effect of post-stimulus experience at the autonomic level after heel prick procedure [6]. Similarly, studies that investigated EEG recordings during pain exposure relied on heel lance procedures and inoculation to assess brain synchronization in nociceptive processing, as reported in [7], [8]. Specifically, the authors have investigated differences between tactile and noxious stimuli in evoked related potentials (ERPs) on the one hand and the involvement of autonomic circuitry in pain processing on the other hand. Furthermore, handling of patients induces also sleep disturbances that can generate long bradycardias and apneas [9].

However, there is no automated method to classify exposure to stress in premature infants based on physiological signal background. What is more common is pain-patterns classification based on biopotentials information in the adult population. Misra [10] has developed a SVM model to classify low-pain and high-pain using EEG recordings in adults. In their study, they have shown that θ and γ power increase in the prefrontal region and β power in the sensorimotor region contralateral to the stimulus, which can predict the pain patterns with an accuracy of 89.58%. Vijayakumar [11] used a wavelet-transform of EEG to derive power in multiple frequency bands and discriminated multiple pain stimuli in adults by means of random forest classifiers. Other modalities have also been investigated to automatically classify pain patterns in the adult population. Gruss [12] has investigated patterns in EMG under thermal stimulation. A noteworthy aspect was not just the automatic approach to discriminate between baseline and pain patterns, but also the reported accuracy in function of the pain intensity. Another example comes from Brown who used fMRI data to develop a SVM model to determine the absence and presence of pain during heat stimuli [13].

Besides machine learning approaches, the physiological reaction to pain has been investigated by means of the autonomic response or EEG connectivity patterns in adults. Loggia [14] has shown the magnitude of the heart-rate and skin-conductance (SC) sweeps increase with increasing thermal pain stimulation. Functional cortical connectivity can also vary according to stress or pain disorders, as reported by De Tommaso [15] and Imperatori [16]. Migraine patients tend to have a greater connectivity at resting state than controls in laser-induced pain experiments [15]. In addition, post-traumatic stress disorder seems to lead to a higher power in the θ band and higher connectivity in the α band [16].

The objective of this study is to develop a classification model to detect the presence of stress exposure in premature infants. A binary model to discriminate between stress and low-stress patterns was derived by multiple features extracted for background physiological activity. As mentioned earlier, perinatal stress is here intended as pain-related stress. Since there is no clear definition of stress in the NICU and there is no consensus on the level under which the effect can be considered negligible or low, different levels of accumulation of procedural pain have been investigated to define stress.

2. Materials and methods

2.1. Patient sample

The current study used data from the Resilience Study, a prospective longitudinal, cohort study conducted in the Neonatal Intensive Care Unit (NICU) of the University Hospitals Leuven, Belgium. Parents of preterm infants born before 34 weeks gestational age (GA) and/or with a birth weight less than 1500 g were informed within the first three days after birth. Exclusion criteria were parents age < 18 years, absence or limited knowledge of Dutch or English, medical (somatic or psychiatric) condition in the parent(s) that impeded participation, and the presence of a major congenital malformation or central nervous system pathology (grade 3 or grade 4 intraventricular hemorrhage or periventricular leukomalacia) at the time of consent. Preterm infants (n=136) were included in the study between July 2016 and 2018.

The research protocol was examined and approved by the Ethical Committee of University Hospitals Leuven, Belgium. The study was performed in accordance with the Guidelines for Good Clinical Practice (ICH/GCP) and the latest version of the Declaration of Helsinki. It was registered at Clinical Trials.gov (NCT02623400).

2.2. Data acquisition

Patients' pain levels were daily recorded at the cot-side. Pain scores were assessed using the Leuven Pain Scale (LPS), a validated multidimensional pain scale for preterm infants [17], [18]. The LPS assigns scores (0,1 and 2) to seven categories (facial expression, cry-

ing, irritability, drowsiness, muscle tension, comfort, and heart rate). Thus, LPS scores vary between 0 and 14, with higher scores corresponding to higher pain intensity. A score of 4 is regarded as the critical threshold for distinguishing a comfortable condition from an uncomfortable one (not pain-free and/or acute pain). The inter-rater reliability was 0.88 [17],[18]. Pain assessment was done routinely by bed-side nurses that were familiar with the LPS and noted it in the patient's electronic medical file. As part of the clinical routine, LPS was scored hourly in preterm infants receiving intensive care, and every three hours in infants receiving intermediate care.

Following the definition of stress exposure, stress has been regarded as the experienced pain in the day before the experimental recording, that is the presence of non-zero LPS in the patient record. As already mentioned, a common threshold for uncomfortable conditions for LPS is 4. However, in line with [12], multiple thresholds for experience of pain in the day before the recording were tested: $LPS > 0$, $LPS > 1$, $LPS > 4$ to define a patient under stress conditions. The main assumption is that a greater threshold should lead to a better stress classification performance.

EEG and ECG measurements were simultaneously measured at three different time points for at least 3 hours. Importantly, infants had to be clinically stable at the time of recording. The first recording took place around 5 days after birth (5days). The second recording was planned around 34 weeks postmenstrual age (PMA) (34w). For infants born at 33-34 weeks GA, only one of the two recordings from birth was performed. A last recording consisted of a 24-hour polysomnography (PSG) that was conducted in the week before discharge home. The parents only consented for the polysomnography and opted out for additional EEG measurements. In the course of their NICU stay, some infants were transferred to level II units in hospitals closer to home. Therefore, not all infants have multiple recordings and some LPS measures are missing. However, we strived to readmit infants to our hospital for the PSG measurement. A total of 245 recordings had corresponding pain scores available and were analyzed. Table 1 summarizes the clinical characteristics of patients at each measuring point. EEG data were collected according to the 10-20 system using nine monopolar electrodes (F_{p1}, F_{p2} , C_3, C_4 , C_z, T_3, T_4, O_1, O_2) and monitored with the OSG system (OSG BVBA, Brussel). Each EEG signal was referenced to C_z, which was then excluded from further analysis, leaving a total amount of 8 channels. ECG was then used to derive the tachogram or HRV signal as subsequent R-peak to R-peak intervals (RRi). The R-DECO toolbox [19] was used for R-peaks detection.

2.3. Preprocessing and sleep-stage analysis

EEG was band-pass filtered between 0.5 and 20 Hz and independent component analysis was used to remove EOG artefacts. To determine whether the channels were contami-

Table 1. Summary of patient data set at different time points: GA (gestational age), birth weight (in g), PMA (postmenstrual age) at EEG and ECG recording. Data are median [interquartile range].

	5days (n=118)	34w (n=67)	PSG (n=117)
GA(*weeks*)	31.14 [28.86-32.43]	28.86 [26.86-30.71]	30.29 [27.29-31.71]
Birth Weight (*g*)	1475 [1120-1725]	1140 [900-1480]	1225 [950-1540]
PMA(*weeks*)	32.14 [30-33.43]	34.14 [33.86-34.29]	38.43 [37.29-39.57]
LPS	1 [0-3]	0 [0-2]	0 [0-2]

nated by movement-related or non-cortical artefacts, multiple criteria were applied on a window-basis in the feature extraction step. Criteria were as follows: standard deviation below 50 μV, absolute difference sample-to-sample below 50 μV and absolute amplitude below 200 μV [20]. If more than 4 channels had an artifact in a window, that window was excluded.

In order to compensate the effect of premature ventricular contractions, the RR intervals were corrected as discussed in [19]. In addition, the R-peak locations were also used to derive the modulation signal m(t) via the integral pulse frequency modulation (IPFM) model, which also accounts for premature contractions [21].

In order to assess whether sleep-wake ciclicity can influence stress classification, sleep states were automatically derived from the EEG recordings. Sleep is characterized by the development of behavorial states with the maturation of the infant and can be divided in quiet-sleep, active sleep and awake [22]. Generally, the EEG signal is relatively more discontinuous during quiet sleep, while the other two states present a more continuous tracing as well as a higher variability of the cardiorespiratory pattern and more body movements [23],[24]. Most of the data-driven algorithms focus on the detection of QS, while the other two states are normally merged in one single state defined as non-quiet sleep. The probability of Quiet Sleep (QS) was derived from the EEG signal with 2 data-driven approaches: a Convolutional Neural Network (CNN) model by [25] and the CLASS algorithm reported in [26]. These two methodologies generate a profile for QS occurrence from the full-channel EEG activity, which was used to derive two 20 minutes window, one associated to QS and the other to non-quiet sleep information (nQS). Each of the window was respectively located around the maximum or the minimum of the probability profile for QS and nQS, considering epoch 10 minutes before and after each prominence.

2.4. Univariate features extraction

Multiple features were extracted on a single channel basis. These attributes span from the classical spectral analysis to more refined nonlinear approaches and are meant to describe the presence of dysmature EEG and HRV. Following the definition by Pavlidis

[27], a dysmature EEG is generally characterized by discontinuity, persistence of slow-waves and general lack of smoothness, while a HRV of a dysmature neurovegetative system is characterized by lower complexity and and a prominent slow-wave baseline. Different studies have shown that pain elicitation can generate burst-type activity in premature infants [28] and the patients that stay longer inside the NICU are the candidates to have a greater negative impact on the neurodevelopment [1]. Therefore, the relationship among dysmature EEG, dysmature HRV and stress can be investigated. In the next section, a general overview of the algorithm to derive those features will be presented.

2.4.1. Power features The power spectral density of the tachogram was computed with both the Welch-periodogram and the continuous wavelet transform. For the periodogram, 70% was chosen as overlap between windows, while analytical Morlet was picked as the mother wavelet for the wavelet transform [29]. In case of EEG, the power was only computed using the Welch Periodogram for the following frequency bands: $\delta_1 = (0.5 - 2]\ Hz$, $\delta_2 = (2 - 4]\ Hz$, $\theta = (4 - 8]\ Hz$, $\alpha = (8 - 16]\ Hz$ and $\beta = (16 - 20]\ Hz$. The power-features were computed in non-overlapping windows of 30 sec and 4 sec subwindows and it was grand-averaged along QS or nQS channel wise [30], [31]. In case of HRV, the absolute powers of high-frequency (HF), low-frequency (LF) and very-low frequency band (VLF) were computed as sum of the PSD bins in the following frequency bands: $HF = (0.2-4]\ Hz$, $LF = (0.08-0.2]\ Hz$, $VLF = (0.0033-0.08]\ Hz$ [32], which slightly differs from the frequency bands definition reported in [33]. The relative power indices were also derived as $\frac{VLF}{LF}$,$\frac{LF}{HF}$,$\frac{LF}{LF+VLF}$,$\frac{LF}{HF+LF}$. Both, the modulated signal $m(t)$ and the resampled HRV, were used for this analysis. The spectral density was computed in QS and nQS epochs using the entire 20 minutes [34],[35]. The Welch periodogram subwindow was set to 5 minutes. In case of the wavelet transform, the time-frequency values were then averaged in both sleep epochs.

2.4.2. Multifractality A more discontinuous signal is normally characterized by a lower entropy, since its pattern has higher predictability or memory-persistence. This means that the signal can be easily predicted by its past sample and, therefore, a more discontinuous signal is characterized by higher regularity. Normally, those signals are characterized by a long-range autocorrelation property and a power-law spectrum, whose rate of decay is controlled by the Hurst exponent. In the specific case that a time series is characterized by one exponent, this signal is also defined as self-similar (it repeats itself over time) or monofractal. However, regular or discontinuous signals might have multiple exponents to control the degree of regularity over time. Those signals are known as multifractals. In general, signals with a high-degree of regularity are known as fractals or scale-free signals. An efficient way to estimate the Hurst exponent H is based on the wavelet transform [36]. Wendt proposed to estimate the spectrum of singularities via wavelet leaders, which measures the different Hurst exponents in the signal and their associated fractal dimension. The c_1,c_2,c_3 parameters respectively represent the location

of the maximum, the width and the asymmetry of the singularity spectrum. Therefore, c_1 is usually considered the main Hurst exponent from the multifractal signal, while c_2 and the difference between the maximal and minimal Hurst exponent are variational indices to represent the dispersion of fractals inside the signal. Further details of the methodology are reported in [36], which also describes the WLBFM toolbox to estimate the fractal parameters in MATLAB.

In case of EEG, the triplet c_1,c_2,c_3 , the maximal and minimal Hurst exponents and their difference ΔH were estimated for each EEG channel in non-overlapping windows of 150 s and the results were averaged along the 20 minutes window channel wise for both QS and nQS. The window 150 s was picked as epoch to assess the long-term correlation of EEG, according to the findings reported by [37],[38],[39].

In case of HRV, the triplet c_1,c_2,c_3, the maximal and minimal Hurst exponents and ΔH were derived for the entire 20 minutes window for both QS and nQS epochs, in order to assess the slower oscillations which drives the HRV power-law spectrum [35],[34].

2.4.3. Multiscale Entropy The persistence of slow-waves, the discontinuity and the predictability of a signal can be measured with sample entropy ($SampEn$), which counts m-long templates matching within a certain tolerance r (usually defined as 20% the standard deviation of the signal) to assess the probability of having similar patterns [40]. However, in order to take into account the irregularities across different scales, Costa el al. [41] proposed the multiscale sample entropy (MSE), which measures SampEn at different scales τ using a coarse-grained version of the signal of interest. In physiological signal processing, $SampEn$ has successfully been used in the prediction of sepsis and assessment of sleep quality in premature infants [42], [40], [40], while the multiscale entropy has been used in prediction of congestive heart failure [41] as well as in sleep-state detection in newborns [37].

In this study, both $SampEn$ and MSE were computed to predict stress levels. In particular, $SampEn(m, r)$ and its variations quadratic sample entropy and coefficient of sample entropy $QSE(m)$ and $COSE(m)$ were computed with m spanning in the range $m = 2, 3, 6$, following the findings by Lake [40] and Li [43], while the MSE curve is derived for the scales $\tau = 1, ..20$, as originally discussed by [41]. The complexity index was then derived via the area under the MSE curve, $CI = \sum_\tau MSE(\tau)$. In addition, the MSE at scale 3 and 20 ($MSE(\tau = 3)$ and $MSE(\tau = 20)$) were also considered [44]. It is important to remember that $SampEn$ and the MSE at scale 3 represent the information at small scales or high frequency, while MSE at scale 20 represents the information at longer scales or lower frequency. CI is a general measure of irregularity across scales. In case of EEG, those features were computed in 150 sec non-overlapping windows and grand-averaged along the time course of each sleep state, while the entire 20 minutes window was used for entropy features of the HRV. As already mentioned,

the length of the epoch was derived according to the entropy analysis [37] and the multifractal analysis by [38],[39].

2.4.4. Neural features The dysmature EEG features can also be quantified by means of the NEURAL toolbox [45], available on GitHub. The approach proposed by the authors yields an exhaustive range of features that can be obtained from amplitude, burst and spectral information, connectivity analysis and the range EEG (rEEG). As mentioned by O'Toole [45], the range EEG (rEEG) was proposed as an alternative to amplitude-integrated EEG (aEEG), since a clear definition of the amplitude integrated EEG is missing and different EEG machines implement different algorithms for its estimation. The range EEG represents a filtered, compressed and rescaled version of the EEG: it is normally band-pass filtered between [2-15] Hz, compressed and rectified in the time domain with a windowing procedure and rescaled in a linear-log amplitude scale. More details are reported in [45]. Among all the attributes, it is worth to mention that rEEG was used to quantify the level of dysmaturity by several morphology indices, such as its lower and upper margin, its standard deviation and the rEEG asymmetry. These two margins represent respectively the 5th and 95th percentiles of the rEEG, while the rEEG asymmetry expresses the difference in distance from the median to these 2 margins. More details are reported in [45]. These features were derived only for the EEG and they were obtained using windows of 2 s with 50% overlap for spectral features and 2 s windows without overlap for amplitude features. All the indices were then grand-averaged for each sleep state.

2.5. Multivariate features

Features were also extracted in a multivariate fashion to describe how signals interact in case of stress. Specifically, EEG cortical connectivity and brain-heart interaction were investigated. The former was estimated by means of functional connectivity methodologies [31], while the latter was derived via the wavelet coherence [46] and the phase dependent dynamics [47].

2.5.1. EEG connectivity Functional cortical connectivity was analysed by means of lagged and instantaneous connectivity as well as synchronization of oscillatory processes from the EEG channels [31], [48], [49]. The greatest advantage of lagged connectivity is the resistance to volume conduction and low spatial resolution [50]. Therefore, the connectivity among EEG channels was estimated with coherence as $C_xy(f) = \frac{P_{xy}(f)}{P_{xx}(f),P_{yy}(f)}$, where $P_{xy}(f)$ is the cross-spectrum between two channels x and y, while the $P_{xx}(f)$, $P_{yy}(f)$ are the autospectra of the two signals. The instantaneous connectivity is derived via magnitude squared coherence, which represents the squared magnitude of the coherency, i.e. $k^2_{xy}(f) = |C_{xy}(f)|^2$. The lagged connectivity is derived via imaginary coherence, which represents the imaginary part of the coherence, i.e. $I_{xy}(f) = \mathcal{I}(C_{xy}(f))$. Synchronization was computed with phase-locking approaches,

which require to define the phase ϕ of the main frequency bands. Therefore, each signal was decomposed via a discrete wavelet transform according to the following dyadic split δ_1, δ_2, θ, α and β and the phase of each frequency band was then derived via the Hilbert-Transform. The synchrony between the phase $\phi_x(t)$ of x and the phase $\phi_y(t)$ of y was then derived via the phase locking value $PLV_{xy} = \frac{1}{n}\left|\sum e^{j\phi_{xy}(t)}\right|$ and phase lag index $PLI_{xy} = \frac{1}{n}\left|sgn(\phi_{xy}(t))\right|$, where $\phi_{xy}(t)$ represents the phase difference between $\phi_x(t)$ and $\phi_y(t)$ and sgn is the sign function. The statistical validity of each coupling was then tested with amplitude adjusted Fourier transform surrogates. Specifically, each coupling must be greater in value than the coupling estimated for 19 surrogates, in order to guarantee a level of statistical significance $\alpha = 0.05$. In case of coherency, the surrogates approach generates a threshold $T(f)$ in function of the frequency, which was used to zeroed the frequency bins in both $k^2_{xy}(f)$ and $I_{xy}(f)$ that didn't pass the surrogate testing. In case of the magnitude squared coherence, the $k^2_{xy}(f)$ was averaged in the bands δ_1, δ_2, θ, α and β. In case of the Imaginary Coherence, the maximal amplitude of $I(C_{xy}(f))$ was then considered in the same frequency bands [48]. The EEG connectivity coherency was estimated using a Welch-periodogram approach for non-overlapping 30 s window. The subwindow was set to 4 secs and overlap to 70%, according to the findings by [29],[31]. Similarly, The phase locking value and phase lag index were computed in non-overlapping 30 s window, based on the previous analysis on premature infants EEG connectivity [31], [49].

2.5.2. Brain-heart interaction: Wavelet Coherence The interaction between brain and heart can be estimated as the correlation in the frequency domain between the envelope of δ power and the heart-rate variability. The method by Piper consists of the time-frequency coherence between the δ oscillations derived with the continuous wavelet transform and the HRV [46]. In order to match the temporal scale, both signals were resampled at 8 Hz. The continuous wavelet coherence was then computed as $C_{BH}(t,f) = \frac{s_{BH}(t,f)}{s_B(t,f)s_H(t,f)}$, where $s_{BH}(t,f)$ is the cross-scalogram between the brain time-course and heart time-course, $s_B(t,f)$ is scalogram of the brain time-course and $s_H(t,f)$ that of the heart time-course. The different scalograms were computed with analytic Morlet as mother wavelet. In order to assess the discriminatory power of the brain-heart, the coherence was investigated in the following frequency bands: VLF, LF and the combined band of VLF and LF band $VLFLF = (0.033{-}0.2]\ Hz$. The main reason is the frequency shift that HRV undergoes with maturation of the neurovegetative system from VLF to LF band [44]. In addition, stress information is commonly associated to the low-frequency band of HRV. Therefore, the brain-heart interaction has been tested with LF oscillations as heart time-course. Similarly to the scalp connectivity, the coupling value was derived as mean of the squared time-frequency coherence in the considered band or the absolute maximum of the imaginary part of the time-frequency coherence. The continuous wavelet coherence was derived for the entire 20 minutes in both QS and nQS epochs and the investigation was limited to channels C_3 , C_4 and HRV, following

the studies in adults by Faes et al. [51] and in premature infants by Pfurtscheller [52]. In order to assess the statistical significance, we used amplitude-adjusted Fourier transform surrogates similarly to EEG functional connectivity.

2.5.3. Brain-heart interaction: Phase dependent dynamics The dependent dynamics between two systems can be investigated via phase reconstruction, as shown by Rosenblum et al. [47]. Their methodology estimates the phase dynamics of interacting oscillators and provides a directionality index, which describes the direction and the intensity of the coupling. The directionality index is a value bounded between [-1,1], which gives an integrated measure of how strongly a system drives (a value close to 1) or how sensitive it is to be driven (close to -1). The main limit of this methodology is the estimation of the phases, since different methods exists with different underlying hypotheses. The phase of a signal can be estimated via the Hilbert Transform in case of an oscillatory process (e.g. a band-passed EEG signal in the delta band). In case of a point-process like the HRV, the phase of the signal has to be derived taking into account this pulse-type of information. Therefore, the modulatory signal $m(t)$ was considered in the reconstruction of the phase dynamics between brain and heart. In addition, since both EEG and m(t) are wide frequency band signals, the phases of the signals were derived as phases of the main carrier frequency from those signals. The main carrier frequency is defined as the frequency with highest power in the signal. In case of the neonatal EEG, the burst activity around 1 Hz is considered the main carrier frequency of the cortical activity [53], while the pacemaker of the modulatory activity is centered on 0.1 Hz, which is between VLF and LF bands [52]. A method to derive the phase of the carrier frequency is the ridge wavelet-transform, which is wavelet transform that finds the carrier frequency by looking at the maximum of the spectrum in each time-point [54]. The directionality index was derived only for the channel C_3 and $m(t)$ and the entire 20 minutes window was used to derive the phases of the two signals and compute the interaction. The significance for the directionality index was tested with Cycle-Phase permutation testing, which randomly permutes cycles of a signal phase to generate surrogates [55]. 19 surrogates were derived to reach a level of statistical significance $\alpha = 0.05$. However, the results of the directionality index strictly depends on the order of finite Fourier series to estimate the phase dependencies [47]. Therefore, the order was tested in the range [1:10]. The ridge-wavelet transform was estimated with MODA toolbox [56], while the directionality index was derived with the DAMOCO toolbox [57] (both implemented in MATLAB).

2.5.4. Graph theory A graph is a diagram of points or nodes, whose relations are defined by the connecting lines among them. If each node represents a signal, the relationship among the time series can be described by a weighted network, where each link underlines the existence of coupling and the weight represents the intensity of coupling. Specifically, a multivariate methodology generates a coupling matrix $M \times M$ defined as $A_{ij} = C_{x_i \leftrightarrow x_j}$, where $C_{x_i \leftrightarrow x_j}$ is the general coupling intensity, M is the

number of processes and $i, j = 1, .., M$. The matrix $A = A_{ij}$ can be interpreted as an adjacency matrix of an undirected graph, with weighted edges [58]. If the direction of interaction is not specified (as underlined by $x_i \leftrightarrow x_j$), A_{ij} is symmetric since its entries represent statistical correlations without any specific direction. Those specific entries, A_{ij}, define the flow of information among nodes in the network and therefore its topology. Consequently, a list of topological indices have been thoroughly discussed to describe the architecture of a network [58],[59]. Among the topological measures, the path length, the clustering coefficient and the eccentricity are the most common indices to assess the network's level of integration. Specifically, the path length is the average shortest path between two nodes in the network. The lower the value the higher the integration in the network. The eccentricity of a node represents the maximum distance from one node to any other node in the graph, while the clustering coefficient is defined as the average of all weighted triangles around a node and mirrors the graph coupling density. A triangle is here intended as the smallest non-trivial motif that a node can form with two neighboring nodes. Alongside the topological indices, the amount of superfluous connections of the adjacency matrix is introduced. Any graph can risk to be overly connected or under-connected. Even after surrogate testing, some connections can emerge as significant, but they result from physiological redundant connectivity [61]. Therefore, we can test the resilience, which is the capacity of the network to keep the global connectivity high even if some connections are removed. Therefore, suppose we order all the connections in descending order based on the intensity of the coupling. Suppose also the set of original weights of A is defined as w_{ij}^0. The number n_{sup} of superfluous connections is derived as the number that maximize the following quantity

$$\max_n H(w_{ij}(n)) + E(w_{ij}(n)) = -\sum_{ij} w_{ij}(n) \log(w_{ij}(n)) + \sum_{ij} (w_{ij}(n) - w_{ij}^0)^2$$

where is the entropy of the matrix A where n weights were removed. The values $w_{ij}(n)$ represent the remaining non-zero weights, while $E(w_i j(n))$ is the squared error between the new matrix A and the original matrix. The number n_{sup} represents the number of removed connections that maintain the global connectivity high without significant deviation from the original matrix. The optimization is to remove as many connections as possible, which results in an increasing error and decreasing connectivity entropy. In an extreme scenario, a redundant network keeps the global entropy high even if there are few non-zero coupling in the matrix; and, in general, they will have a higher number n_{sup} of superfluous connections. In this study, graph were applied in two scenarios. The first one was EEG scalp connectivity, where the number M of processes was set to the number of EEG channels ($M = 8$). Specifically, the adjacency matrix was derived every 30 s. The topological indices were computed for each window and averaged along the entire 20 minutes of QS and nQS. The second case relates to brain-heart connectivity computed with wavelet coherence. In this scenario, M shrunk to 3 because only two EEG channels (C_3 and C_4) and HRV were considered. Since

the coupling method is based on a time-frequency approach, the adjacency matrix was computed for each time sample of the 20 minutes window and the derived topological indices were then grand-averaged for each sleep state.

2.6. Sleep-based Classification

A customized software tool was developed using MATLAB built-in machine learning libraries to classify patients into stress or absence of stress conditions. To summarize, the following groups of features were derived for each patient and for both QS and nQS:

- Power features for both HRV and EEG: 7 features for each HRV time-series and 5 features for each EEG channel
- Entropy features: 12 features for each EEG channel and HRV time-series
- Regularity features: 6 features for each EEG channel and HRV time-series
- NEURAL features for EEG: 30 for all EEG channels
- EEG – connectivity topological indices: 76 features for all EEG channels
- Brain-heart connectivity topological indices: 18 features for all EEG channels and HRV combined and 10 directionality indices.

Considering the number of channels, the different frequency bands and the different type of methodologies, 800 features in total were extracted.

The study investigated if there was any association between those features and stress experience in the NICU. As mentioned earlier, the presence of stress was defined as experience of pain the day before the recording. In order to test the different level of pain in the stress modeling, the following thresholds were considered $LPS > 0$, $LPS > 1$, $LPS > 4$ in the definition of the presence of stress (similarly to [12]). Multiple classification methods were tested such as SVMs and linear discriminant analysis (LDA), but subspace ensemble with LDA has been found superior in separating the two classes. Therefore, the results reported here are only for subspace-LDA [60]. Subspace LDA is an ensemble method like random forest, which uses a different learner (LDA, instead of a decision tree) and makes an inherent feature selection to find the best classification subspace or feature subsets to separate the data.

Given that the number of features should be below one tenth of the training dataset size and the datapoints in the different datasets can go as low as 50, the subspace of features has been restricted to 5 [61]. However, before tuning of the model, a basic filtering approach was applied. Features were included in the subspace ensemble algorithm if intra-feature correlation was below 90% and they had the highest ratio between intra and inter variance ratio [61]. In addition, the subject's PMA is an important feature covariate given the maturational changes in both EEG and HRV features. Therefore, if there was a significant Pearson correlation between the selected feature and PMA

($p < 0.05$), the linear correction on the generic feature F $F_{corr} = F - b_1 * age - b_0$ was applied, where F_{corr} was the baseline value deprived of age effects. However, this correction works at best if and only if there is a linear dependence between age and the considered feature. The association between the selected features and the stress experience was also tested via a Kruskal-Wallis test, both for the binary variable $LPS > 0$ and the three groups $LPS = 0$, $LPS = [1 - 4]$, $LPS = 4$.

The classification accuracy was estimated with a leave-one-patient-out (LOPO) scheme and the model was tuned in each training-set with a 10-fold cross-validation. The final accuracy of all test sets was obtained by pooling the class estimates of all patients together. Therefore, only one set of performance indices was obtained for each classifier. Specifically, each model was assessed via the misclassification error (E(%)), the area under the curve (AUC) and the Cohen's kappa between machine learning labels on the pooled test-set obtained with the LOPO scheme.

The stress classification was tested in the three datasets based on 3 monitoring groups ($5days$, $34w$, PSG) and both sleep states. The stress classification was also tested on 4 different regroupings of the recordings based on PMA: ≤ 32 $weeks$, $(32-34]$ $weeks$, $(34-36]$ $weeks$, > 36 $weeks$. The reason to test the classification performance on those different groups is twofold. On the one hand, this classification might still be influenced by age in the monitoring groups (especially in the first group, where the frequency of hands-on care and the heterogeneity of PMA and gestational age are the highest). On the other hand, if a classifier has to be implemented at the cot-side, clinicians tend to be interested in a classifier that considers the age of the patient instead of the monitoring group. This age division follows clinical and physiological definition. The 32 weeks is the threshold to define early prematurity, while 36 weeks is the end of full gestation. The period that goes from 32 to 34 weeks is characterized by synaptogenesis from the subcortical areas towards cortical areas, while the phase after 34 weeks experiences the development of the white matter and the reduction in size of the subcortical plate [62].

3. Results

The results for the stress classification for the monitoring groups are reported in Figure 1 and 2, respectively for QS and nQS. The left panel of both Figures reports the AUC, while the right one shows the kappa score. The AUC, intended as measure for accuracy, is above 0.7 for both QS and nQS. Specifically, the AUC of two sleep states respectively lies in the range [0.73-0.97] and [0.70-0.90]. However, the description of the classifier would be incomplete without reporting the agreement between the ground truth and the predicted labels, expressed by Cohen's kappa. The kappa scores span the range [0.24-0.68] for QS and [0.18-0.68] for nQS, which suggests a moderate association between features and the outcome variable, but also an increased AUC due to an unbal-

anced design of the dataset. In fact, the most noticeable aspect is that kappa does not necessarily increase with increasing pain threshold, but it can also decrease (as shown in the PSG group), which can be due to the presence of an unbalanced dataset.

The results for age groups are reported in Figure 3 and 4. Similarly to the monitoring groups, AUC can respectively span in the range [0.78-0.93] for nQS and [0.77-0.96] for QS and the kappa score span the range [0.18-0.59] for nQS and [0.20-0.63] for QS. As mentioned in the monitoring groups, the AUC can increase beyond 0.7, however Cohen's kappa shows a moderate association between features and outcome variable or the possibility of an unbalanced dataset. Moreover, kappa score can either increase or decrease in function of the pain score and it especially decreases in groups above 36 PMA weeks in nQS. No substantial differences were noticed between the two sleep groups.

Figure 5 shows how the different groups of features contribute to the classification of stress. Results are here only reported for $LPS > 0$. In both sleep states, two main aspects emerge. HRV features consistently underperform, especially in nQS, with AUC never above 0.62. In addition, there is a consistent predominance of EEG and Scalp connectivity features (CONN) for stress discrimination in the young and old age group (respectively $\leq$ 32 weeks and $>$ 36 *weeks*), while brain-heart interaction features can outperform the EEG or connectivity features in the period from 32 weeks to 36 weeks. This is also confirmed by feature's topoplots and boxplots, displayed in figures from 6 to 10. Young patients show a marked decrease in complexity and phase-lag-index eccentricity in θ band in case of stress, as reported in figure 6 and 7. It is important to remember that the lower the eccentricity, the higher the integration in the network (and therefore the connectivity). This persistence of slow-wave activity, higher regularity and connectivity are also present in term babies: figures 9 and 11 report a higher power in δ_2 band and connectivity in δ_1 band for $PMA > 36$ *weeks*. In the period between 32 and 36 weeks, EEG is characterized by a higher rEEG lower margin as well as a higher rEEG asymmetry, as shown in figure 8. On top of that, a higher brain-heart synchrony is present in case of stress, as shown by the lower eccentricity in figure 10. Figure 12 specifically reports the results related to the selected univariate EEG features in a three group fashion, i.e. $LPS = 0$, $LPS = [1 - 4]$, $LPS > 4$, and for almost the entire age range investigated in this study. Those boxplots confirm that the reported differences for $LPS > 0$ persist even if the trend is investigated in three categories. In particular, the EEG shows a marked decrease in complexity with increasing level of stress for $PMA \leq 32$ *weeks* as well as an increased power in δ_2 band for $PMA > 36$ *weeks* and an increased range EEG asymmetry band for PMA $= PMA = (32 - 34]$ *weeks*.

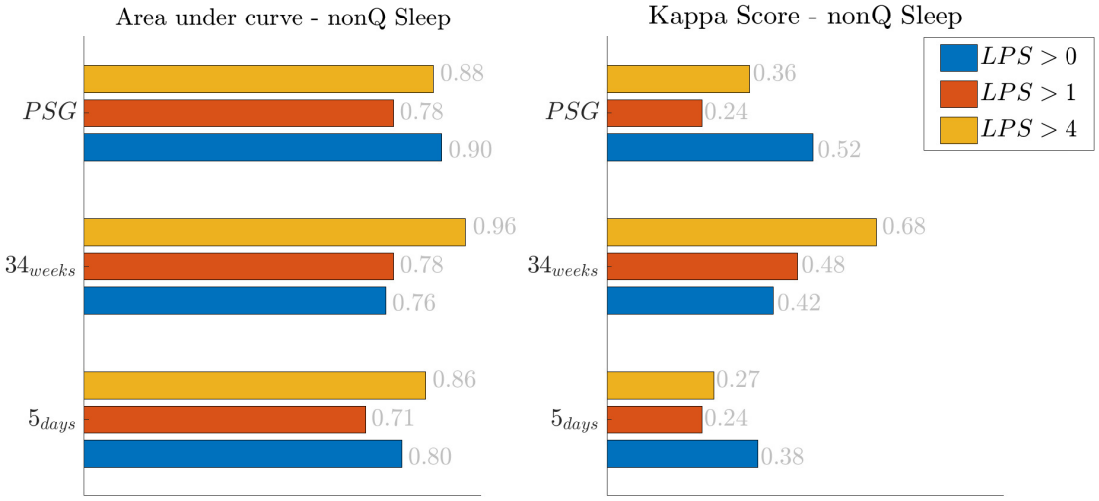

Figure 1. Performance for the stress classifier in function of the Leuven Pain Score (LPS) collected the day before the recording. The results are reported for the three monitoring group for the non-quiet sleep epoch: the recordings within 5 days from birth (5_{days}), the recordings at 34 *weeks* (34_{weeks}) and the polysomnographies at discharge (PSG). The left panel displays the area under the curve, while the right one reports the kappa score. The legend reports the threshold applied on the LPS to define the stress group ($LPS > 0$, $LPS > 1$, $LPS > 4$).

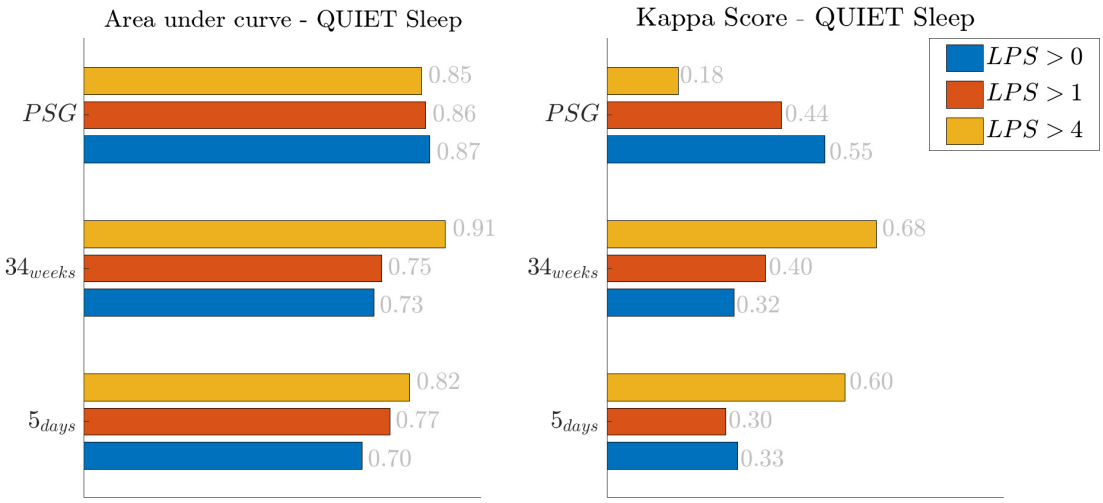

Figure 2. Performance for the stress classifier in function of the Leuven Pain Score (LPS) collected the day before the recording. The results are reported for the three monitoring group for the quiet sleep epoch: the recordings within 5 days from birth (5_{days}), the recordings at 34 *weeks* (34_{weeks}) and the polysomnographies at discharge (PSG). The left panel displays the area under the curve, while the right one reports the kappa score. The legend reports the threshold applied on the LPS to define the stress group ($LPS > 0$, $LPS > 1$, $LPS > 4$).

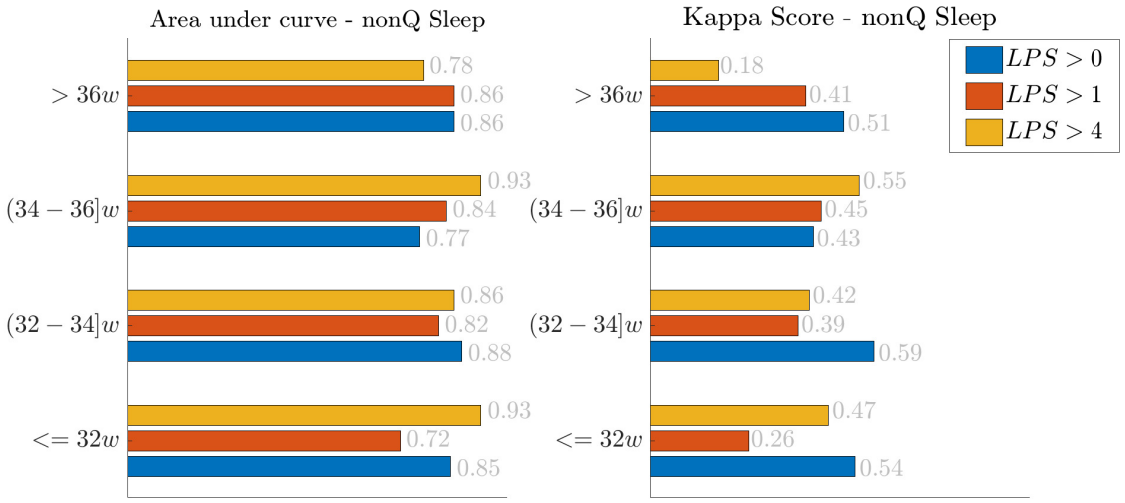

Figure 3. Performance for the stress classifier in function of the Leuven Pain Score (LPS) collected the day before the recording. The results are reported for the 4 post-menstrual age groups for the non-quiet sleep epoch: ≤ 32 *weeks*, $(32 - 34]$ *weeks*, $(34 - 36]$ *weeks*, > 36 *weeks*. The left panel displays the area under the curve, while the right one reports the kappa score. The legend reports the threshold applied on the LPS to define the stress group ($LPS > 0$, $LPS > 1$, $LPS > 4$).

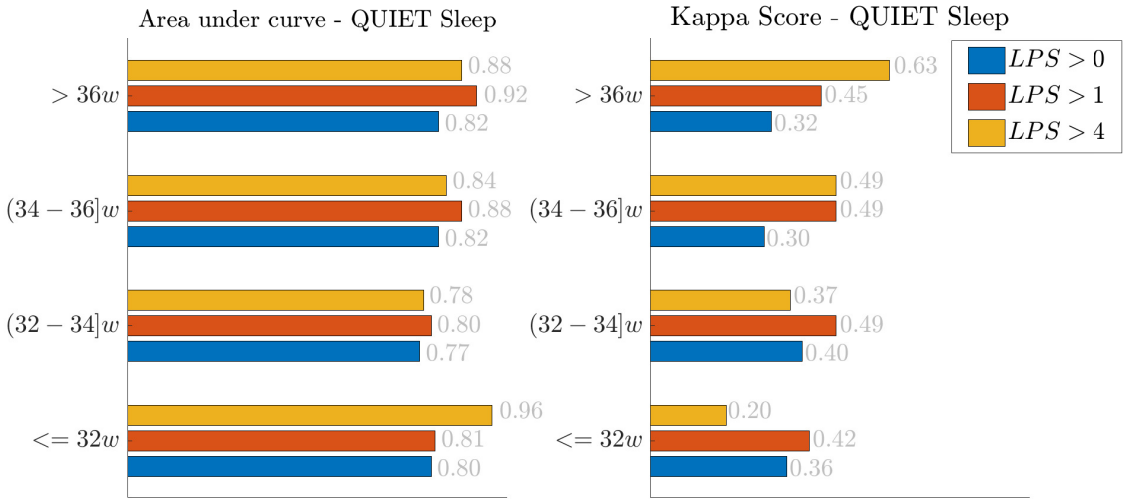

Figure 4. Performance for the stress classifier in function of the Leuven Pain Score (LPS) collected the day before the recording. The results are reported for the 4 post-menstrual age groups for the quiet sleep epoch: ≤ 32 *weeks*, $(32 - 34]$ *weeks*, $(34 - 36]$ *weeks*, > 36 *weeks*. The left panel displays the area under the curve, while the right one reports the kappa score. The legend reports the threshold applied on the LPS to define the stress group ($LPS > 0$, $LPS > 1$, $LPS > 4$).

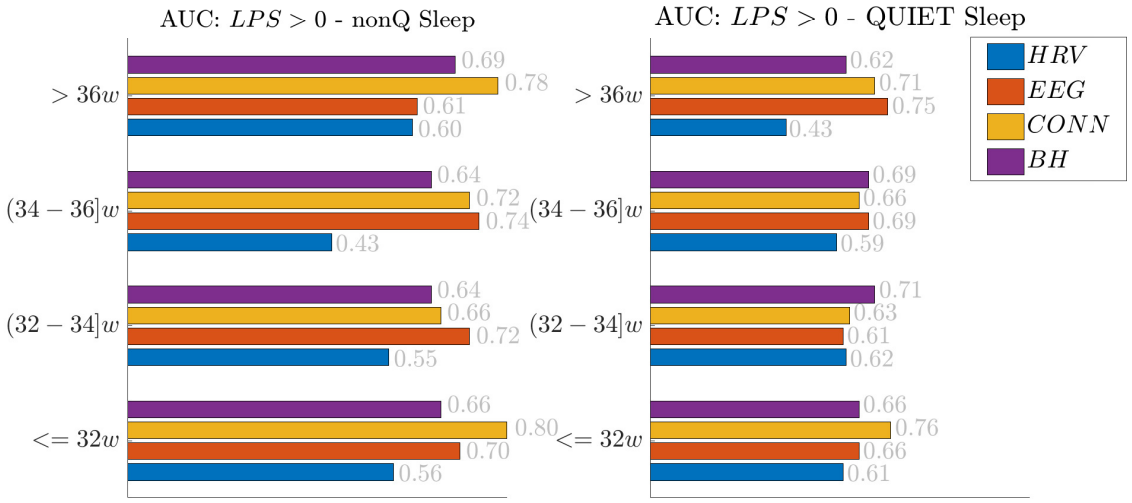

Figure 5. Classification performance for each set of features in discriminating stress with threshold $LPS > 0$. The results are reported for the 4 post-menstrual age groups for the quiet sleep epoch: ≤ 32 *weeks*, $(32-34]$ *weeks*, $(34-36]$ *weeks*, > 36 *weeks*. The left panel displays the area under the curve in non-quiet sleep, while the right one reports the AUC for quiet sleep. The legend reports the four sets of attributes : heart-rate variability features (HRV), EEG features (EEG), EEG connectivity features (CONN), brain-heart synchrony features (BH).

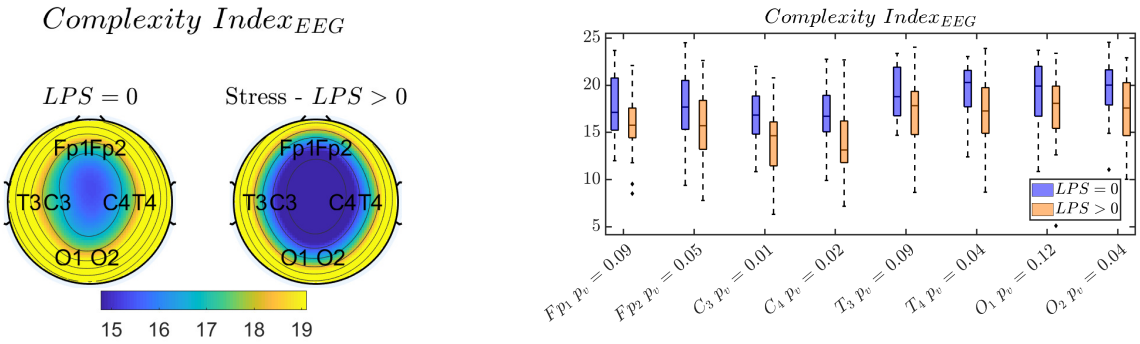

Figure 6. EEG complexity shows a marked decrease in case of stress during nQS for patient with *post-menstrual age* < 32 *weeks*. The left panel reports the topoplot of the complexity index on the scalp (area under MSE curve), while the right one reports the boxplot of the complexity index for each channel (p-value reported with Kruskal-Wallis test). The reduction of EEG complexity can be explained by the greater discontinuity in case of stress.

4. Discussion

In the current study, we performed analysis of brain activity and heart-rate in order to extract features to develop a stress classifier for premature infants at cot-side. Physiological signals, such as EEG and HRV, were collected together with the pain-scale data. The main novelty of study lies on the absence of intrusive methods or pain elicitation

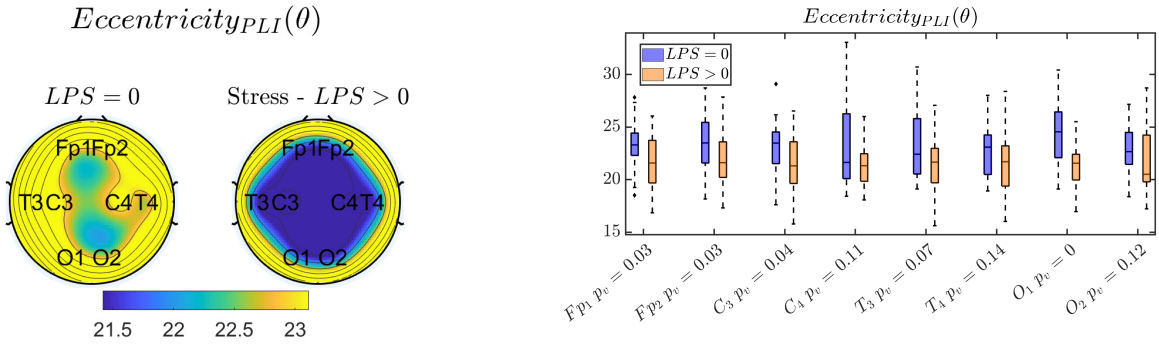

Figure 7. The synchrony in the θ band shows a marked increase in case of stress during nQS for patient with *post-menstrual age* < 32 *weeks*. The left panel displays the topoplot for the eccentricity on the scalp (computed with *phase lag index* in θ band), while the right one reports the boxplot of eccentricity for each channel (p-value reported with Kruskal-Wallis test). The eccentricity is a measure of distance, therefore the lower the eccentricity the higher the connectivity.

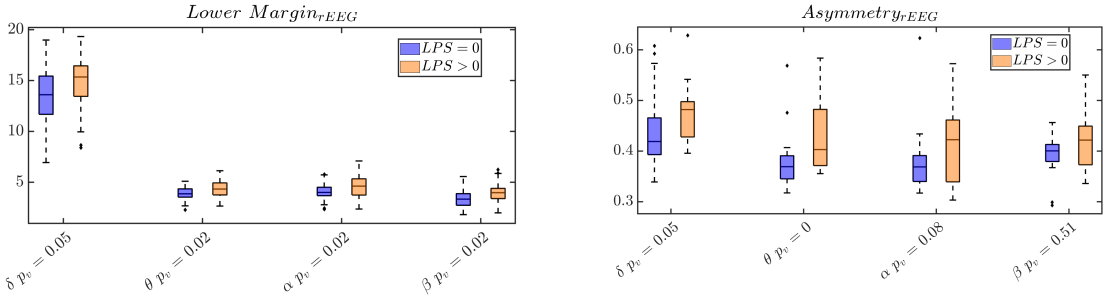

Figure 8. The rEEG showed a marked increase in case of stress during nQS for patients with *post-menstrual age* (PMA) between 32 and 36 weeks. The left panel reports the boxplot of rEEG lower margin for the different frequency bands in patients with PMA between 32 and 34 weeks, while the right one reports the boxplot of rEEG asymmetry for the different frequency bands with PMA between 34 and 36 weeks (p-value reported with Kruskal-Wallis test).

protocols to develop the stress classifier. Three main findings can be reported. First, discontinuous EEG is related with a higher stress load, especially at young age. Second, a higher scalp and brain-heart connectivity are biomarkers of stress experience. Third, the stress classification results are comparable to the pain-pattern classification literature.

The results in figure 6 and 9 shows a lower complexity of EEG for premature infants (PMA ≤ 32 *weeks*) and higher δ_2 power in age-matched premature patients (PMA > 36 *weeks*). In addition, the subjects show a higher rEEG margin and asymmetry in the period from 32 to 36 weeks, as displayed in figure 8. These findings suggest the perinatal stress might increase EEG discontinuity or dysmaturity, which refers to a type of EEG signal characterized by persistence of slow-wave (intended as higher power in δ band and signal regularity), discontinuity (intended as burst pattern which increases

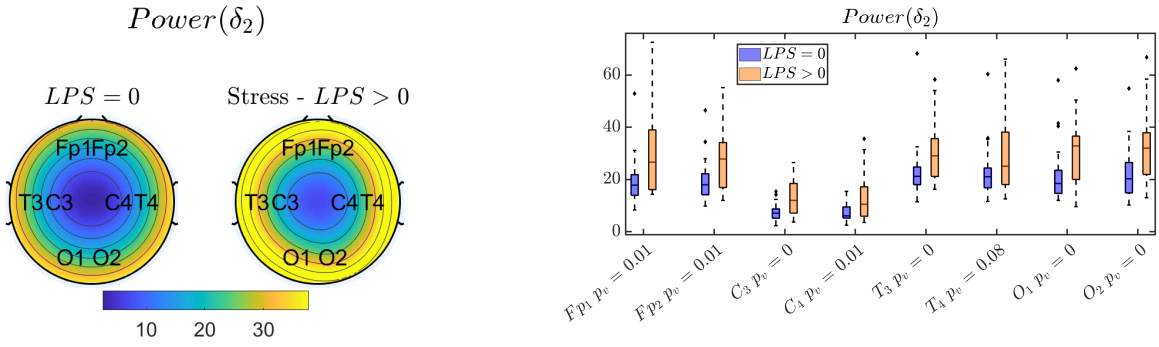

Figure 9. The δ_2 power showed a marked increase in case of stress ($LPS > 0$) during QS for patients with post-menstrual age > 36 *weeks*. The left panel displays the topoplot of the δ_2 power on the scalp, while the right one reports the boxplot of δ_2 power for each channel (p-value reported with Kruskal-Wallis test).

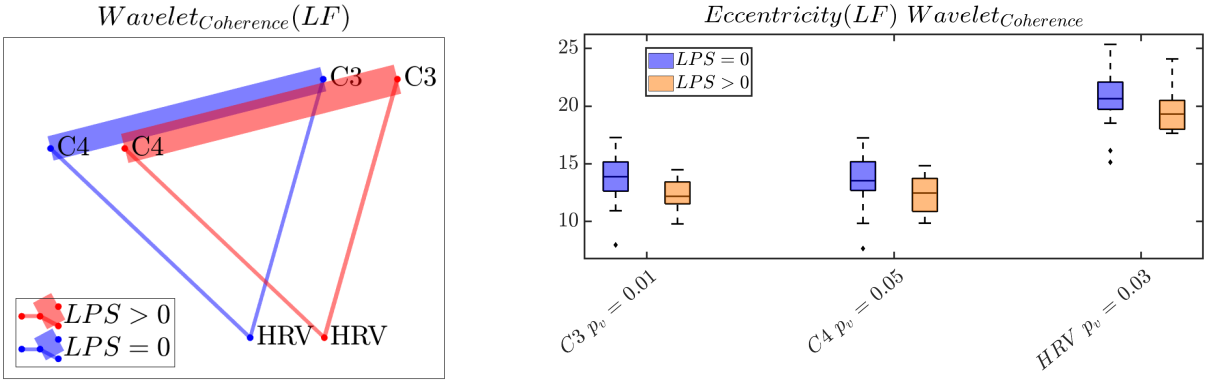

Figure 10. The Brain-heart connectivity measured via wavelet coherence in the LF band during nQS for patients with a *post-menstrual age* between 34 and 36 weeks. Stress exposure increases coupled dynamics between brain and heart. However, most of the linked dynamics is shared between EEG channels C_3 and C_4. The left panel reports the graph of brain-heart network, while the right panel displays a boxplot for its eccentricity.

regularity) and presence of transient waveforms [27]. The reasons to have a dysmature EEG might be multiple. Figure 5 shows predominance of connectivity and EEG features for the younger patients, who have EEG complexity as one of the most discriminative feature (Figure 6). This might simply be due to the fact that younger patients are the one to experience higher stress and therefore a discontinuous cortical activity might be one of the most discriminative features. However, according to Fabrizi [63], the full development of pain circuitry is completed by 35 weeks PMA, while infants at 28 weeks tend to have immature evoked potentials following painful stimuli. Under pain elicitation, Fabrizi et al. observed delta-brushes and a 10-fold increase in the δ_2 band in premature infants [63], which could explain results in Figures 6, 8, 9. Consequently, the higher discontinuity might be driven by a higher stress exposure, which is also a hypoth-

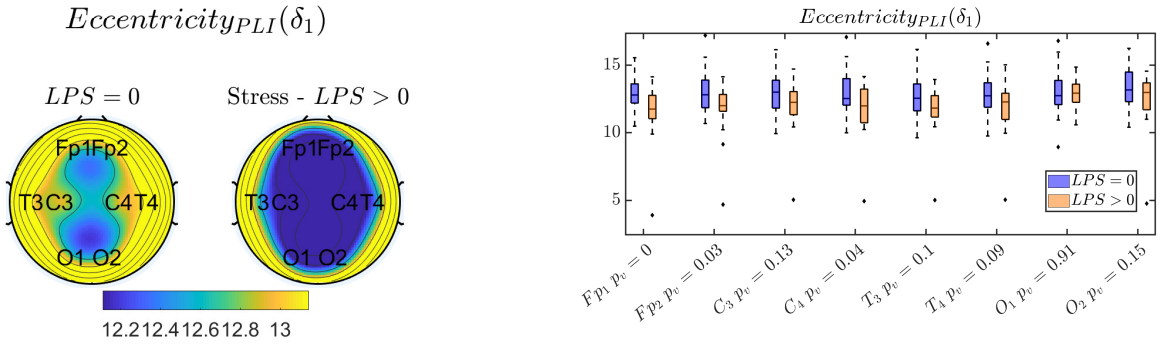

Figure 11. The synchrony in the δ_1 band shows a marked increase in case of stress during nQS for patients with *post-menstrual age* > 36 *weeks*. The left panel reports the topoplot of the eccentricity (computed with phase lag index in δ_1 band) on the scalp, while the right one reports the boxplot of the eccentricity for each channel (p-value reported with Kruskal-Wallis test). The eccentricity is a measure of distance, therefore the lower the eccentricity the higher the connectivity.

esis supported by Figure 12. Although all reported features in the paper were tested in a three-group fashion, this figure specifically reports the results related to the selected univariate EEG features, which best illustrate the relationship between stress and EEG dysmaturity. If data are grouped in three stress categories ($LPS = 0$, $LPS = [1-4]$, $LPS > 4$), EEG complexity, δ_2 power and asymmetry show a trend related with stress levels, as also shown by the Kruskal-Wallis test and the multicomparison tests with the sign (*) for $p \leq 0.05$ and the sign (**) for $p \leq 0.01$. In other words, those boxplot trends might further suggest that an "increase of dysmaturity" can also be due to early-life pain experience. Although conclusions cannot be drawn without a further study with a multiclass framework, the majority of significant multiple comparison tests reveal differences between $LPS = 0$ and $LPS = [1-4]$, which might suggest that signs of discomfort are enough to detect differences in physiological data. This might also be linked to the asymmetric distribution reported in Table 1 and it suggests that stress can be discriminated with $LPS > 0$. However, the lack of significant comparison tests between $LPS = [1-4]$ and $LPS > 4$ can be simply due to a lower number of intense pain patients ($LPS > 4$), which is also highlighted by the asymmetric distribution of Table 1. In addition, the infants cortical responses are known to be related to pain intensity and animal models have shown that power in θ band can also increase in case of noxious stimuli, which is further confirmed in the lower margin from rEEG in the θ band (figure 8) and higher connectivity in θ band (figure 7).

Concerning the greater cortical and brain-heart synchrony in case of stress load, the more intense EEG interactions can be the result of a delayed development since the functional cortical connectivity as measured by EEG decreases with maturation [49], [64]. After 32 weeks PMA, neural fibers growth is dominated by interhemispheric connections to form the core of the adult white-matter. At the same time, those fibers grow

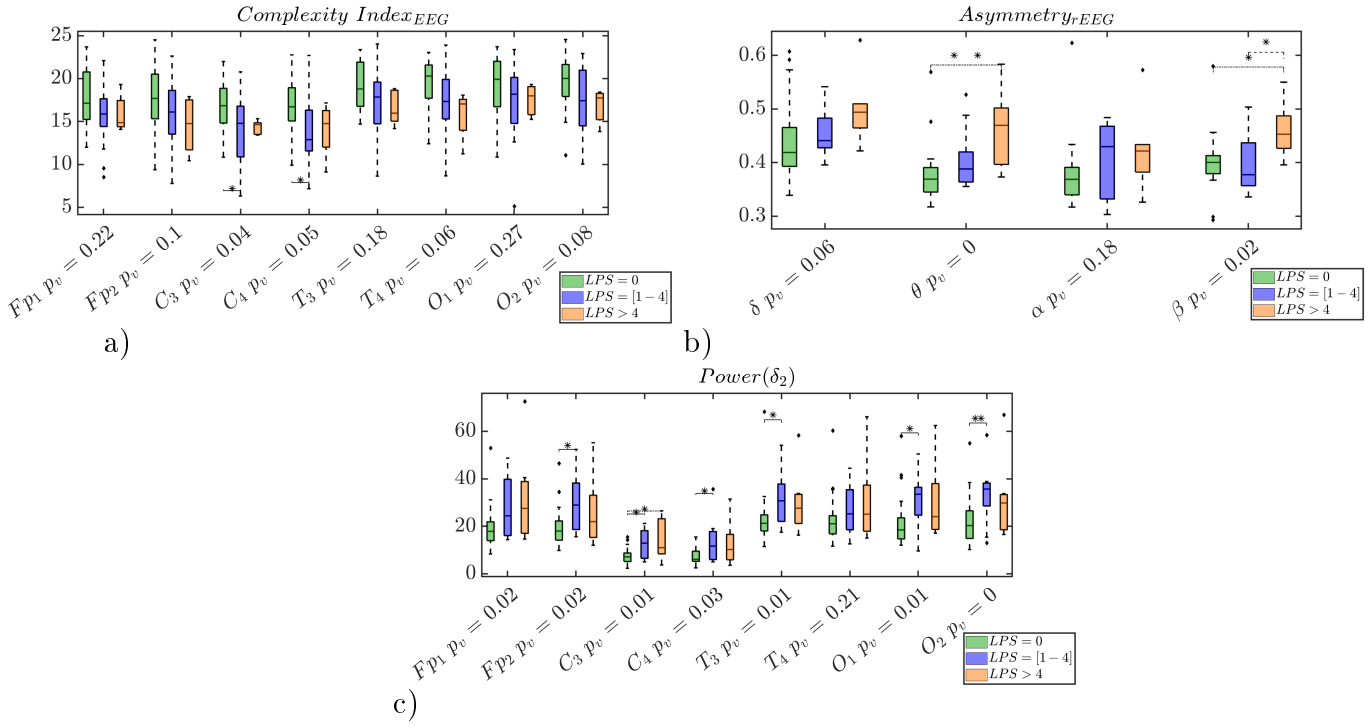

Figure 12. The figure shows the EEG complexity in nQS (panel a) for patients with post-menstrual age ≤ 32 *weeks*, the range EEG asymmetry in nQS (panel b) for patients in the range between 34 and 36 weeks and the *delta*$_2$ power during QS for patients with post-menstrual age > 36 *weeks* (panel c). Data are reported in three groups $LPS = 0$, $LPS = [1-4]$, $LPS = 4$. The p-values are obtained with a Kruskal-Wallis test and the multicomparison tests significance is expressed by (**) for $p \leq 0.01$ and (*) for $p \leq 0.05$. Similarly to the binary comparison, higher stress values are related to higher *delta*$_2$ power and *asymmetry* as well as a lower complexity. These results show how stress can increase the level of dysmaturity of EEG with a higher level of cumulated pain.

into the cortical plate together with thalamo-cortical connections, while the subcortical plate diminishes in size [64]. Those events, together with cortex gyrification, lead to the establishment of cortical specialization and they are accompanied by profound changes in EEG, which should be less discontinuous [65] and less functionally connected [64]. The persistent low-frequency rhythms and connectivity can be a sign of delayed development or emergence of altered cognitive processing [66]. Those findings seem to support the fact that stress does not only increase the power, but also the connectivity, in the low-frequency bands (as reported by figures 7, 11, and 10).

Interestingly, the autonomic activity seems less discriminative in detecting stress exposure, while figure 5 shows that the brain-heart interaction have better performance in stress classification. Stress exposure might increase the synchrony between the cardiovascular activity and the cortical activity, as reported in figure 10. In addition, Pfurtscheller has shown that burst activity might induce an HR response, which is greater in case of closer burst-activity [67]. One might speculate the stress might modu-

late cortical activity, which can modulate the cardiovascular system. Similar discussion were reported for stroke [68] and epilepsy [69], but further testing is required to prove the presence of a synchronized physiological network in case of stress.

Based on the results provided from figure 1 to 4, substantial differences between sleep states are not noticed, as also previously reported by Grunau, since EEG responses to noxious stimuli are not sleep dependent [1]. If these results are compared with the literature about pain classification [10], [12], the reported stress classifiers reached similar performances in terms of accuracy. However, several aspects should be highlighted. The pain classification literature usually does not report kappa scores, while this study suggests a moderate association between the predicted labels and the true labels (the Cohen's kappa mostly lies in the range [0.4-0.6]). Gruss et al [12] reported Cramers's V to support the accuracy rates of thermal pain classification. Although their results suggest a strong association between their features and the level of pain (the variable V mostly lies in the range [0.6-0.8]), Gruss et al. focused on evoked pain, while this study focuses on perinatal stress detection based on physiological background activity. In addition, unlike Gruss [12], an increasing pain threshold does not always lead to an increase of classification performance. In particular, a decreased Kappa is observed for a higher pain threshold in old patients (> 36 *weeks*) and higher kappa for higher pain threshold in younger patients (≤ 36 *weeks*), which can indicate either a milder experience of pain or a more difficult discrimination of stress at older age. This result is line with findings by [70], which indicates a predominant effect of early procedural pain in brain development. Therefore, PMA becomes an essential factor in stress discrimination. This is not only due to the reduced amount of experienced pain later on in the NICU stay, but also the interplay between pain and development which can shift the set of features necessary to discriminate stress. Consequently, age correction is not only required for clinical purposes, but it is fundamental since the physiological response to perinatal stress differs with patient's development [70]. However, the effect of post-menstrual age and sleep-wake cyclicity were thoroughly minimized. The analysis based on QS/nQS, the age correction and the division in age groups according to Pavlidis et al. [65] limited the influence of development on the stress classifier. It should also be mentioned that the stress definition was based on the pain score of the day before and represents the cumulated effect of the previous day. Therefore, the classification can be barely influenced by the sleep staging of the recording under investigation and the effect of cumulative pain or stress is expected to influence the physiology of the infant, as also shown in case of pain stimulation [7].

In a nutshell, the current study suggests that preterm infants that show greater stress have a more dysmature or discontinuous EEG, as reported by the lower complexity (Figure 6), the lower rEEG asymmetry (Figure 8) and higher power in δ_2 band (Figures 9 and 12). Although the decrease in EEG complexity and the better performance of the stress classifier in the youngest patients (Figures 1, 2, 3, 4) might point out that patients

under stress are most vulnerable at lower gestational age, the stress results seem not only to depend on the age of the patient, but they can vary based on the threshold used to define stress. In this perspective, the increase in δ_2 power might be related to the increase of the cortical activity due to accumulation of pain (Figure 9). Furthermore, stress exposure seems to also induce a further synchronization among the different modalities, with an increase in EEG connectivity in θ and δ_1 band (Figures 7 and 11)) and a higher brain-heart synchrony (Figure 10). Connectivity seems also to fundamentally contribute to the stress classification, as highlighted in Figure 5. Up to our knowledge, this is the first study that tries to automatically classify stress in preterm infants in an unobtrusive way, which can justify the moderate association between features and stress. Our results show that the relative contribution of the different features and modalities of the stress classification model change with the infants' age. Therefore, pain scores as well postmenstrual age are essential factors to take into account for stress discrimination in the NICU.

5. Conclusion

The current study is the first to discriminate background stress in premature infants based on physiological data. The main novelty of study lies on the absence of intrusive methods or pain elicitation protocols to investigate physiological data under stress conditions. This approach could provide insights in the relationship between neurodevelopment and stress impact in the neonatal intensive care unit. Alongside a well-established literature about EEG responses to noxious stimuli and procedural pain, the findings in this study show a possible relationship between dysmature EEG and stress. In addition, a more synchronized cortical activity as well as a tighter communication between the slow-wave cortical activity and the cardiovascular activity have been found in case of stress, while the autonomic activity does not show a clear link with perinatal stress. The current findings suggest that an automatic tool to investigate disorganized EEG can be used not only for brain development, but can also be helpful to assess the level of stress in infants at the cot-side.

6. Acknowledgements

This research is supported by Bijzonder Onderzoeksfonds KU Leuven (BOF): The effect of perinatal stress on the later outcome in preterm babies (# C24/15/036). The research leading to these results has received funding from EU H2020 MSCA-ITN-2018: 'Integrating Functional Assessment measures for Neonatal Safeguard (INFANS)', funded by the European Commission under Grant Agreement # 813483. This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information. This research is supported by Flanders AI Research Program. M.L. is a SB PhD fellow at Fonds voor Wetenschappelijk Onderzoek-Vlaanderen (FWO), supported by Flemish government.

7. References

[1] Ruth Eckstein Grunau. Neonatal pain in very preterm infants: long-term effects on brain, neurodevelopment and pain reactivity. *Rambam Maimonides medical journal*, 4(4):e0025, 2013.

[2] Jeffrey M Perlman and Joseph J Volpe. Suctioning in the Preterm Infant: Effects on Cerebral Blood Flow Velocity, Intracranial Pressure, and Arterial Blood Pressure. *PEDIATRICS*, 72(3):329–334, 1983.

[3] Manon Ranger, C. Céleste Johnston, and K. J.S. Anand. Current Controversies Regarding Pain Assessment in Neonates, oct 2007.

[4] Laura Jones, Lorenzo Fabrizi, Maria Laudiano-Dray, Kimberley Whitehead, Judith Meek, Madeleine Verriotis, and Maria Fitzgerald. Nociceptive Cortical Activity Is Dissociated from Nociceptive Behavior in Newborn Human Infants under Stress. *Current Biology*, 27(24):3846–3851.e3, dec 2017.

[5] Susanne Brummelte, Ruth E. Grunau, Vann Chau, Kenneth J. Poskitt, Rollin Brant, Jillian Vinall, Ayala Gover, Anne R. Synnes, and Steven P. Miller. Procedural pain and brain development in premature newborns. *Annals of Neurology*, 71(3):385–396, mar 2012.

[6] Xiaomei Cong, Regina M. Cusson, Stephen Walsh, Naveed Hussain, Susan M. Ludington-Hoe, and Di Zhang. Effects of Skin-to-Skin Contact on Autonomic Pain Responses in Preterm Infants. *The Journal of Pain*, 13(7):636–645, jul 2012.

[7] Rebeccah Slater, Alan Worley, Lorenzo Fabrizi, Siân Roberts, Judith Meek, Stewart Boyd, and Maria Fitzgerald. Evoked potentials generated by noxious stimulation in the human infant brain. *European Journal of Pain*, 14(3):321–326, mar 2010.

[8] Rebeccah Slater, Laura Cornelissen, Lorenzo Fabrizi, Debbie Patten, Jan Yoxen, Alan Worley, Stewart Boyd, Judith Meek, and Maria Fitzgerald. Oral sucrose as an analgesic drug for procedural pain in newborn infants: A randomised controlled trial. *The Lancet*, 376(9748):1225–1232, oct 2010.

[9] Daphna Yasova Barbeau and Michael D. Weiss. Sleep Disturbances in Newborns. *Children*, 4(10):90, oct 2017.

[10] Gaurav Misra, Wei-en Wang, Derek B. Archer, Arnab Roy, and Stephen A. Coombes. Automated classification of pain perception using high-density electroencephalography data. *Journal of Neurophysiology*, 117(2):786–795, feb 2017.

[11] Vishal Vijayakumar, Michelle Case, Sina Shirinpour, and Bin He. Quantifying and Characterizing Tonic Thermal Pain Across Subjects From EEG Data Using Random Forest Models. *IEEE Transactions on Biomedical Engineering*, 64(12):2988–2996, dec 2017.

[12] Sascha Gruss, Roi Treister, Philipp Werner, Harald C. Traue, Stephen Crawcour, Adriano Andrade, and Steffen Walter. Pain Intensity Recognition Rates via Biopotential Feature Patterns with Support Vector Machines. *PLOS ONE*, 10(10):e0140330, oct 2015.

[13] Justin E. Brown, Neil Chatterjee, Jarred Younger, and Sean Mackey. Towards a Physiology-Based Measure of Pain: Patterns of Human Brain Activity Distinguish Painful from Non-Painful Thermal Stimulation. *PLoS ONE*, 6(9):e24124, sep 2011.

[14] Marco L. Loggia, Mylène Juneau, and Catherine M. Bushnell. Autonomic responses to heat pain: Heart rate, skin conductance, and their relation to verbal ratings and stimulus intensity. *Pain*, 152(3):592–598, mar 2011.

[15] Marina de Tommaso, Gabriele Trotta, Eleonora Vecchio, Katia Ricci, Frederik Van de Steen, Anna Montemurno, Marta Lorenzo, Daniele Marinazzo, Roberto Bellotti, and Sebastiano Stramaglia. Functional connectivity of EEG signals under laser stimulation in migraine. *Frontiers in Human Neuroscience*, 9(NOV), nov 2015.

[16] Claudio Imperatori, Benedetto Farina, Maria Isabella Quintiliani, Antonio Onofri, Paola Castelli Gattinara, Marta Lepore, Valentina Gnoni, Edoardo Mazzucchi, Anna Contardi, and Giacomo Della Marca. Aberrant EEG functional connectivity and EEG power spectra in resting state post-traumatic stress disorder: A sLORETA study. *Biological Psychology*, 102(1):10–17, 2014.

[17] Karel Allegaert, Dick Tibboel, Gunnar Naulaers, Denise Tison, Annick De Jonge, Monique Van Dijk, Christine Vanhole, and Hugo Devlieger. Systematic evaluation of pain in neonates: effect on the number of intravenous analgesics prescribed. *European Journal of Clinical Pharmacology*, 59(2):87–90, jun 2003.

[18] Karel Allegaert, Gunnar Naulaers, Sophie Vanhaesebrouck, and Brian J Anderson. The paracetamol concentration-effect relation in neonates. *Pediatric Anesthesia*, 23(1):45–50, jan 2013.

[19] Jonathan Moeyersons, Matthew Amoni, Sabine Van Huffel, Rik Willems, and Carolina Varon. R-DECO: An open-source Matlab based graphical user interface for the detection and correction of R-peaks. *bioRxiv*, page 560706, feb 2019.

[20] Joseph R. Isler, Raymond I. Stark, Philip G. Grieve, Martha G. Welch, and Michael M. Myers. Integrated information in the EEG of preterm infants increases with family nurture intervention, age, and conscious state. *PLOS ONE*, 13(10):e0206237, oct 2018.

[21] Raquel Bailón, Ghailen Laouini, César Grao, Michele Orini, Pablo Laguna, and Olivier Meste. The integral pulse frequency modulation model with time-varying threshold: Application to heart rate variability analysis during exercise stress testing. *IEEE Transactions on Biomedical Engineering*, 58(3 PART 1):642–652, mar 2011.

[22] M. André, M. D. Lamblin, A. M. D'Allest, L. Curzi-Dascalova, F. Moussalli-Salefranque, S. Nguyen The Tich, M. F. Vecchierini-Blineau, F. Wallois, E. Walls-Esquivel, and P. Plouin. Électroencéphalographie du nouveau-né prématuré et à terme. Aspects maturatifs et glossaire. *Neurophysiologie Clinique*, 40(2):59–124, 2010.

[23] Anneleen Dereymaeker, Kirubin Pillay, Jan Vervisch, Maarten De Vos, Sabine Van Huffel, Katrien Jansen, and Gunnar Naulaers. Review of sleep-EEG in preterm and term neonates. *Early Human Development*, 113:87–103, oct 2017.

[24] Ofelie De Wel, Mario Lavanga, Alexander Caicedo, Katrien Jansen, Gunnar Naulaers, and Sabine Van Huffel. Decomposition of a Multiscale Entropy Tensor for Sleep Stage Identification in Preterm Infants. *Entropy*, 21(10):936, sep 2019.

[25] Amir Hossein Ansari, Ofelie De Wel, Mario Lavanga, Alexander Caicedo, Anneleen Dereymaeker, Katrien Jansen, Jan Vervisch, Maarten De Vos, Gunnar Naulaers, and Sabine Van Huffel. Quiet sleep detection in preterm infants using deep convolutional neural networks. *Journal of Neural Engineering*, 15(6), sep 2018.

[26] Kirubin Pillay, Anneleen Dereymaeker, Katrien Jansen, Gunnar Naulaers, Sabine Van Huffel, and Maarten De Vos. Automated EEG sleep staging in the term-age baby using a generative modelling approach. *Journal of Neural Engineering*, 15(3), feb 2018.

[27] Elena Pavlidis, Rhodri O. Lloyd, and Geraldine B. Boylan. EEG-A Valuable Biomarker of Brain Injury in Preterm Infants. In *Developmental Neuroscience*, volume 39, pages 23–35. S. Karger AG, jul 2017.

[28] Madeleine Verriotis, Pishan Chang, Maria Fitzgerald, and Lorenzo Fabrizi. The development of the nociceptive brain. *Neuroscience*, 338:207–219, dec 2016.

[29] Radoslav Bortel and Pavel Sovka. Approximation of statistical distribution of magnitude squared coherence estimated with segment overlapping. *Signal Processing*, 87:1100–1117, 2007.

[30] Martha G. Welch, Raymond I. Stark, Philip G. Grieve, Robert J. Ludwig, Joseph R. Isler, Joseph L. Barone, and Michael M. Myers. Family nurture intervention in preterm infants increases early development of cortical activity and independence of regional power trajectories. *Acta Paediatrica*, 106(12):1952–1960, dec 2017.

[31] Julián J. González, Soledad Mañas, Luís De Vera, Leopoldo D. Méndez, Santiago López, José M. Garrido, and Ernesto Pereda. Assessment of electroencephalographic functional connectivity in term and preterm neonates. *Clinical Neurophysiology*, 122(4):696–702, 2011.

[32] Maya David, Michael Hirsch, Jacob Karin, Eran Toledo, and Solange Akselrod. An estimate of fetal autonomic state by time-frequency analysis of fetal heart rate variability. *Journal of Applied Physiology*, 102(3), 2007.

[33] AJ Camm, M Malik, JT Bigger, and B Günter. Task force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation and clinical use. *Circulation*, 93(5):1043–1065, 1996.

[34] Muriel Doret, Jiří Spilka, Václav Chudáček, Paulo Gonçalves, Patrice Abry, and HE Stanley. Fractal Analysis and Hurst Parameter for Intrapartum Fetal Heart Rate Variability Analysis: A Versatile Alternative to Frequency Bands and LF/HF Ratio. *PLOS ONE*, 10(8):e0136661, aug 2015.

[35] P Abry, H Wendt, S Jaffard, H Helgason, P Goncalves, E Pereira, C Gharib, P Gaucherand, and M Doret. Methodology for multifractal analysis of heart rate variability: From LF/HF ratio to wavelet leaders. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 106–109. IEEE, aug 2010.

[36] Herwig Wendt, Patrice Abry, and Stéphane Jaffard. Bootstrap for Multifractal Analysis. *IEEE Signal Processing Magazine*, 24(4):38–48, 2007.

[37] Ofelie De Wel, Mario Lavanga, Alexander Dorado, Katrien Jansen, Anneleen Dereymaeker, Gunnar Naulaers, and Sabine Van Huffel. Complexity Analysis of Neonatal EEG Using Multiscale Entropy: Applications in Brain Maturation and Sleep Stage Classification. *Entropy*, 19(10):516, sep 2017.

[38] D. Popivanov, V. Stomonyakov, Z. Minchev, S. Jivkova, P. Dojnov, S. Jivkov, E. Christova, and S. Kosev. Multifractality of decomposed EEG during imaginary and real visual-motor tracking. *Biological Cybernetics*, 94(2):149–156, 2006.

[39] Olga E Dick. Multifractal analysis of the psychorelaxation efficiency for the healthy and pathological human brain. *Chaotic Model Simul.*, 1(December 2011):219–227, 2012.

[40] Douglas E Lake and J Randall Moorman. Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices. *American journal of physiology. Heart and circulatory physiology*, 300(1):H319–H325, 2011.

[41] Madalena Costa, Ary L. Goldberger, and C.-K K. Peng. Multiscale Entropy Analysis of Complex Physiologic Time Series. *Physical Review Letters*, 89(068102), 2002.

[42] Douglas E. Lake, Joshua S. Richman, M. Pamela Griffin, and J. Randall Moorman. Sample entropy analysis of neonatal heart rate variability. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 283(3):R789–R797, sep 2002.

[43] Duan Li, Xiaoli Li, Zhenhu Liang, Logan J. Voss, and Jamie W. Sleigh. Multiscale permutation entropy analysis of EEG recordings during sevoflurane anesthesia. *Journal of Neural Engineering*, 7(4), aug 2010.

[44] Dirk Hoyer, Florian Tetschke, Susan Jaekel, Samuel Nowack, Otto W. Witte, Ekkehard Schleußner, and Uwe Schneider. Fetal Functional Brain Age Assessed from Universal Developmental Indices Obtained from Neuro-Vegetative Activity Patterns. *PLoS ONE*, 8(9):1–8, 2013.

[45] John M. O' Toole, Geraldine B. Boylan, John M O' Toole, and Geraldine B. Boylan. NEURAL: quantitative features for newborn EEG using Matlab. *arXiv preprint arXiv:1704.05694*, apr 2017.

[46] D Piper, K Schiecke, B Pester, F Benninger, M Feucht, and H Witte. Time-variant coherence between heart rate variability and EEG activity in epileptic patients: an advanced coupling analysis between physiological networks. *New Journal of Physics*, 16(11):115012, nov 2014.

[47] Michael G. Rosenblum, Laura Cimponeriu, Anastasios Bezerianos, Andreas Patzak, and Ralf Mrowka. Identification of coupling direction: application to cardiorespiratory interaction. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 65(4):11, 2002.

[48] Guido Nolte, Ou Bai, Lewis Wheaton, Zoltan Mari, Sherry Vorbach, and Mark Hallett. Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clinical Neurophysiology*, 115(10):2292–2307, oct 2004.

[49] Mario Lavanga, Ofelie De Wel, A. Caicedo, Katrien Jansen, Anneleen Dereymaeker, Gunnar

Naulaers, and Sabine Van Huffel. A brain-age model for preterm infants based on functional connectivity. *Physiological Measurement*, 39(4):044006, apr 2018.

[50] Cornelis J. Stam, Guido Nolte, and Andreas Daffertshofer. Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Human Brain Mapping*, 28(11):1178–1193, nov 2007.

[51] L Faes, D Marinazzo, F Jurysta, and G Nollo. Linear and non-linear brain–heart and brain–brain interactions during sleep. *Physiol. Meas*, 36:683, 2015.

[52] Gert Pfurtscheller, Andreas R. Schwerdtfeger, Annemarie Seither-Preisler, Clemens Brunner, Christoph Stefan Aigner, Joana Brito, Marciano P. Carmo, and Alexandre Andrade. Brain–heart communication: Evidence for "central pacemaker" oscillations with a dominant frequency at 0.1 Hz in the cingulum. *Clinical Neurophysiology*, 128(1):183–193, jan 2017.

[53] F. Wallois. Synopsis de la maturation des activités cérébrales chez le nouveau-né prématuré. *Neurophysiologie Clinique*, 40(2):125–126, 2010.

[54] D. Iatsenko, P. V.E. McClintock, and A. Stefanovska. Extraction of instantaneous frequencies from ridges in time-frequency representations of signals. *Signal Processing*, 125:290–303, aug 2016.

[55] Gemma Lancaster, Dmytro Iatsenko, Aleksandra Pidde, Valentina Ticcinelli, and Aneta Stefanovska. Surrogate data for hypothesis testing of physical systems, jul 2018.

[56] Philip Clemson, Gemma Lancaster, and Aneta Stefanovska. Reconstructing Time-Dependent Dynamics. *Proceedings of the IEEE*, 104(2):223–241, feb 2016.

[57] Björn Kralemann, Arkady Pikovsky, and Michael Rosenblum. Reconstructing effective phase connectivity of oscillator networks from observations. *New Journal of Physics*, 16(8):085013, aug 2014.

[58] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, mar 2009.

[59] JA Bondy and USR Murty. *Graph theory with applications*. Springer Berlin Heidelberg, 1976.

[60] Tin Kam Ho. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844, 1998.

[61] Yi Wei Chen and Chih Jen Lin. Combining SVMs with various feature selection strategies. *Studies in Fuzziness and Soft Computing*, 207:315–324, 2006.

[62] Ivica Kostović and Miloš Judaš. The development of the subplate and thalamocortical connections in the human foetal brain. *Acta Paediatrica, International Journal of Paediatrics*, 99(8):1119–1127, mar 2010.

[63] Lorenzo Fabrizi, Madeleine Verriotis, Gemma Williams, Amy Lee, Judith Meek, Sofia Olhede, and Maria Fitzgerald. Encoding of mechanical nociception differs in the adult and infant brain. *Scientific Reports*, 6, jun 2016.

[64] M. M. Myers, P. G. Grieve, R. I. Stark, J. R. Isler, M. A. Hofer, J. Yang, R. J. Ludwig, and M. G. Welch. Family Nurture Intervention in preterm infants alters frontal cortical functional connectivity assessed by EEG coherence. *Acta Paediatrica, International Journal of Paediatrics*, 104(7):670–677, jul 2015.

[65] Elena Pavlidis, Rhodri O. Lloyd, Sean Mathieson, and Geraldine B. Boylan. A review of important electroencephalogram features for the assessment of brain maturation in premature infants, sep 2017.

[66] Gennady G Knyazev. EEG delta oscillations as a correlate of basic homeostatic and motivational processes. *Neuroscience and Biobehavioral Reviews*, 36:677–695, 2011.

[67] Klaus Pfurtscheller, Günther Bauernfeind, Gernot R. Müller-Putz, Berndt Urlesberger, Wilhelm Müller, and Gert Pfurtscheller. Correlation between EEG burst-to-burst intervals and HR acceleration in preterm infants. *Neuroscience Letters*, 437(2):103–106, may 2008.

[68] Zhili Chen, Poornima Venkat, Don Seyfried, Michael Chopp, Tao Yan, and Jieli Chen. Brain–Heart Interaction. *Circulation Research*, 121(4):451–468, aug 2017.

[69] Karin Schiecke, Britta Pester, Diana Piper, Franz Benninger, Martha Feucht, Lutz Leistritz, and

Herbert Witte. Nonlinear Directed Interactions Between HRV and EEG Activity in Children With TLE. *IEEE Transactions on Biomedical Engineering*, 63(12):2497–2504, dec 2016.

[70] Emma G. Duerden, Ruth E. Grunau, Ting Guo, Justin Foong, Alexander Pearson, Stephanie Au-Young, Raphael Lavoie, M. Mallar Chakravarty, Vann Chau, Anne Synnes, and Steven P. Miller. Early procedural pain is associated with regionally-specific alterations in thalamic development in preterm neonates. *Journal of Neuroscience*, 38(4):878–886, jan 2018.