Evaluation of maturation in preterm infants through an ensemble machine learning algorithm using physiological signals

C.S. Leon, S. Cabon, H. Patural, G. Gascoin, Cyril Flamant, J. Roue, G. Favrais, A. Beuchee, P. Pladys, Guy Carrault

Evaluation of maturation in preterm infants through an ensemble machine learning algorithm using physiological signals

Cristhyne León, Sandie Cabon, Hugues Patural, Géraldine Gascoin, Cyril Flamant, Jean-Michel Roué, Géraldine Favrais, Alain Beuchée, Patrick Pladys, and Guy Carrault

Abstract—This study was designed to test if heart rate variability (HRV) data from preterm and full-term infants could be used to estimate their functional maturational age (FMA), using a machine learning model. We propose that the FMA, and its deviation from the postmenstrual age (PMA) of the infants could inform physicians about the progress of the maturation of the infants. The HRV data was acquired from 50 healthy infants, born between 25 and 41 weeks of gestational age, who did not present any signs of abnormal maturation relative to their age group during the period of observation. The HRV features were used as input for a machine learning model that uses filtering and genetic algorithms for feature selection, and an ensemble machine learning (EML) algorithm, which combines linear and random forest regressions, to produce as output a FMA. Using HRV data, the FMA had a mean absolute error of 0.93 weeks, 95% CI [0.78, 1.08], compared to the PMA. These results demonstrate that HRV features of newborn infants can be used by an EML model to estimate their FMA. This method was also generalized using respiration rate variability (RRV) and bradycardia data, obtaining similar results. The FMA, predicted either by HRV, RRV or bradycardia, and its deviation from the true PMA of the infants, could be used as a surrogate measure of the maturational age of the infants, which could potentially be monitored non-invasively and in real-time in the setting of neonatal intensive care units.

Index Terms—Ensemble machine learning, genetic algorithm, premature infants, heart rate variability, respiration rate, bradycardia.

I. INTRODUCTION

Heart rate variability (HRV), defined as the variation of the duration of the interval between consecutive heartbeats over a period of time, is a useful and non-invasive tool for evaluating the status of the autonomic nervous system (ANS) [1]. In neonates and infants, HRV measurements could be used to evaluate the process of maturation of their ANS ([2], [3]). HRV might also have diagnostic value for different clinical situations in neonates and infants which are linked to either congenital or acquired autonomic dysregulation. For instance, sudden infant death syndrome (SIDS) [4], late-onset sepsis ([5], [6]), neonatal seizures [7], and hypoxic-ischemic encephalopathy [8] have been associated with altered HRV.

C. León, S. Cabon, A. Beuchée, P. Pladys, and G. Carrault are with Univ Rennes, Inserm, LTSI - UMR 1099, F-35000 Rennes, France.
H. Patural is with Univ Saint-Etienne, CHU Saint-Etienne, EA SNA-EPIS, F-42000 Saint-Etienne, France.
G. Gascoin is with CHU d'Angers, F-49000 Angers, France.
C. Flamant is with CHU Nantes, F-44000 Nantes, France.
J. -M. Roué is with Univ Brest, CHU Brest, LIEN, F-29000 Brest, France.
G. Favrais is with Univ Tours, CHRU Tours, F-37000 Tours, France.

Infants born prematurely (before 37 weeks of gestational age) represent between 5% to 18% of live births world wide [9], and they are more susceptible to some of the aforementioned conditions when compared to term infants ([10]–[16]). Therefore, premature infants constitute a population of particular interest for the evaluation of their HRV as it is linked to the maturation of their ANS.

However, healthy preterm neonates present an altered HRV compared to that of infants born at term [17]. This translates into significant differences in time-domain, frequency-domain, and non-linear HRV measurements [18]. Furthermore, significant differences have been found in the HRV of premature infants with different degrees of prematurity ([19], [20]), with HRV measurements approaching normal values with increased gestational age (GA) [21]. Although the HRV parameters show an improvement with chronological age ([20], [22]), these differences in HRV continue to prevail even when the postmesntrual age (PMA) of the preterm infants reaches term equivalent age ([23], [24]), and for months afterwards [25]. While normative data for the HRV of full-term neonates has been proposed ([2], [3]), this is not the case for preterm newborns.

All these factors pose an obstacle for evaluating the maturation of preterm infants based on their HRV. Moreover, using the HRV of preterm infants as a potentially diagnostic tool for clinical conditions associated to autonomic dysregulation presents a greater difficulty, as even the HRV of healthy preterm infants at theoretical term will seem abnormal if compared to that of healthy neonates born at full-term [26].

Previous studies have reported successfully using machine learning algorithms for the estimation of the maturity of preterm infants. One study used functional magnetic resonance imaging of preterm infants at term equivalent age for predicting the GA of the infants, which the authors proposed as a surrogate measure of the brain maturity [27]. Another study used features derived from the electroencephalogram of preterm infants to predict the PMA, which the authors proposed as a surrogate measure for the brain maturation of the infants [28]. However, while the link between HRV, GA, and brain maturity (particularly of the ANS) has been largely reported in the literature, and the use of HRV and machine learning has been suggested to predict the prognosis of infants in the perinatal period [29], we did not find any previous study that used HRV and GA in combination with machine learning techniques, to evaluate the maturation process of infants during

the period after birth.

Therefore, in this study we propose a method based on the use of machine learning, with HRV measurements and GA as inputs, and a functional maturational age (FMA) as output. The FMA thus predicted, and its deviation from the PMA, which is measured clinically, could potentially help physicians evaluate the maturation of premature infants throughout their stay in the neonatal intensive care unit (NICU). This could also aid to the early detection of abnormalities in the neurological development of the infants as they manifest in the HRV.

This study is framed in the Digi-NewB project [30], which has as one of its objectives improving the evaluation of the maturation process in preterm infants. Digi-NewB has collected not only HRV data from the neonates in its cohort, but also respiration, movement, bradycardia, cry, and sleep data. Therefore, and although the method proposed in this study was initially designed and developed for estimation of the maturation based on HRV features, we wanted the approach and resulting algorithm to be general enough that it could be used for FMA estimation using other types of data available in the project. Thus, we also tested the proposed method on respiration rate variability (RRV) and bradycardia features, which were available in the project at the time of this study. Similarly to HRV, RRV and bradycardia also show different patterns with increasing GA and PMA of the infants ([31]–[33]), as well with certain pathologies in neonates ([34]–[36]).

II. MATERIALS AND METHODS

A. Population

The data used in this study is part of the database of the Digi-NewB cohort (NCT02863978, EU GA n°689260). The cohort prospectively included infants born between 25 and 42 weeks of gestation, hospitalized in the NICU of six university hospitals in western France (University Hospitals of Rennes, Angers, Nantes, Brest, Poitiers, and Tours) in 2017-2019. The collection of data was carried out after approval by the ethics committee (CPP Ouest 6-598) and informed parental consent.

For this study we used a population of preterm and full term infants who did not show signs of abnormal maturation, relative to their GA group, during the observation period. The selection of the population was done based on a three-step clinical evaluation performed by two senior neonatologists. First, they selected the newborns from the Digi-NewB cohort who did not present any of the following exclusion criteria: chest compression for resuscitation at birth; severe neurological lesions (grade 3 or 4 intraventricular haemorrhage, white matter lesions, hypoxic-ischemic encephalopathy); early onset
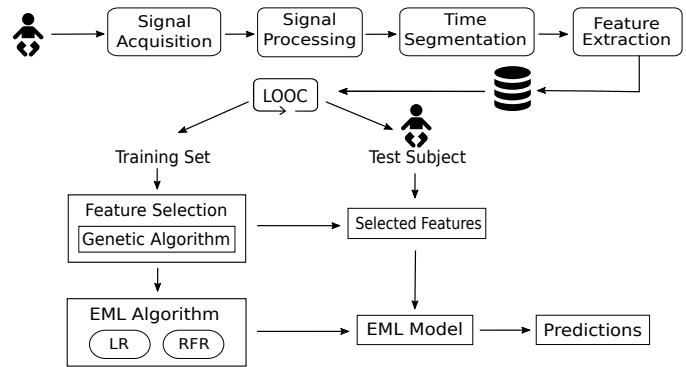

Fig. 1: Proposed approach.

sepsis; late onset sepsis; enterocolitis; severe malformations; and preterm infants with a birth weight lower than the 10th percentile for their GA. Second, both nenonatologists verified that, based on the clinical health reports, the pre-selected infants presented trajectories during the entire period of observation that could be considered normal for their GA. Third, both neonatologists met to verify and share their evaluation, and in case of doubt or disagreement in one case, that infant was excluded from the population.

The population was split into five categories, according to the GA of the infant: extreme preterm (EP), very preterm (VP), late preterm (LP), early term (ET), and full term (FT). The cut-off GA for each group, as well as the number n of infants in each group is detailed in Table I. The definition of GA, PMA, and chronological age that we use throughout this article is that proposed by the American Academy of Pediatrics [37].

The HRV, RRV, and bradycardia population were built from the same base population of 50 infants, and the differences in the number of infants between them is due to the exclusion of neonates which did not have enough recordings of good quality in the case of the RRV population; or, in the case of the bradycardia population, because their data had not been processed yet for the detection of bradycardia episodes and the extraction of the associated features.

For preterm infants, the data was collected from continuous monitoring, during 24 hours a day, for the first three weeks of life. Afterwards, the monitoring was done also for 24 continuous hours, but every ten days. In the case of ET and FT infants, the data was collected only for 24 continuous hours after the third day of life, to avoid the phase of early adaptation to extrauterine life.

B. Proposed Approach

The general approach we propose is described in Figure 1. In general terms, we acquired and processed the raw signals from each patient, in this case the electrocardiogram (ECG) which was used for HRV, RRV, and bradycardia detection. The data is then segmented into shorter periods of time, which length might vary according to the data type. From these periods the features used to describe the HRV, RRV, or bradycardia are extracted.

Given the limited number of infants in our database, we used a subject based leave-one-out cross-validation (LOOC)

Group	GA (weeks)	HRV population (n = 50)	RRV population (n = 48)	Bradycardia population (n = 43)
EP	[24, 28[	9	9	9
VP	[28, 32[	14	14	14
LP	[32, 37[	12	12	11
ET	[37, 39[	6	4	3
FT	≥ 39	9	9	6

TABLE I: Population characteristics

technique. Using the LOOC technique can help minimize over-fitting of the machine learning algorithm, and the literature suggests it generally introduces less, or at most equal, variance and bias as using a K-fold technique [38]. Thus, all the subsequent steps of our algorithm iterate over the entire population, in each iteration one patient is used as the test set (the patient left out), and the rest of the patients are used as the training set.

Next we used a genetic algorithm for feature selection and an ensemble machine learning (EML) algorithm [39] for the estimation of the PMA, which serves as the functional maturational age (FMA). The EML algorithm we propose combines the output of linear regression (LR) and random forest regression (RFR).

The following sections will explain in greater detail the different steps of our proposed approach. For signal processing and feature extraction we will detail only the HRV data, as this data type was the main focus of our study. Details concerning the signal processing to obtain the RRV data were proposed by Navarro et al. [40], and further information about the RRV features used in this study are given in the appendix B. The signal processing for bradycardia detection was described by Altuve et al. [41], and information about the features used for this study are given in appendix A. However, the subsequent steps regarding feature selection and the EML model are general and applicable to the different types of data.

C. Signal Processing and Extraction of the HRV Features

The ECGs were obtained with a sampling rate of 500Hz. R-peak detection was done with a modified version of the Pan and Tompkins algorithm, with filter coefficients adapted for neonates, as proposed in [42]. Afterwards, we extracted the R-R interval (RRI) time series, and then segmented it into 30 minutes periods, each of which was time-stamped so it could be associated to the corresponding PMA of the infant.

For each infant in the database, we selected only the available data corresponding to the first 12 weeks after birth. From each of the 30 minutes segments within the selected data we extracted the HRV features, which are categorized in four different types: time-domain ([43]–[45]), frequency-domain [43], non-linear measurements [43], and visibility graph indexes ([46], [47]). The later are derived from visibility graphs (VG) [48] and the horizontal visibility graphs (HVG) [49] computed from the time series. These HRV features, along with the GA, compose our entire feature set and are shown in Table II. We also retain the time-stamp to be able to calculate the true PMA of the infant associated to the features from each segment. Further detail about the HRV features used in this study and how they were extracted can be found in [6].

D. Data Analysis and Feature Selection

As a first step of our method for automatic feature selection, we calculated the Spearman correlation between all variables and the PMA in the training set, and eliminated all the features with a very weak absolute correlation to the PMA, ($|\rho_{xy}|$ < 0.1). We did not eliminate any features based on their correlation to each other.

Next, we standardized the remaining features, as it is recommended to do when using linear regression models. Then we used a genetic algorithm [50] on the standardized features to find the optimal combination for our EML algorithm. The genetic algorithm was configured as follows:

a) Population: We used a population of 20 chromosomes. In a genetic algorithm context, a chromosome encodes the information to be optimized. For this study, each chromosome had a length equal to the number of features, so each gene corresponded to one feature. The genes were binary, indicating if the corresponding feature was to be included or not in the final feature set. The first generation was initialized randomly, and subsequent generations were generated through a process of crossover and mutations.

b) Cost function: To evaluate the cost (or fitness) of each chromosome, we used the mean absolute error (MAE) between the estimated FMA and the true PMA as cost function. For this, we trained the machine learning algorithm for which we were trying to optimize the features (either LR or RFR), on the feature set corresponding to each chromosome and calculated the resulting MAE of the FMA on the test set. The MAE was then used as the measure of cost, with a lower MAE indicating a lower cost. The algorithm stores the chromosome with the lowest cost as well as its associated cost. When a chromosome with an even lower cost is found, this information is updated.

c) Construction of new generations: After the cost function was calculated for every chromosome of a given generation, a new generation would be constructed by ordering the chromosomes of the last generation from lowest to highest cost, and taking the top 30% (that is, the 30% best performing chromosomes) without changes and passing them on to the next generation. The remaining 70% of the new generation would be constructed by taking the top 50% of the last generation and designating them as the parent population; this population then undergoes a crossover and mutation process to generate their offspring which will pass on to the next generation.

d) Crossover: The algorithm selects two chromosomes from the parent population at random, and performs a crossover operation, in which a new chromosome is created by taking half the genes from one parent, and the remaining half from the other.

e) Mutation: The offspring chromosome resulting from a crossover operation goes through a mutation process, in which 10% of its genes are randomly chosen to change their value.

f) Stopping criteria: The genetic algorithm would stop after reaching a maximum 150 generations or if the minimum cost remained constant for 30 continuous generations.

As the genetic algorithm is applied using the LOOC split of the population, it will result in one optimized set of features for each infant. Thus, to get one unique feature set to use in the EML algorithm, we retain only the features that appear in 50% or more of the optimized feature sets.

Since the ensemble model we propose uses both LR and RFR, we implemented two instances of the genetic algorithm: one with the target to minimize the MAE of the FMA (in weeks) obtained by a LR model, and one with the target to minimize the MAE of FMA estimated by the RFR model.

Category	Feature	Abbreviation
Time-Domain	Mean duration of the RRIs	meanRR
	Standard deviation of the RRIs	sdRR
	Root mean square of the RRIs	RMSSD
	Maximum RRI	maxRR
	Minimum RRI	minRR
	Skewness of the RRI time series	Skewness
	Kurtosis of the RRI time series	Kurtosis
	Acceleration of the heart rate	AC
	Deceleration of the heart rate	DC
Frequency-Domain	Low frequency power (0.02-0.2Hz)	LF
	High frequency power (0.2-2Hz)	HF
	LF in normalized units	LFnu
	HF in normalized units	HFnu
	The ration between LF and HF	LF_HF
Non-linear Measurements	Sample entropy	SampEn
	Approximate entropy	ApEn
	Short-range fractal correlation of the time series	$\alpha1$
	Long-range fractal correlation of the time series	$\alpha2$
	Short term variability derived from the Poincaré plot	SD1
	Long term variability derived from the Poincaré plot	SD2
Visibility Graph Indexes	Mean degree of the nodes in the VG	MD_VG
	Cluster coefficient of the VG	C_VG
	Transitivity of the VG	Tr_VG
	Assortativity of VG	r_VG
	Mean degree of the nodes in the HVG	MD_HVG
	Cluster coefficient derived from the HVG	C_HVG
	Transitivity of the HVG	Tr_HVG
	Assortativity of the HVG	r_HVG
Non-HRV related	Gestational age	GA

TABLE II: Category, description, and abbreviation of all features included in the HRV feature set

Through this technique we obtained two new sets of features: one with the optimal features for LR and one for RFR, both optimized to minimize the MAE.

E. Ensemble Machine Learning

Based on preliminary inspections of the behaviour of the HRV features, we observed that some features displayed a linear relation to the PMA, while others seemed to have a non-linear behaviour. During the early stages of the study we also observed that LR and RFR favored different features. Therefore, we suggest the use of an EML model that combines LR and RFR. This allows the exploitation of both linear and non-linear correlations between the features derived from the physiological data and the PMA of the infants.

To build the EML model we first used the features selected as optimal for LR, by the method explained in the previous section, to train the LR part of the model. The FMA estimated by this model, $\widehat{FMA}_{LR}$, is then added as an additional feature to the set of features selected as optimal for the RFR by the method described in section II-D. Then the RFR is trained using this modified feature set as input, thus completing the training of EML model. Then, predictions are made on the test patient. When there are multiple observations (equivalent to 30-minute segments in the case of the HRV data) for the same day, the median value of all the FMAs corresponding to the same day and, therefore, to the same true PMA, is calculated and given as the predicted FMA for that day. Thus, the algorithm always gives only one FMA per day, regardless of the amount of available observations.

The same process is repeated for every iteration of the LOOC method, until every infant in the population has been used, at some point, as test patient.

F. Evaluation Method

To evaluate the accuracy of the model we compare the estimated FMAs with the PMAs of each infant in terms of the mean absolute error (MAE) in weeks, which is given by:

$$\text{MAE} = \frac{\sum_{n=1}^{N} |FMA_n - PMA_n|}{N} \tag{1}$$

where FMA_n is the age estimated by the model for a certain observation n, while PMA_n is the true PMA corresponding to the same observation. N is the total number of observations or predictions for each infant.

Given that we used a LOOC technique to train and test the model, so each infant in the population was at some point the test patient, we were able to calculate the MAE for each infant. Therefore, we then calculated the mean MAE, its 95% confidence interval (CI), the range of the MAE, and standard deviation (SD) over the entire population.

G. Generalization of the proposed method

In order to make the method described applicable to different types of data available in the project in which this study is framed, we had to make several considerations.

a) Handling missing values: While data types such as HRV and RRV do not have any missing values in any time period for which recordings are available, this might not be the case for every type of data. For instance, in bradycardia data there might be missing values which correspond to features that would usually describe a type of bradycardia episode which did not occur during a period of observation. Thus, it was necessary to include a method that would allow to handle missing values.

The method incorporates different options for how to handle these cases. The user can choose to drop the features for which more than a certain percentage of observations are missing, with the threshold also being adjustable by the user. To impute values on the missing data on the remaining features, the user can choose between using the mean, the median, or the most frequent value for that feature.

As a result, every feature F which has missing values, will be replaced by the feature F', which is constructed as follows:

$$F' = \begin{cases} f & \text{if the value was not missing in F} \\ f_i & \text{otherwise} \end{cases} \tag{2}$$

where f is the value in the original feature F, and f_i is the value calculated by the chosen imputation method.

Whichever imputation method is chosen, for every feature that has at least one missing value a new associated feature, *F_wasMissing*, is added to the dataset. This new feature is constructed as follow:

$$F\_wasMissing = \begin{cases} 1 & \text{if the value was missing in F} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

This feature is added because the information of whether a value was missing or not might be relevant [51]. This is the case for bradycardia, where the number of bradycardia episodes are expected to decline with increased PMA, so more missing values might be related to higher PMA.

b) Feature filtering by Spearman Correlation: In this step we excluded the features with weak Spearman correlation to the PMA, before continuing to the genetic algorithm for feature selection. For this, we set the threshold at 0.1, so that features with an absolute correlation value under this threshold ($|\rho_{xy}| < 0.1$) would be eliminated from the feature set. We set such a low threshold because a combination of features that are weakly correlated to the target variable could contribute to the final prediction.

While datasets with many features, and many of which show a strong correlation with the PMA, might benefit from this filtering step, datasets with less features, or which have a weak correlation to the PMA, might not benefit from this. Thus, we have made this step optional.

c) Handling Categorical Features: While HRV and RRV data only have numerical features, other data types might contain categorical features, or these categorical features might result from the missing data imputation process. Thus, we included the necessary provisions for the method to be able to handle both numerical and categorical features. Under these
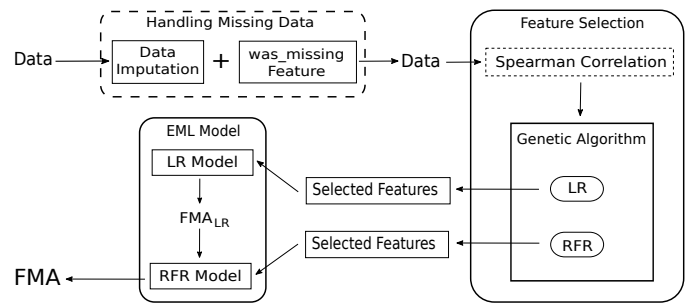

Fig. 2: Overview of the generic tool proposed. Optional phases and steps are represented framed by dashed lines. Phases and steps that are applied regardless of the data type are represented framed by solid lines.

provisions, categorical features will be transformed into binary variables using one-hot encoding.

The resulting generic tool is represented in Figure 2 and can be viewed as a series of blocks or phases. We represent phases or steps within each phase which are optional by framing them in dashed boxes; the phases or steps that are general, regardless of the data type used as input, are framed by solid lines. As observed in the figure, the method we propose takes the data, regardless of type, as input, and produces an estimation of the FMA as output. The first phase of the algorithm handles missing data by the method previously explained. This phase is optional as it applies only to data with missing values. The next phase is feature selection. While this phase is common to all data types, the step concerning filtering out the features with weak Spearman correlation to the target variables, as it was previously explained, is optional. The third phase is the EML model itself, where the FMA estimated by the LR model (FMA_{LR}) is used as an additional input feature by the RFR, thus producing the final estimated FMA.

All the data analysis, feature selection, EML algorithm, and evaluation process were developed in a Python environment.

III. RESULTS

In this section we first focus on the results related to the HRV data, presenting the results of the feature selection method and the EML model for this data. Finally, we present the results obtained when applying the same method to RRV and bradycardia data.

A. Selected HRV Features

The first step of the feature selection process was filtering out the features with a very weak Spearman correlation to the PMA ($|\rho_{xy}| < 0.1$). Figure 3 shows the correlations between all the HRV features, plus the gestational age, and the target variable (PMA). Based on this criteria, at this stage the only HRV features to be eliminated from the feature set were *meanRR*, *maxRR*, and *Tr_VG*.

The next step was using two instances of the genetic algorithm, each to optimize the feature set for training a LR model and a RFR model, respectively. This was done using the LOOC method, so we only retained for the final model
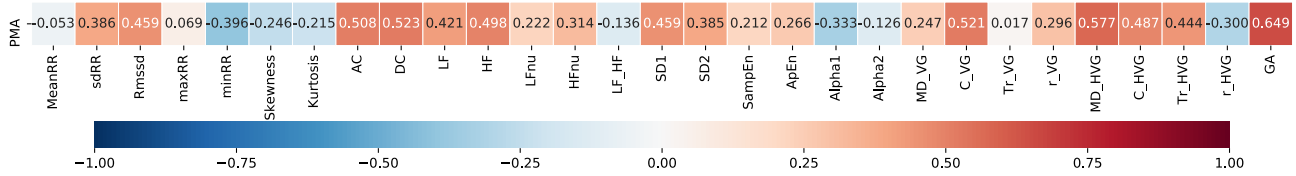
Fig. 3: Spearman correlation between HRV features and the target variable (PMA).
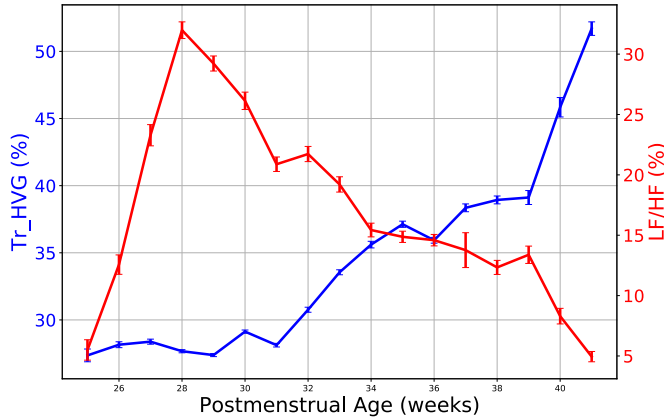

Fig. 4: Weekly average of the Tr_HVG (blue) and the LF_HF (red) features for the entire HRV population. The error bars represent the standard error of the mean (SEM).

the features that were present in at least 50% of the LOOC iterations. The features obtained by this technique are listed in Table III, and in parenthesis we present the percentage of the LOOC iterations for which the feature was chosen. We observe that the LR model favors mostly the time-domain features and visibility graph indexes derived from the HRV, while the RFR model relies mostly on frequency-domain features. From the non-linear measurements, both models use only one feature (ApEn), and also both models use the GA, which is the only feature in the set which is not derived from the HRV.

The use of different features by the two different models indicates that some HRV characteristics might have a linear

correlation to the PMA (mostly time-domain features and visibility graph indexes), while others (mostly the frequency-domain features) might have a non-linear correlation to the PMA. This is exemplified in Figure 4, were we show the weekly average of two features for the entire HRV population. One of the features selected by the genetic algorithm exclusively for the LR model was *Tr_HVG* (shown in blue), for which we observe that its behaviour is, in general, mono-tonically related to the PMA. While in red we display the behaviour of the *LF_HF* feature, one of the features which was selected by the genetic algorithm exclusively for the RFR model and which does not show a linear relation to the PMA.

B. Performance of the EML model on HRV data

We evaluated the accuracy of the EML model to estimate the functional maturational age (FMA) on the test patient in every iteration of the LOOC process. To this end we calculated the MAE, in weeks, of the FMA estimated by the model in relation to the PMA, for every patient. From this we obtained the mean MAE, maximum MAE (Max. MAE), minimum MAE (Min. MAE), the standard deviation (SD), and 95% confidence interval (95% CI) for the entire population. These results are summarized in the first column of Table IV. We observe that the mean MAE over the population is under one week, with the maximum at just over two weeks. This, accompanied by a low SD and a narrow 95% CI suggests that the method is robust for estimating the FMA.

Furthermore, in Figure 5 we observe the scatter plot of the PMA versus the FMA for the entire population, grouped by terms. We observe that most observations fall reasonably close to the dotted line which represents a perfect prediction, for which the PMA and the FMA would be equal.

Most of the observations that are far from this line belong to EP infants (blue). This is also the case for some VP infants (green). While LP (yellow), ET (red), and FT infants (magenta) tend to have less error.

Category	Linear Regression	Random Forest Regression
Time-domain	sdRR (50%) minRR (70%) Skeweness (84%) Kurtosis (56%)	minRR (92%)
Frequency-domain	LFnu (62%) HFnu (58%)	LF (74%) HF (56%) LFnu (100%) HFnu (56%) LF_HF (76%)
Non-linear measurements	ApEn (60%)	ApEn (78%)
Visibility Graph Indexes	MD_VG (72%) C_VG 56%) r_VG (50%) C_HVG (80%) Tr_HVG (76%)	C_HVG (88%)
Non-HRV related	GA (90%)	GA (92%)

TABLE III: List of features selected by the genetic algorithm for the linear regression and random forest regression models. The percentage of LOOC iterations for which the feature was chosen is presented in parenthesis.

Data type	HRV	RRV	Bradycardia
Mean MAE (weeks)	0.93	1.39	1.39
Max. MAE (weeks)	2.01	4.33	4.31
Min. MAE (weeks)	0.03	0.02	0.03
SD (weeks)	0.54	0.93	0.99
95% CI (weeks)	[0.78, 1.08]	[1.12, 1.66]	[1.08, 1.69]

TABLE IV: Performance of the model on the HRV, RRV, and Bradycardia data, as measured by the mean, maximum, and minimum MAE, its SD and 95% CI.
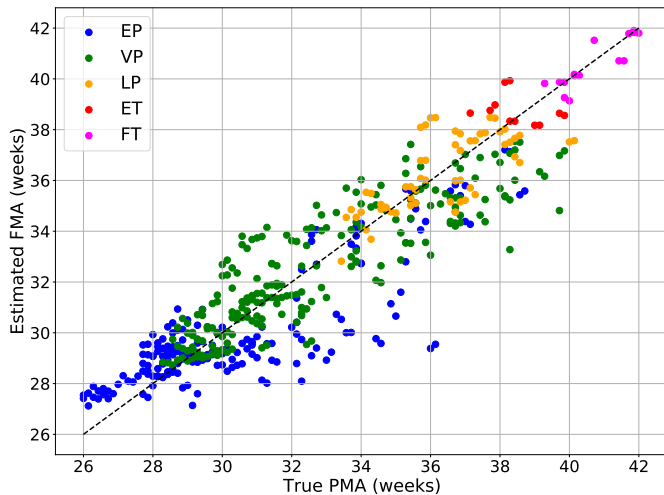
Fig. 5: PMA versus FMA for all infants in the HRV population, grouped by term.

C. Validation of the Model on RRV and bradycardia Data

We also evaluated the performance of the EML model when it was trained and tested on the bradycardia and the RRV data, respectively. The list of the selected features for these data types can be found in the Appendix A and Appendix B, respectively. The results regarding the estimation of the FMA are summarized in Table IV. We observe that both data types have very similar results. The mean MAE is low, at under 1.4 weeks for both cases. However, the performance is not as good as the one obtained on the HRV data, where we had a mean MAE of 0.93 weeks. Likewise, the Max. MAE is also higher for these data types, with a Max. MAE at around 4.3 weeks. On the other hand, the Min. MAE remains quite low and the result is comparable to that obtained with the HRV data. The SD for these data remains under one week, and the 95% CI also remains reasonably narrow. These results suggest that the method we propose for FMA estimation can be used on different data types with reliable results.

The higher error obtained when using the RRV and bradycardia data might be explained by the fact that these data types have, in general, a weaker correlation to the PMA as compared to the HRV data. This is particularly true in the case of RRV data, for which we opted to not perform the feature filtering by Spearman correlation, as it reduced the feature set to only a few variables. Another factor which might contribute to a lower performance is that both the RRV and bradycardia population (n=48 and n=43, respectively) were smaller than the HRV population (n=50), which translates into fewer data to train the model on, a difference that is specially marked for the bradycardia population, which is 14% smaller than the HRV population.

In Figure 6 we present the PMA versus FMA scatter plot for all the infants in the RRV and bradycardia populations. Although the results are more dispersed than for the HRV data, the general behaviour is the same, with EP (in blue) and VP (in green) infants presenting a smaller deviation in the FMA for smaller PMAs, and the deviation increasing as they

approach the term equivalent age.

IV. DISCUSSION

In this study we developed an automated estimator of the FMA for preterm and full-term infants, which can be used as a surrogate measure of the maturational age. A FMA similar to the PMA would suggest that the infant is following a normal maturational pattern, while a big deviation of the FMA with regard to the PMA would indicate to clinicians that there is a disruption in the maturation of the infant. The estimator is based on an EML model which is capable of making estimations using different types of data as input, which could be extracted from the heart rate monitoring of the infants, which is typically done in a NICU. While we focused our study on using HRV data to estimate the FMA, we also showed that the model we propose can make estimations using RRV and bradycardia data.

Previous studies took similar approaches, using brain imaging [27] or electroencephalogram recordings of preterm infants [28] in combination with machine learning algorithms to estimate a functional maturational age of the infants. However, we did not find any previous study that used HRV data to produce such an estimation of the FMA. Thus, this study proposes a novel and non-invasive approach to the evaluation of the maturation of infants using HRV data. Furthermore, all the previous studies we found that use machine learning to estimate a maturational age of infants were dedicated to a specific data type. We did not find any previous studies that propose a generic method for estimation of the maturational age using machine learning that can be applied to different types of data acquired in a NICU setting.

Another novel feature of our study is that it includes visibility graph indexes to characterize the HRV. This is a network-based analysis that has recently begun to be applied to HRV [52], and its use for the analysis of the HRV of premature infants has only recently been suggested ([6], [53]).

The EML model we propose combines LR and RFR to make the final predictions. This allows the model to exploit information from the features in the data, whether their correlation with PMA is linear or non-linear. To select the optimal feature set to train the EML model we used genetic algorithms. The feature selection process was done separately for the LR and for the RFR portions of the EML model, so that each element of the EML model would only use the features optimal for it. In the case of HRV, this yielded interesting results, showing that while LR predominantly favored time-domain and visibility graph features, RFR favored frequency-domains features, suggesting that the later have a non-linear correlation to the PMA of the infants.

An important characteristic of the proposed method is that in the population we used to train and validate the model we included preterm infants with varying degrees of prematurity, none of which manifested clinical signals of abnormal maturation for their GA. Consequently, this method is well suited to evaluate infants regardless of their GA. This is specially important in the case of the preterm infants because, to our knowledge, no normative HRV data has been proposed for this population.

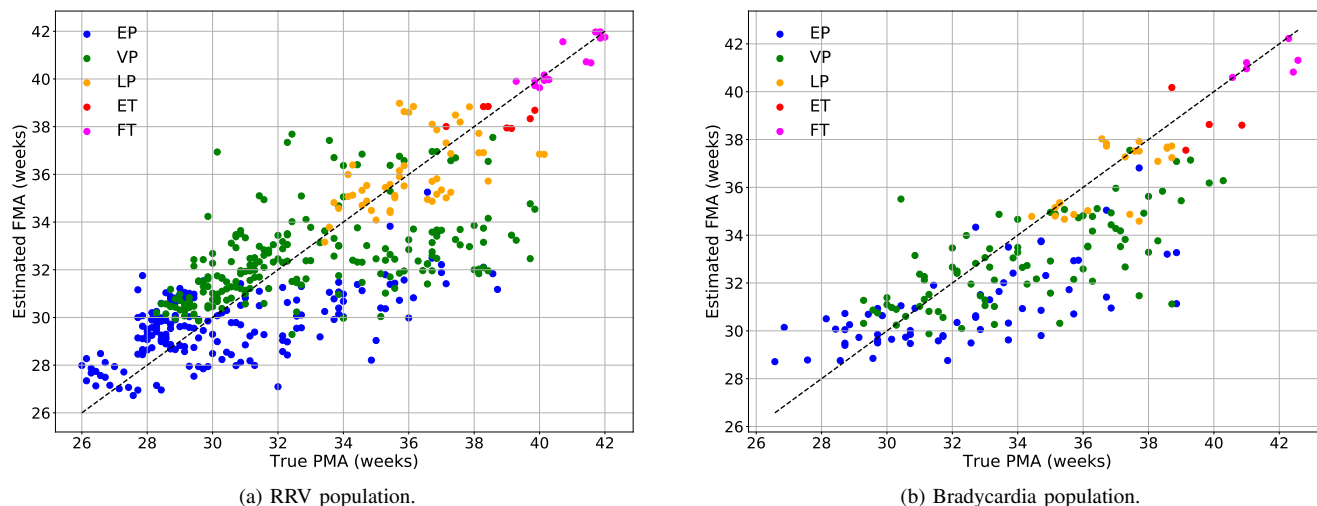(a) RRV population.   (b) Bradycardia population.

Fig. 6: True PMA versus Estimated FMA for all infants in the RRV and bradycardia population, grouped by term

The performance of the model was very accurate, with a mean MAE under one week when using the HRV data, and under 1.4 weeks when using the RRV and bradycardia data. These low errors are in agreement with the results we expected from the model on a population of infants with no signs of abnormal maturation relative to their GA, such as the one we used in this study. For this population we expected the PMA and FMA to be similar, translating into a low mean MAE, which signals that the model we propose is capable of accurately tracking a normal maturation pattern. Part of the error in the estimations done by our model might be accounted for by the inherent error associated to the PMA, due to the difference between the day of conception and the day of the last menstrual period.

However, another interesting observation was that EP and VP infants presented more deviation in their FMA when compared to their PMA. This might be explained by the fact that, as mentioned before, there are differences in the HRV of preterm infants at term equivalent age and that of infants born at term ([23], [24]). This explanation is supported by the fact that, for both the EP and the VP groups, the observations with the biggest difference between the PMA and the FMA occur for PMAs that are closer to term equivalent age. Thus, the increased deviation of the FMA with regard to the PMA as the preterm infants reach term equivalent age reflects this expected delay in their maturation. We consider this result to be further evidence that the model we propose offers a good measure of the maturational age of the infants, and that it is not simply overfitting to the PMA, as it is able to capture these well documented maturational delays for the EP and VP infants.

This suggests that the model we propose might be a reliable tool for estimating the FMA of neonates based on different physiological signals, and use it as surrogate measure of their maturational age. Changes in HRV, RRV and bradycardia, based on which our model estimates the FMA, reflect the maturational trajectories of autonomic regulation of heart rate and control of respiration which could be impacted by genetics,

environment and pathologies. We think that the quantification of the functional trajectories proposed in this study could be tested as a useful tool to individually evaluate neonatal maturation and to personalize care. In a NICU setting, the estimated FMA and its deviation from the PMA could help physicians evaluate the maturation process of the infants in real time and without need for invasive or additional tests.

V. LIMITATIONS OF THE PRESENT STUDY

Our study was limited to the first 12 weeks of chronological age due to constraints of the available data; however, further studies could be done to determine if this method could be applied to evaluate infants during a longer period of time. Also, a retrospective analysis to determine if premature infants whose FMA presented more deviation from the PMA display any signs of delayed maturation later in life is under study.

This study focused on the evolution of healthy infants during their hospitalization in NICU. In the future, we aim to test the method we propose on a population of infants who, during their hospitalization in NICU, presented complications associated to negative neurodevelopmental outcomes, such as a brain injury or sepsis. This would allow us to test the hypothesis that for these infants the FMA estimated by the EML model should differ significantly more from the PMA than it does for the population of healthy infants presented here.

Another limitation arises from the fact that we only obtained data from the healthy full-term infants around the time of their birth. Therefore, while the proposed method could serve to evaluate the maturation of full-term infants at birth, it could not be used to evaluate the progress of their maturation after birth as it does with preterm infants.

VI. CONCLUSION

In this study we propose an automated, non-invasive method for estimation of the maturation of infants during their first months of life, based on machine learning using different

HRV, RRV or bradycardia features. As these features are extracted from the heart rate or respiration rate monitoring of the patients, this method has the potential to be used in real time and as a bed-side tool in NICU settings, given that the EML model would require only a minimum of 30 minutes of continuous heart rate or respiration rate monitoring to produce an estimation of the FMA. The method we propose used genetic algorithms to find the optimal features for the machine learning algorithm, and combines linear regression and random forest regression to estimate the FMA of the infant.

We propose this estimated FMA as a surrogate measure of the maturation of the infants. An FMA similar to the PMA, which is measured clinically, could signal that the maturation of the infant is following a normal patterns for their GA, while a big difference between FMA and PMA could suggest a disruption in the infant's maturation. This information could assist clinicians in making decisions regarding ventilatory support, discharge, sleep management, and environmental care of the infants.

This study was done as part of the Digi-NewB project, which is a greater research project that aims to improve care for neonates. Digi-NewB has collected not only HRV, RRV, and bradycardia information, but also movement, cry, and sleep data from the infants in its cohort. In the future, we aim to integrate the analysis presented here with the results from the analysis of the other data types available in the project. The final goal is to produce an integrated decision support system that can, in real-time, help physicians evaluate the maturation process of the infants in different axes based on all the data available.

Further studies must be done with a population that includes infants with abnormal maturation patterns, to test the hypothesis that the model would be able to detect such disruptions in the maturation. In the future, we would also like to evaluate the feasibility of using the tool we have proposed here in a real-life and real-time setting.

APPENDIX A
BRADYCARDIA FEATURES

For the extraction of the bradycardia features, the bradycardia episodes were classified in four groups, depending on the method used for bradycardia detection and the threshold of beats per minute (BPM) considered:

a) Yellow (Yel): Bradycardia episodes detected by the Philips monitor and which triggered a yellow alarm.

b) Red: Bradycardia episodes detected by the Philips monitor and which triggered a red alarm.

c) Under 100 BPM: Bradycardia episodes detected from the ECG signal using the method described by Altuve et al. [41], and for which the heart rate fell under 100 BPM.

d) Under 80 BPM: Bradycardia episodes detected from the ECG signal by the method previously cited, and for which the heart rate fell under a threshold 80 BPM.

All four groups of bradycardia data were segmented into six-hour periods, from which several features were extracted:

a) Length (len): Duration of each bradycardia episode.

Feature	LR	RFR
lenYel (median)	Yes	Yes
$T_{DiffStart}$Yel (median)	No	Yes
TminYel (median)	No	Yes
lenYel (SD)	Yes	No
minBPM$_{Prev}$Yel (SD)	No	Yes
T_minYel (SD)	Yes	Yes
lenRed (median)	Yes	Yes
T_minmRed (median)	No	Yes
lenRed (SD)	No	Yes
minBPM$_{First}$Red (SD)	Yes	Yes
T_minRed (SD)	Yes	Yes
Slope$_{Prev}$100 (median)_wasMissing	Yes	No
Slope$_{First}$100 (median)_wasMissing	No	Yes
minBPM$_{Prev}$100 (median)_wasMissing	Yes	No
T$_{DiffStart}$100 (median)_wasMissing	No	Yes
len100 (median)_wasMissing	No	Yes
minBPM$_{First}$100 (median)_wasMissing	No	Yes
Slope$_{First}$100 (SD)_wasMissing	Yes	Yes
T$_{DiffStart}$100 (SD)_wasMissing	Yes	Yes
minBPM$_{First}$100 (SD)_wasMissing	Yes	No
len100 (SD)_wasMissing	No	Yes
T_min100 (SD)_wasMissing	Yes	Yes
Slope$_{First}$80 (median)_wasMissing	Yes	Yes
Slope$_{Prev}$80 (median)_wasMissing	Yes	Yes
T$_{DiffStart}$80 (median)_wasMissing	Yes	Yes
len80 (median)_wasMissing	Yes	Yes
T_min80 (median)_wasMissing	Yes	No
minBPM$_{Prev}$80 (median)_wasMissing	No	Yes
minBPM$_{First}$80 (median)_wasMissing	No	Yes
Slope$_{First}$80 (SD)_wasMissing	Yes	Yes
Slope$_{Prev}$80 (SD)_wasMissing	No	Yes
minBPM$_{Prev}$80 (SD)_wasMissing	Yes	No
minBPM$_{First}$80 (SD)_wasMissing	Yes	No
T_min80 (SD)_wasMissing	Yes	No
GA	Yes	Yes

TABLE V: Bradycardia features included in either the LR or RFR feature set. The name of the features correspond to the name or abbreviation of the variable (len, T$_{DiffStart}$, etc.), followed by the bradycardia group (Yel, Red, 100, and 80), and in parenthesis if it is the median value or the SD. Features with _wasMissing at the end indicate these are categorical features added to indicate if the value in the original feature was missing or not.

b) T_min: Time between the moment when the bradycardia episode was detected, and the time of the minimum BPM for that episode.

c) $T_{DiifStart}$: Time between two consecutive bradycardia episodes.

d) minBPM$_{Prev}$: The difference, in BPM, between the heart rate previous to the detection of the bradycardia, and the minimum BPM during the episode.

e) minBPM$_{First}$: The difference, in BPM, between the heart rate at the moment of detection of the bradycardia, and the minimum BPM during the episode.

f) Slope$_{Prev}$: Describes how fast the heart rate fell from the normal value previous to the bradycardia episode, to the minimum BPM during the episode. It is calculated by the

Feature Category	Feature	Feature Description	LR	RFR
$\mathbf{T}_{tot}$	Mean_T_{tot}	Mean of the T_{tot} time series	Yes	No
	Kurt_T_{tot}	Kurtosis of the T_{tot} time series	Yes	No
	Sk_T_{tot}	Skewness of the T_{tot} time series	Yes	No
	Med_T_{tot}	Median Value of the T_{tot} time series	Yes	No
	SD1_T_{tot}	Standard deviation of the points perpendicular to the line of symmetry of the Poincaré plot	Yes	No
	SD2_T_{tot}	Standard deviation of the points along the line of symmetry of the Poincaré plot	Yes	No
	SD2xSD1_T_{tot}	The multiplication of SD2_T_{tot} and SD1_T_{tot}	Yes	No
	SD2/SD1_T_{tot}	The ration between of SD2_T_{tot} and SD1_T_{tot}	Yes	No
	Rejection rate	The number of rejected T_{tot} values due to the respiration being saturated	Yes	No
$\mathbf{T}_{in}$	Median_T_{in}	Median value of the T_{in} time series	No	Yes
	SlopeInsp	Median value of the slope of the inspiration phase	Yes	No
$\mathbf{T}_{ex}$	Median_T_{ex}	Median value of the T_{ex} time series	Yes	No
Apnea	nbApnea3s	Number of apneas defined as cessation of breathing for more than three seconds	Yes	Yes
	nbApnea2cyc	Number of apneas defined as cessation of breathing with more than two missed respiratory cycles	Yes	Yes
	nbApnea3s2cyc	Number of apneas with cessation of breathing for more than three seconds and more than two missed respiratory cycles	No	Yes
	nbApnea10s	Number of apneas defined as cessation of breathing for more than ten seconds	Yes	No
Non-RRV related	GA	Gestational age	Yes	Yes

TABLE VI: Category, name, and description of all RRV features included in either the LR or RFR feature set.

following equation:

$$\text{Slope}_{Prev} = \frac{\text{minBPMn}_{Prev}}{\text{T\_min} + 0.1\text{s}} \tag{4}$$

g) Slope$_{First}$: Describes how fast the heart rate fell during the bradycardia episode. It is calculated by the following equation:

$$\text{Slope}_{First} = \frac{\text{minBPM}_{First}}{\text{T\_min}} \tag{5}$$

Finally, for each of these variables, in each of the four groups of detected bradycardia episodes, the median value and standard deviation (SD) over the six-hour segment were calculated. Those median values and SD were the features used as input for the FMA estimation algorithm described in this study.

Some types of bradycardia episodes studied might not have been present in all six-hour segments in the database, resulting in missing or null values. Thus, some categorical features to indicate whether a value was missing or not were generated by the algorithm when handling the missing values, as explained in Section II-G.

The final features sets selected by the genetic algorithm for the LR and RFR model, including both features from the original feature set, and features generated during the handling of missing data, are presented in Table V.

APPENDIX B
RRV FEATURES

For extracting the features used to describe the RRV, the respiration signal was acquired from clinically applied ECG leads using the trans-thoracic impedance. The signal thus obtained was first segmented into 30-minutes periods. From each segment, three time series were computed:

a) Inspiration phase (T$_{in}$): This time series reflects the variability in the duration of the inspiration phase, and was computed as the time differences between consecutive minima and maxima in the respiration signal.

b) Expiration phase (T$_{ex}$): This time series reflects the variability in the duration of the expiration phase, and was computed as the time differences between consecutive maxima and minima in the respiration signal.

c) Breathing cycle (T$_{tot}$): This time series reflects the variability in the duration of the entire breathing cycle, and was computed as the time differences between consecutive minima in the respiration signal.

From each of the these time series different features were extracted, resulting also in three category of features. A fourth category of RRV features was used to quantify the number of episodes of apnea.

The category, name, and a brief description of the features is presented in Table VI. For simplicity, we only present the features that were selected by the genetic algorithm for either the LR or RFR model, or both. The last two columns of the table represent whether the feature was included or not in the LR and/or RFR feature sets.

REFERENCES

[1] U. Rajendra Acharya, K. Paul Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: A review," *Medical and Biological Engineering and Computing*, vol. 44, no. 12, pp. 1031–1051, 2006.

[2] E. Longin, T. Schaible, T. Lenz, and S. König, "Short term heart rate variability in healthy neonates: Normative data and physiological observations.," *Early Human Development*, vol. 81, no. 8, pp. 663–671, Aug. 2005.

[3] H. Patural, V. Pichot, S. Flori, A. Giraud, P. Franco, P. Pladys, A. Beuchée, F. Roche, and J.-C. Barthelemy, "Autonomic maturation from birth to 2 years: Normative values," *Heliyon*, vol. 5, no. 3, e01300, 2019.

[4] V. L. Schechtman, R. M. Harper, K. A. Kluge, A. J. Wilson, H. J. Hoffman, and D. P. Southall, "Heart rate variation in normal infants and victims of the sudden infant death syndrome," *Early Human Development*, vol. 19, no. 3, pp. 167 –181, 1989.

[5] R. Joshi, D. Kommers, L. Oosterwijk, L. Feijs, C. van Pul, and P. Andriessen, "Predicting neonatal sepsis using features of heart rate variability, respiratory characteristics, and ecg-derived estimates of infant motion," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 3, pp. 681–692, 2020.

[6] C. León, G. Carrault, P. Pladys, and A. Beuchée, "Early detection of late onset sepsis in premature infants using visibility graph analysis of heart rate variability," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 4, pp. 1006–1017, 2021.

[7] R. Statello, L. Carnevali, D. Alinovi, F. Pisani, and A. Sgoifo, "Heart rate variability in neonatal patients with seizures," *Clinical Neurophysiology*, vol. 129, no. 12, pp. 2534 –2540, 2018.

[8] T. Al-Shargabi, R. B. Govindan, R. Dave, M. Metzler, Y. Wang, A. du Plessis, and A. N. Massaro, "Inflammatory cytokine response and reduced heart rate variability in newborns with hypoxic-ischemic encephalopathy," *Journal of Perinatology*, vol. 37, no. 6, pp. 668–672, 2017.

[9] H. Blencowe, S. Cousens, M. Z. Oestergaard, D. Chou, A.-B. Moller, R. Narwal, A. Adler, C. Vera Garcia, S. Rohde, L. Say, and J. E. Lawn, "National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: A systematic analysis and implications.," *The Lancet*, vol. 379, no. 9832, pp. 2162–2172, Jun. 2012.

[10] R. Y. Moon, R. S. Horne, and F. R. Hauck, "Sudden infant death syndrome," *The Lancet*, vol. 370, no. 9598, pp. 1578 –1587, 2007.

[11] M. H. Malloy, "Prematurity and sudden infant death syndrome: United states 2005–2007," *Journal of Perinatology*, vol. 33, no. 6, pp. 470–475, 2013.

[12] B. J. Stoll, N. Hansen, A. A. Fanaroff, L. L. Wright, W. A. Carlo, R. A. Ehrenkranz, J. A. Lemons, E. F. Donovan, J. E. T. Ann R. Stark, W. Oh., C. R. Bauer, S. B. Korones, S. Shankaran, A. R. Laptook, D. K. Stevenson, L.-A. Papile, and W. K. Poole, "Late-Onset Sepsis in Very Low Birth Weight Neonates: The Experience of the NICHD Neonatal Research Network," *Pedriatics*, vol. 110, no. 2, pp. 285–291, Aug. 2002.

[13] M.-H. Tsai, J.-F. Hsu, S.-M. Chu, R. Lien, H.-R. Huang, M.-C. Chiang, R.-H. Fu, C.-W. Lee, and Y.-C. Huang, "Incidence, clinical characteristics and risk factors for adverse outcome in neonates with late-onset sepsis.," *Pediatric Infectious Disease Journal*, vol. 33, no. 1, e7–e13, Jan. 2014.

[14] G. M. Ronen, S. Penney, and W. Andrews, "The epidemiology of clinical neonatal seizures in newfoundland: A population-based study," *The Journal of Pediatrics*, vol. 134, no. 1, pp. 71 –75, 1999.

[15] R. D. Sheth, G. R. Hobbs, and M. Mullett, "Neonatal seizures," *Journal of Perinatology*, vol. 19, no. 1, pp. 40–43, 1999.

[16] K. R. Gopagondanahalli, J. Li, M. C. Fahey, R. W. Hunt, G. Jenkin, S. L. Miller, and A. Malhotra, "Preterm hypoxic-ischemic encephalopathy," *Frontiers in Pediatrics*, vol. 4, p. 114, 2016.

[17] S. Cardoso, M. J. Silva, and H. Guimarães, "Autonomic nervous system in newborns: A review based on heart rate variability.," *Childs Nervous System*, vol. 33, no. 7, pp. 1053–1063, Jul. 2017.

[18] F. A. Selig, E. R. Tonolli, E. V. C. M. d. Silva, and M. F. d. Godoy, "Heart rate variability in preterm and term neonates.," *Arq Bras Cardiol*, vol. 96, no. 6, pp. 443–449, Jun. 2011.

[19] E. Longin, T. Gerstner, T. Schaible, T. Lenz, and S. König, "Maturation of the autonomic nervous system: Differences in heart rate variability in premature vs. term infants," *Journal of perinatal medicine*, vol. 34, pp. 303–8, Feb. 2006.

[20] R. Joshi, D. Kommers, C. Guo, J.-W. Bikker, L. Feijs, C. van Pul, and P. Andriessen, "Statistical modeling of heart rate variability to unravel the factors affecting autonomic regulation in preterm infants," *Scientific reports*, vol. 9, no. 1, pp. 7691–7691, May 2019.

[21] T Gerstner, J Sprenger, T Schaible, C Weiss, and S Koenig, "Maturation of the autonomic nervous system: Differences in heart rate variability at different gestational weeks.," German, *Z Geburtshilfe Neonatol*, vol. 214, no. 1, pp. 11–14, Jan. 2010.

[22] T. Nakamura, H. Horio, S. Miyashita, Y. Chiba, and S. Sato, "Identification of development and autonomic nerve activity from heart rate variability in preterm infants," *Bio Systems*, vol. 79, no. 1-3, pp. 117–124, 2005.

[23] H. Patural, V. Pichot, F. Jaziri, G. Teyssier, J.-M. Gaspoz, F. Roche, and J.-C. Barthelemy, "Autonomic cardiac control of very preterm newborns: A prolonged dysfunction.," *Early Human Development*, vol. 84, no. 10, pp. 681–687, Oct. 2008.

[24] E. Helander, N. Khodor, A. Kallonen, A. Värri, H. Patural, G. Carrault, and P. Pladys, "Comparison of linear and non-linear heart rate variability indices between preterm infants at their theoretical term age and full term newborns," in *EMBEC & NBC 2017*, H. Eskola, O. Väisänen, J. Viik, and J. Hyttinen, Eds., Singapore: Springer Singapore, 2018, pp. 153–156.

[25] K. L. Fyfe, S. R. Yiallourou, F. Y. Wong, A. Odoi, A. M. Walker, and R. S. C. Horne, "The effect of gestational age at birth on post-term maturation of heart rate variability.," *Sleep*, vol. 38, no. 10, pp. 1635–1644, Oct. 2015.

[26] S. B. Mulkey, S. Kota, C. B. Swisher, L. Hitchings, M. Metzler, Y. Wang, G. L. Maxwell, R. Baker, A. J. du Plessis, and R. Govindan, "Autonomic nervous system depression at term in neurologically normal premature infants.," *Early Human Development*, vol. 123, pp. 11–16, Aug. 2018.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2021.3093096, IEEE Journal of Biomedical and Health Informatics

12

[27] C. D. Smyser, N. U. Dosenbach, T. A. Smyser, A. Z. Snyder, C. E. Rogers, T. E. Inder, B. L. Schlaggar, and J. J. Neil, "Prediction of brain maturity in infants using machine-learning algorithms," *NeuroImage*, vol. 136, pp. 1 –9, 2016.

[28] N. J. Stevenson, L. Oberdorfer, N. Koolen, J. M. O'Toole, T. Werther, K. Klebermass-Schrehof, and S. Vanhatalo, "Functional maturation in preterm infants measured by serial recording of cortical activity," *Scientific Reports*, vol. 7, no. 1, p. 12 969, 2017.

[29] M. Chiera, F. Cerritelli, A. Casini, N. Barsotti, D. Boschiero, F. Cavigioli, C. G. Corti, and A. Manzotti, "Heart rate variability in the perinatal period: A critical and conceptual review.," *Frontiers in neuroscience*, vol. 14, p. 561 186, 2020.

[30] Digi-NewB. (2020). "Digi-newb a new generation monitoring system in neonatology," [Online]. Available: http://www.digi-newb.eu/ (visited on 03/01/2021).

[31] F Rusconi, M Castagneto, L Gagliardi, G Leo, A Pellegatta, N Porta, S Razon, and M Braga, "Reference values for respiratory rate in the first 3 years of life.," *Pediatrics*, vol. 94, no. 3, pp. 350–355, Sep. 1994.

[32] S. Fleming, M. Thompson, R. Stevens, C. Heneghan, A. Plüddemann, I. Maconochie, L. Tarassenko, and D. Mant, "Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: A systematic review of observational studies.," *The Lancet*, vol. 377, no. 9770, pp. 1011–1018, Mar. 2011.

[33] S. A. Lorch, L. Srinivasan, and G. J. Escobar, "Epidemiology of apnea and bradycardia resolution in premature infants," *Pediatrics*, vol. 128, no. 2, e366–e373, 2011.

[34] M. A. Mohr, K. D. Fairchild, M. Patel, R. A. Sinkin, M. T. Clark, J. R. Moorman, D. E. Lake, J. Kattwinkel, and J. B. Delos, "Quantification of periodic breathing in premature infants," *Physiological measurement*, vol. 36, no. 7, pp. 1415–1427, Jul. 2015.

[35] B. R. Greene, P. de Chazal, G. Boylan, R. B. Reilly, C. O'Brien, and S. Connolly, "Heart and respiration rate changes in the neonate during electroencephalographic seizure," *Medical and Biological Engineering and Computing*, vol. 44, no. 1, pp. 27–34, 2006.

[36] M. S. Miller, K. M. Shannon, and G. T. Wetzel, "Neonatal bradycardia," *Progress in Pediatric Cardiology*, vol. 11, no. 1, pp. 19 –24, 2000.

[37] "Age terminology during the perinatal period," *Pediatrics*, vol. 114, no. 5, pp. 1362–1364, 2004.

[38] Y. Zhang and Y. Yang, "Cross-validation for selecting a model selection procedure," *Journal of Econometrics*, vol. 187, no. 1, pp. 95–112, Jul. 2015.

[39] J. Mendes-Moreira, C. Soares, A. M. Jorge, and J. F. D. Sousa, "Ensemble approaches for regression: A survey," *ACM Computer Surveys*, vol. 45, no. 1, Dec. 2012.

[40] X. Navarro, F. Porée, A. Beuchée, and G. Carrault, "Artifact rejection and cycle detection in immature breathing: Application to the early detection of neonatal sepsis," *Biomedical Signal Processing and Control*, vol. 16, pp. 9 –16, 2015.

[41] M. Altuve, G. Carrault, A. Beuchée, P. Pladys, and A. I. Hernández, "On-line apnea-bradycardia detection using hidden semi-markov models," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 4374–4377.

[42] M. Doyen, D. Ge, A. Beuchée, G. Carrault, and A. I. Hernández, "Robust, real-time generic detector based on a multi-feature probabilistic method," *PLOS ONE*, vol. 14, no. 10, pp. 1–22, Oct. 2019.

[43] F. Shaffer and J. P. Ginsberg, "An Overview of Heart Rate Variability Metrics and Norms," *Frontiers in Public Health*, vol. 5, p. 258, 2017.

[44] M. Bolanos, H. Nazeran, and E. Haltiwanger, "Comparison of heart rate variability signal features derived from electrocardiography and photoplethysmography in healthy individuals," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2006, pp. 4289–4294.

[45] A. Bauer, J. W. Kantelhardt, P. Barthel, R. Schneider, T. Mäkikallio, K. Ulm, K. Hnatkova, A. Schömig, H. Huikuri, A. Bunde, M. Malik, and G. Schmidt, "Deceleration capacity of heart rate as a predictor of mortality after myocardial infarction: cohort study," *The Lancet*, vol. 367, no. 9523, pp. 1674–1681, May 2006.

[46] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. V. Boas, "Characterization of complex networks: A survey of measurements," *Advances in Physics*, vol. 56, no. 1, pp. 167–242, 2007.

[47] M. E. J. Newman, "Assortative Mixing in Networks," *Physical Review Letters*, vol. 89, no. 20, Oct. 2002.

[48] L. Lacasa, B. Luque, J. Luque, and J. C. Nuño, "The visibility graph: A new method for estimating the Hurst exponent of fractional Brownian motion," *EPL (Europhysics Letters)*, vol. 86, no. 3, p. 30 001, May 2009.

[49] B. Luque, L. Lacasa, F. Ballesteros, and J. Luque, "Horizontal visibility graphs: Exact results for random time series," *Physical Review E*, vol. 80, no. 4, p. 046 103, Oct. 2009.

[50] M. Zhao, C. Fu, L. Ji, K. Tang, and M. Zhou, "Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5197–5204, 2011.

[51] R. H. H. Groenwold, "Informative missingness in electronic health record systems: The curse of knowing," *Diagnostic and Prognostic Research*, vol. 4, no. 1, p. 8, 2020.

[52] X. Sun, Y. Zhao, and X. Xue, "Analyzing spatial characters of the ecg signal via complex network method," in *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*, vol. 3, Oct. 2011, pp. 1650–1653.

[53] T. Nguyen Phuc Thu, A. I. Hernández, N. Costet, H. Patural, V. Pichot, G. Carrault, and A. Beuchée, "Improving methodology in heart rate variability analysis for the premature infants: Impact of the time length," *PLOS ONE*, vol. 14, no. 8, pp. 1–14, Aug. 2019.