IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 19 June 2024, accepted 11 July 2024, date of publication 15 July 2024, date of current version 25 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3428993

RESEARCH ARTICLE

HRV-Based Regression Analysis in Newborns With Sepsis: Forecasting BAYLEY-III Scores at 6 and 12 Months

LORENZO FRASSINETI[1], (Member, IEEE), GIOVANNA BERTINI[2], FEDERICO CALÀ[1], (Member, IEEE), CARLO DANI[3], ANGELA PARENTE[1], CARLOS ALBERTO REYES GARCÍA[4], (Senior Member, IEEE), CLAUDIA MANFREDI[1], SILVIA LORI[5], SIMONETTA GABBANINI[5], CLARA LUNARDI[3], CESARINA COSSU[5], MARIA BASTIANELLI[5], CATERINA COVIELLO[2], AND ANTONIO LANATÀ[1], (Member, IEEE)

[1]Department of Information Engineering, Università degli Studi di Firenze, 50139 Florence, Italy
[2]Division of Neonatology, Careggi University Hospital, 50134 Florence, Italy
[3]Department of NEUROFARBA, Careggi University Hospital, Università degli Studi di Firenze, 50134 Florence, Italy
[4]Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Puebla 72840, Mexico
[5]Neurophysiology Unit, Neuro-Musculo-Skeletal Department, Careggi University Hospital, 50134 Florence, Italy

Corresponding author: Lorenzo Frassineti (lorenzo.frassineti@unifi.it)

This work was supported in part by RSI—Programma Operativo Regionale (POR) Fondo europeo di sviluppo regionale (FESR) TOSCANA 2014/2020 BANDO n. 2 DDR n. 3389 30/07/2014 Project: DI-ASD CUP Sviluppo Toscana under Grant 3389.30072014.068000080; in part by the Fondazione Cassa di Risparmio di Firenze, Florence, Italy, under Project 2018.0976; and in part by Project PE8:"Conseguenze e sfide dell'invecchiamento-AGE-IT-Ageing Individuals in an Ageing Society through the Building Institutional, Biomedical and Technological Solutions for a Successful Italian Aging Society" under Grant CUP B83C22004800006.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of Careggi University Hospital, Firenze, Italy (Local Ethical Committee) under Approval No. DOL1 01/2018, and performed in accordance with the Declaration of Helsinki.

ABSTRACT The early detection of neurodevelopmental disorders in newborns is of utmost importance in clinical practice. Recently, to predict the neurodevelopment scores in preterms, Artificial Intelligence (AI) methods have been proposed mainly based on Electroencephalographic (EEG) or heart rate variability (HRV) analysis. In this work, HRV measures of preterm newborns with and without Sepsis are computed and used as input features of AI regression models. The study assesses the reliability of such features in predicting BAYLEY-III scores obtained during the clinical follow-up at 6- and 12-months. Forty-eight preterms (gestational age 27.8±1.8 weeks) were involved, 27 of which were diagnosed with Sepsis. HRV analysis was performed on ECG signals recorded at the corrected term age. BAYLEY-III score prediction was implemented, considering HRV features as input predictors of ensemble regression models. Models were validated using the Leave-One-Subject-Out (LOSO) framework. Encouraging results were achieved, with a Mean Absolute Error (MAE) < 5 points for the Sepsis group in the BAYLEY-III cognitive and language scales at 6- and 12-months. Preliminary results suggested that the autonomic nervous system development may be linked to central nervous system maturation. HRV features, and AI regression models could predict alterations that affect the correct neurodevelopment of newborns.

INDEX TERMS BAYLEY-III, entropy, HRV, neonatal sepsis, neurodevelopment, preterm, support vector regression.

I. INTRODUCTION

Almost 15 million infants are born each year prematurely, about 10% of the worldwide neonatal population [1]. In Neonatal Intensive Care Units (NICUs), at least 33% of hospitalizations are related to preterm newborns. The preterm birth rate, defined as the ratio between preterm births and the number of newborns born alive, increased from 9.6% in 2005 [2] to 11.1% in 2010 [3]. The risk of death increases in newborns with gestational age (GA) lower than 34 weeks [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Yiqi Liu.

Moreover, neonatal death mainly occurs during the first week of life (almost 80% [5]). The survival rate of preterm newborns varies worldwide: in developed countries, the percentage of survived newborns after the first week, with a GA lower than 28 weeks, is 90%. Instead, in underdeveloped countries, this percentage drops to 10% [6].

A preterm birth may carry several complications, such as Necrotizing EnteroColitis (NEC) or Broncho Pulmonary Dysplasia (BPD), which are the most common causes of neonatal death, with almost 3 million deaths per year [3]. Among the death causes, Sepsis represents the third most common cause for newborns [7].

In adults, Sepsis is defined as a life-threatening organ dysfunction caused by a dysregulated response to infection [8]. For newborns, still, a unique and shared definition of neonatal Sepsis does not exist [9], [10], [11].

The clinical signs of a neonatal infection are numerous, unspecific, and challenging to be detected. Recently, improvements have been achieved in managing and treating preterm newborns with Sepsis, and the survival procedures increased the percentage of survivals. However, preterm newborns are more sensitive to neurodevelopmental delay or diseases when compared to at-term newborns [12]. More than 25% of newborns with GA between the 28th and 32nd week show a neurodevelopment delay, usually related to several degrees of impairment [5]. Recently, it has been argued that Sepsis may negatively impact the neurodevelopment of surviving newborns. It may cause significant alterations of the cerebral networking in the neonatal period and could harm brain development [13], [14], [15], [16]. Furthermore, the early detection of neurodevelopmental disorders or delay is of utmost importance in clinical practice, the first two years of life being the most vulnerable and critical period of neurodevelopment [17]. Thus, the newborn at risk or with sepsis-related damages should be identified as soon as possible to define the best neuro-rehabilitative program [17].

The clinical staff often uses neurodevelopment scales such as the Bayley Scales of Infant and Toddler Development or Griffiths Mental Development Scales [18] to monitor neurodevelopment and detect abnormal behaviours. These scales generally consist of a list of tests and tasks the clinicians administer at different follow-up periods, usually from 3-6 months after birth up to 18-24 months after birth (considering the corrected age for preterms). Recently, Artificial Intelligence (AI) methods were proposed to predict neurodevelopment scores, mainly based on Electroencephalographic (EEG) signals recorded from newborns during or immediately after their stay in NICU [19], [20]. Thus, these methods could support the clinical staff in the early detection of newborns at risk of neurodevelopmental disorders. Moreover, neurodevelopment itself is altered by preterm birth [21], and changes could be detected by the analysis of the Autonomic Nervous System (ANS) [22]. Precisely, Heart Rate Variability (HRV) analysis reflects changes in ANS activity and can provide information about the development of newborns [23]. HRV analysis can be obtained by Electrocardiographic (ECG) signals that are less invasive and cheaper than EEG. Furthermore, previous studies have suggested the associations between HRV and neurodevelopmental outcome [24]. However, only a few studies have investigated HRV dynamics and AI methods to predict neurodevelopmental scores [23]. This work is the first study concerning HRV analysis, BAYLEY-III scores and AI regression models for newborns suffering from Sepsis. We evaluated if HRV parameters applied as input features of regression models may provide a reliable estimate of the BAYLEY-III score during follow-ups at 6 and 12 months for preterm newborns with and without Sepsis. This work is organized as follows: Section II describes the dataset, the pre-processing applied to ECG signals, the extracted features, and the validation scheme applied to regression models to evaluate their performance in predicting BAYLEY-III scores. In Section III, the results are shown. Sections IV and V discuss the results and conclusions about using HRV features to predict neurodevelopmental scores.

II. MATERIAL AND METHODS

A retrospective dataset of ECG recordings was collected at the Neuro-physiopathology and Neonatal Intensive Care Unit of Careggi University Hospital (Firenze, Italy). ECG signals were recorded using the Nemus ICU Galileo NT Line system - EB Neuro S.p.A. (Firenze, Italy). The dataset was collected in the years 2018 - 2023. Preterms with a GA lower than 37 weeks were initially included in the study. Infants with congenital brain malformations or severe brain injury defined as the occurrence of intraventricular haemorrhage $\geq$ 3 grade and cystic periventricular leukomalacia were excluded. ECG signals were recorded within time windows of about 54$\pm$9 minutes, with a sampling frequency of 128Hz. The ECG signal is recorded with pre-gelled disposable surface electrodes placed on intact and clean skin of a bone surface. The recording follows the peripheral bipolar lead D1. The landmarks are identified between the right clavicle (negative pole) and the left clavicle (positive pole). The study was performed according to the Declaration of Helsinki and approved by the Institutional Review Board of Careggi University Hospital, Firenze, Italy (Local Ethical Committee approval code: DOL1 01/2018), for all newborns, parents or guardians provided written informed consent. It involved 64 preterm newborns with gestational age (GA) between 24 and 31 weeks (27.8$\pm$1.8 weeks). None of the considered newborns had heart disease, which could affect the study results. The age of newborns at the time of ECG recordings was between 37 and 43 weeks (38.5$\pm$1.5 weeks). However, only 48 out of 64 subjects were considered, as the ECG signals of 16 subjects were corrupted by noise. Moreover, 27 out of 48 patients had Sepsis during hospitalization. The distribution of cases between Sepsis (S) and No-Sepsis (NS), as far as the age at the time of the recording (i.e. the postmenstrual age PMA), gender, and the GA, are shown in Figure 1. Further details about the subjects involved in the study are reported in Table 1. Pearson's χ^2 test

and Mann-Whitney tests confirmed no statistical differences as far as gender, GA and age are concerned between the two groups considered: S and NS (Table 1, Mann-Whitney test, level of confidence 0.05). Because this is a retrospective study, the exact PMA when these infants had Sepsis was not recorded for all the infants. Therefore, this information was not included in the study.

The BAYLEY-III scales were administered by an expert psychologist at Careggi University Hospital, Firenze, Italy. Each BAYLEY-III scale comprises different items, and their administration is flexible but with a standard order. Moreover, the number of items varies according to the subject's age. For example, the cognitive scale (maximum number of items 91) evaluates elements such as the development of spatial exploration, memory, manipulation, the relationship between objects and concepts, and information comprehension. It is usually the first administered scale and requires a good capacity of participation of the involved subject. According to the BAYLEY-III tests, scores below 85 and 70/75 denote a moderate or severe impairment, respectively [18], while scores between 85/90-100 and higher are associated with normality conditions. The cognitive, language, and motor scores were collected at 6-month and 12-month follow-ups. Table 2 shows the BAYLEY-III scores. As this study is still ongoing, only the scores at 6 and 12 months of life were available for a large enough number of infants.

Regarding ECG, the Bayley-II scores of each domain were available for the whole dataset at six months, while at 12 months, scores were available for 38 subjects only. As shown in Table 2, statistically significant differences in BAYLEY-III scores were found between the two groups for the cognitive and language scales at 6- and 12-months, respectively (Mann-Whitney test, level of significance 0.05).

A. HRV ANALYSIS

For HRV analysis, a sliding time window of 5 minutes duration without overlap was applied on the entire recording [25]. Then, inter-beat-interval (IBI) time series were obtained using the Pan-Tompkins algorithm [26]. Besides the exclusion of the subjects with ECGs corrupted by noise, for the remaining subjects, the time windows with ectopic beats or QRS detection errors were visually evaluated and excluded before the HRV feature extraction step.

The evaluation was done considering at the same time the tachogram, the original 5-minute ECG window and its corresponding R peak detections. This analysis was performed using the MATLAB version of the Pan-Tompkins algorithm proposed in [27].

According to [28], [29], and [30], 82 HRV features were extracted to characterize the newborn's ANS. Specifically, the following features for all the 5-minute windows of each subject were considered:

- Time-domain features: Heart Rate (HR), standard deviation of RR intervals (SDRR) and successive differences (SDSD), percentage of successive RR intervals >50ms (pNN50), root mean square of successive RR interval differences (RMSSD), HRV triangular index (TRI), Triangular Interpolation of the NN interval histogram (TINN), Poincarè plot standard deviation along the line of identity (SD2), SD1/SD2, where SD1 is the Poincarè standard deviation perpendicular to the line identity (SD1SD2ratio), and correlation dimension (CD).
- Frequency-domain features: the absolute power of Very Low Frequency (VLF), Low Frequency (LF) and High Frequency (HF), relative power for LF and HF (pLF and pHF), total power (TP) and LF to HF ratio (LF/HF).
- Entropy-domain features: approximate entropy (ApEn), multiscale sample entropy from scale 1 to scale 20 (MSE1 $\cdots$ MSE20) [31], multiscale distribution entropy from scale 1 to scale 20 (MDE1$\cdots$MDE20) [31], multiscale fuzzy entropy from scale 1 to scale 20 (MFE1 $\cdots$ MFE20) [31] and Bubble Entropy [32]. The Complexity Index (CI) [33] for MSE, MDE and MFE was also computed.

Time- and frequency domain features were computed using the tool proposed in [34]. According to [31], [35], and [36], the frequency ranges were adapted to the neonatal case, the LF range was set to 0.04Hz - 0.3Hz, and the HF range to 0.3Hz $-$1.3Hz. For the VLF, the following range was considered: 0.003Hz - 0.04Hz. For multiscale entropy, the coarse-grained procedure was applied [31], [33]. For all the entropy indexes, including ApEn, the embedding dimension was set to m=2. For MSE, a threshold value r=0.2 of the standard deviation of the epoch was considered. For MDE, the number of bins was set as B=512. For MFE, the exponent n was equal to 2 [31]. After the extraction of HRV features, the following statistics descriptors were applied to each subject: mean, median, standard deviation (std), kurtosis, skewness, and interquartile range (iqr) [28]. Thus, each HRV feature provided six different statistics descriptors. A matrix of size Nx492 was obtained, where N is the number of patients (48 at 6 months and 38 at 12 months), and 492 is the number of statistics descriptors extracted from the original HRV features. This high number of features was considered because, to the best of our knowledge, this is the first attempt at employing HRV features to a regression problem to predict the BAYLEY-III scores in newborns with Sepsis. Combined with feature selection methods, these features are evaluated as input for the HRV-based regression models described in section II-B.

In Figure 2, a diagram resumes the main steps for the HRV feature extraction described in the current subsection. Moreover, in Figure 2, an illustrating example for each step from a single subject is shown (as HRV features, the HR one and its statistics descriptor are reported).

B. REGRESSION ANALYSIS

Only the training features were normalised before performing the regression model validation (Z-score). Figure 3 shows the implemented training and validation process workflow.

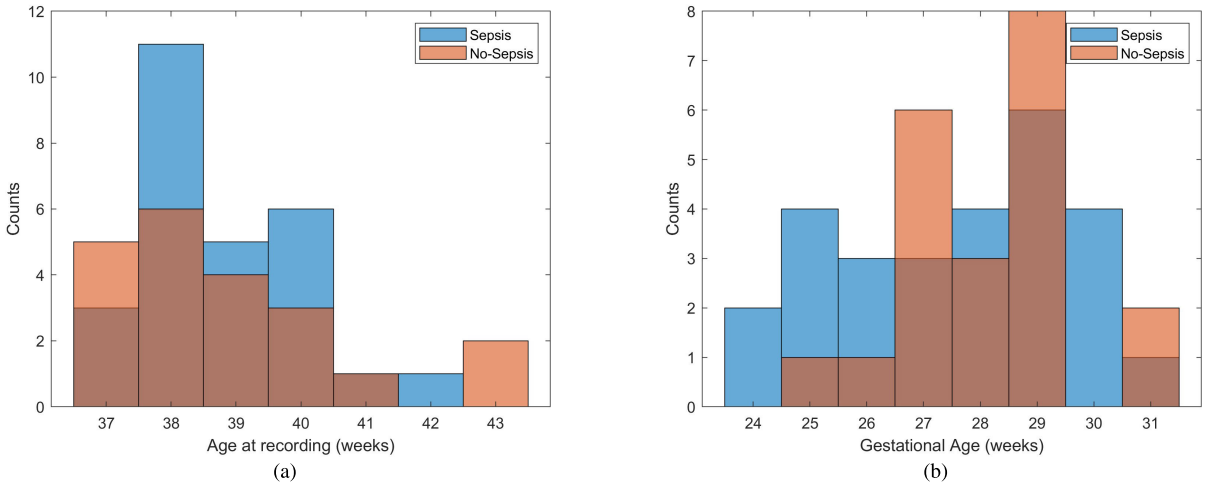


FIGURE 1. Histogram distribution of the ECG cohort. (a) The age in weeks at the time of ECG recordings for Sepsis and No-Sepsis cases. (b) the GA distribution for Sepsis and No-Sepsis cases.

TABLE 1. ECG dataset details, μ=mean, σ=standard deviation.

Group	GA $\mu\pm\sigma$ weeks	Age at ECG (PMA) $\mu\pm\sigma$ weeks	Cases (Male/Female)
S	27.5 ± 2.0	38.7 ± 1.3	27 (17/10)
NS	28.1 ± 1.5	38.8 ± 1.7	21 (8/13)

TABLE 2. BAYLEY-III scores for all the sub-scales for both groups: Sepsis (S) and No-Sepsis (NS). The p-value is related to the Mann-Whitney test. Iqr = interquartile range, m=months. (*) denotes a significant p-value (< 0.05).

	Group			
BAYLEY-III scores	S	NS	p-value	Cases (with Sepsis)
Cognitive 6m	90 (15)	100 (11)	0.01*	48 (27)
Language 6m	77 (12)	83 (11)	0.06	48 (27)
Motor 6m	87 (12)	91 (12)	0.14	48 (27)
Cognitive 12m	93 (20)	102 (15)	0.15	38 (27)
Language 12m	86 (10)	90 (8)	0.01*	38 (27)
Motor 12m	79 (9)	86 (10)	0.29	38 (27)

Moreover, the validation sets were modified according to the training statistics to avoid data leakage. Then, the highly correlated predictors were removed by applying a threshold equal to |0.80| to the Pearson correlation coefficient of all the HRV predictors [37], keeping only one for each correlation pair. The remaining features were ranked using the F-tests for regression (FSR) [38] or the ReliefF algorithm [39]. FSR tests the hypothesis that the response values grouped by the variable predictor values are drawn from populations with the same mean against the alternative hypothesis that the population means differ.

Thus, the reordered HRV features were used as input predictors of support vector regression (SVR) models (linear and Gaussian) or Ensemble regression models (Bag, LsBoost) [40]. Briefly, SVR represents the version of the Support Vector Machine (SVM) for regression problems [41]. They allow a generalization of the classification problem by estimating a continuous-valued multivariate function that defines relationships between input and output variables (in our case, the HRV features and the BAYLEY-III scores). The transition from SVM to SVR is possible by introducing an ϵ-region (i.e. a margin of tolerance) around the so-called ϵ-tube function [42], [43]. This function adapts the optimization problem for SVR, defining a convex ϵ insensitive loss function that will be minimized to find the flattest tube that contains most of the training observations [41]. Then, the multivariate function is built from the loss function and the geometrical properties of the tube. As in SVM, all the training observations outside the tube's boundary are called support vectors and define the hyperplane. SVR models deal with both linear and nonlinear feature space. For nonlinear functions, the observations can be mapped in the so-called kernel space by applying a nonlinear transformation (i.e. the kernel function). This work evaluated the linear kernel and the Gaussian kernel for SVR models. For a more detailed presentation of SVR theory, please see [41], [42]. Besides

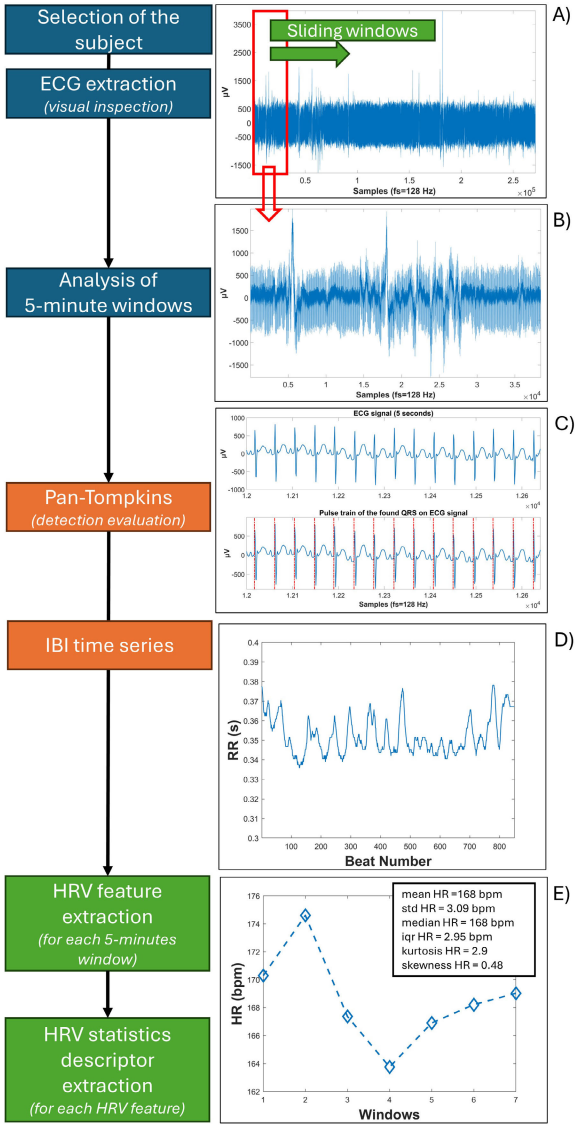


FIGURE 2. Diagram of the main steps described in Section IIA for the HRV analysis applied on a single subject. A) The original ECG signal for one subject. B) One of the 5-minute ECG windows extracted from this subject. C) A detail of the R-peak detections from the 5-minute window reported in B) - only 5 seconds are shown for resolution reasons. D) The tachogram of the 5-minute window. E) The HR values for each window extracted from the subject - a legend of the related statistics descriptors is reported in the top right corner.

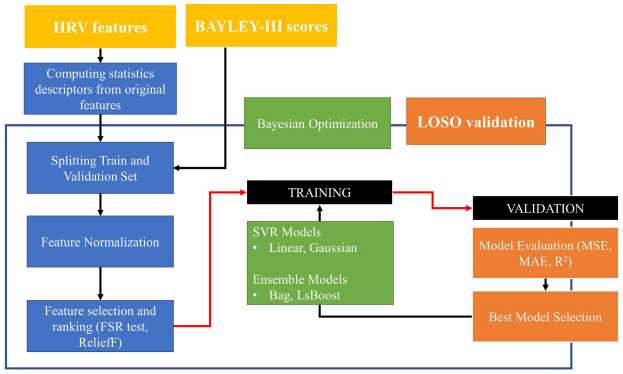


FIGURE 3. Workflow of the regression model training and validation procedure.

SVR models, in this work, ensemble models were evaluated. The first kind of model was the Bootstrap aggregation (Bagging, i.e. Bag) regressor ensemble model [44], [45]. In this work, the decision tree was considered a weak learner. Moreover, the predictors for each decision split were randomly selected for every tree in the ensemble following the random forest technique. In our case, the final prediction of BAYLEY-III scores was defined as the average over predictions from individual weak learners. Furthermore, in this work, the Least-Squares Boosting (LsBoost) regressor ensemble was also evaluated [46]. The LsBoost fits to minimize the mean squared error between the target variable (Y, in our case, the BAYLEY-III scores) and the combined predictions of the weak learners (Y_{pred}). As for the Bag ensemble, the regression trees were used as weak learners (B). Weak learners are trained in parallel in bagging while they learn sequentially in boosting. The algorithm starts with an initial estimation of the aggregated prediction of the target variable ($\widetilde{Y}$) as a function of predictor variables (here, the HRV features). During each training iteration, the LsBoost fits a new learner to the difference between the observed response and the aggregated prediction of all learners defined in previous steps, according to a learning rate parameter η (where $0 < \eta < 1$ as defined in equation 1). Furthermore, the LsBoost algorithm identifies misclassified observations and adjusts their weights ω to minimize the training error. Finally, LsBoost combines multiple regression models in a weighted manner.

$$Y_{pred}(HRV_{features}) = \widetilde{Y} + \eta \sum_{m=1}^{M} \omega_m B_m(HRV_{features}) \quad (1)$$

where M is the total number of weak learners for a model. Like a forward selection, it started with models that considered only the first feature ranked by F-tests or ReliefF and ended with models that considered all the features. This operation was performed during validation both for SVR and Ensemble models.

To evaluate the developed models, the Leave-One-Subject-Out (LOSO) validation was performed using the Bayesian Hyperparameter Optimization, searching for the model with the lowest Mean Absolute Error (MAE, equation 2).

For validating neonatal ML applications, the LOSO framework is widely accepted in the presence of data scarcity [47]. Thus, choosing LOSO allowed us to evaluate and estimate how the models may work in the presence of unseen data and different subjects. Certainly, using LOSO reduces possible bias in performance but may increase their variance simultaneously, depending on the dataset considered. However, this work aims not to test a system already validated in clinical practice but to validate a

preliminary first version of the models on a retrospective cohort. Therefore, the recruitment of further subjects as dedicated test sets should be considered to validate the methods in clinical practice. The number of iterations for Bayesian Optimization was set up to 500. Furthermore, the MAE metric was divided into MAE for the S group (hereinafter MAE1, the MAE of regression models applied only on subjects with Sepsis) and MAE for the NS group (hereinafter MAE0). The MAE is defined as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2}$$

where n is the number of observations evaluated, y_i the $i-th$ observed values and $\hat{y}_i$ is the corresponding value predicted by the regression model. In our case, the predicted values were the BAYLEY-III scores for each scale. During the LOSO validation, the following hyperparameters were optimized:

- For the SVR models: box constraint and kernel scale [29]. The Box constraint and Kernel scale were evaluated during the Bayesian Optimization in the $10^{-5} - 10^{5}$ range.
- For the Ensemble models: number of learning cycles, minimum leaf size, maximum number of splits and learn rate (only for LsBoost models). During the Bayesian Optimization, the learning rate was evaluated in the range 0.001 - 1, the number of learning cycles between 10 and 300, the minimum leaf size between 1 and 50, the maximum number of splits between 1 and the size of the current training set (e.g., for the cognitive scales at 6-months, the upper limit was set to 47).
- For both models (SVR and Ensemble): number of features, number of neighbours of ReliefF [39]. During the Bayesian Optimization, the number of neighbours of ReliefF was evaluated in the range of 1 - 15.

The Mean Squared Error (MSE) for both groups and the R^2 parameter was computed to have a complete overview of the models' performance. The MSE is defined as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3}$$

R^2, also known as the coefficient of determination, is a measure of the goodness of fit of a regression model. The R^2 formula is reported in equation 4.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{4}$$

where SS_{res} is the residual sum of squares and SS_{tot} is the total sum of squares. An ideal SVR model should obtain an R^2 equal to 1. This validation scheme was repeated for all the BAYLEY-III scales, generating HRV-based regression models. In other words, the output of each regression model was the predicted BAYLEY-III outcome compared with its real neurodevelopmental scores. Therefore, six different models were developed (three for the cognitive, language and motor scales at 6 months and another three at 12 months.) Moreover, to evaluate the level of agreement [48] between predictions and real BAYLEY-III outcomes, the Bland-Altman and correlation plot analysis [49] were assessed for each HRV-based model. For these evaluations, the Pearson correlation coefficient (r), the related p-value, the reproducibility coefficient (RPC) and the limits of agreement were investigated. The results related to regression analysis are shown in Section III.

III. RESULTS

This section reports the main results of the regression analysis described in section II-B. Table 3 shows the results of the LOSO validation for the BAYLEY-III scales using HRV features. In Table 3 also ΔMAE is reported, defined as the ratio between the MAE and the difference between the maximum and minimum values of the BAYLEY-III scores for each scale.

For the 6-month scores, the lowest MAE value was reached for the cognitive scale (MAE=4.8). It was obtained by a Bag regressor (Ensemble Model), with FSR as the feature selection method and using the following hyperparameters: number of learning cycles=83; minimum leaf size=1; maximum number of splits=9 and number of features=2. Specifically, the two features used are MSE2 (iqr) and the MDE18 (std). Note that for such a scale, the MAE was < 5 points both for the S and the NS group. No statistically significant differences were found in residuals between S and NS groups. However, as shown in Table 3, MAE1 obtained lower values than MAE0.

Furthermore, we analyzed MSE2 (iqr) and MDE18 (std) individually in terms of R^2 to confirm if this combination of features via regression is necessary. Interestingly, MSE2 (iqr) obtained a non-significant R^2 (0.02), instead MDE18 (std) obtained a statistically significant R^2 (0.27, p-value e-04), but still lower than the regression model (Table 3, $R^2 = 0.44$).

Instead, for the 12-month follow-up, the language scale was the best (Table 3, MAE=5.3 and R^2=0.32). It was obtained by an LsBoost regressor (Ensemble Model), with FSR as the feature selection method and using the following hyperparameters: learn rate=0.5488, number of learning cycles=14, minimum leaf size=12, maximum number of splits=23 and number of features=2. Specifically, the two features used are LF (mean) and MFE12 (mean).

As for the 6-month follow-up, we analyzed LF (mean) and MFE12 (mean) individually in terms of R^2. We did not obtain statistically significant R^2 (0.05 and 0.01, respectively) for both of them. A comparable result was also achieved for the cognitive scale (MAE=6.0 and R^2=0.27). Figure 4 shows the best HRV-based model prediction results obtained with the LOSO validation among all 6-month follow-ups. Specifically, Figure 4 illustrates the results for the Cognitive Scale at 6 months, comparing the predicted score with the real outcomes. The left plot shows the correlation plot, while on the right, the Bland-Altam plot [49] is reported. In Figure 4, NS subjects are shown with (○, blue circles) and S subjects

TABLE 3. Results of the LOSO validation for all the BAYLEY-III scales considered, using HRV features (m=months). (*) denotes a significant F-value (level of significance 0.05).FS = feature selection method, nf=number of features used in the best model.

BAYLEY-III scales	Regressor metrics					Model	FS (nf)
	MSE	MAE (MAE0-MAE1)	ΔMAE (%)	R^2	F-value		
Cognitive 6m	34	4.8 (4.8 – 4.7)	13%	0.44	25.95*	Bag	FSR (2)
Language 6m	56	5.7 (5.4 – 6.0)	14%	0.08	9.43*	Bag	ReliefF (2)
Motor 6m	76	6.2 (6.6 – 5.9)	18%	0.10	39.43*	LsBoost	corr (98)
Cognitive 12m	87	6.0 (7.5 – 4.9)	15%	0.27	34.49*	SVR	ReliefF(8)
Language 12m	51	5.3 (7.1 – 4.0)	17%	0.32	31.63*	LsBoost	FSR (2)
Motor 12m	47	5.6 (6.6 – 4.8)	21%	0.02	0.76	Bag	FSR (5)

with (◇, red diamonds), respectively. Similarly, Figure 5 shows the correlation plot and the Bland-Altam plot for the best HRV-based models obtained with LOSO validation among all the 12-month follow-ups. Specifically, the results related to the Language Scale at 12 months are shown.

Furthermore, a statistical analysis evaluated significant differences between S and NS for each HRV statistics descriptor introduced in Section II-A. Firstly, the Shapiro-Wilk test was performed for each comparison to check the normality assumption. When the normality assumption was violated, the non-parametric Mann-Whitney test was considered; otherwise, the two-sample t-test was used. The level of significance was 0.05. In Table 4, the statistically significant HRV statistics descriptors are resumed. Noteworthy, most of them were from the Entropy Domain. Instead, only pLF from the other domains was found to be significant. For clarity, most of these features were correlated with each other (e.g. mean and median features for the same HRV parameter). To remove redundancy from the dataset, feature selection methods were considered during the development of the regressor models (as explained in Section II-B). This analysis strengthens that differences between S and NS subjects might exist regarding HRV values and ANS functioning as already reported in [50] and [51].

IV. DISCUSSION

This work reports a first attempt to apply HRV-based regression models to predict the BAYLEY-III scores of newborns with and without Sepsis at different follow-ups. The research evaluates if the newborn's electroclinical characteristics at the corrected term age could predict their neurodevelopmental scores. Also, preterm newborns with sepsis episodes during their hospitalization were taken into account, as Sepsis could affect their neurodevelopment. Results suggest that the analysis of the ECG signal recorded when the preterm newborn reached the term age could support neonatologists or pediatric neurologists in the early detection of newborns at risk of neurodevelopmental delays.

Results shown in Table 3 suggest that, for some BAYLEY-III scales, HRV-based regression analysis could predict the neurodevelopmental scores of the newborns at 6 and 12 months of life. For the cognitive scale at 6 months and language scale at 12 months, regression models could predict the BAYLEY scores with an MAE lower or close to 5 points for the Sepsis group.

Interestingly, as reported in Figure 4, for the Cognitive scale at 6-month a significant correlation was found between the predicted score and the real BAYLEY-III outcome (rho=0.68, p-value=1.3e-07). Instead, a weaker significant correlation was found for the Language score at 12 months (Figure 5 rho=0.60, p-value=7.7e-05). This might suggest that these HRV models are more accurate for estimating the neurodevelopmental scale closer to the physiological recordings.

Furthermore, Table 3 shows that the motor scores always gave the worst performance on all the metrics considered. It suggests that such scales might not be predictable by regression methods, and thus, any estimation of such scores must be considered with caution. However, it was suggested that BAYLEY-III motor scales, for the considered follow-ups, tend to underestimate later impairment rates [52].

This study is one of the first studies, which considers neurodevelopmental scores, regression analysis, and Sepsis. Similarly, Alotaibi et al. [20] proposed a regression model to predict BAYLEY-III scores on newborns with hypoxic-ischemic encephalopathy. They obtained an MAE of 12.07 on cognitive scores, confirming that the regression analysis could be a valuable support in the neurodevelopment assessment. Furthermore, several studies confirmed that HRV features could predict neurodevelopment-related scales [22], [24], [53], although most of them considered only univariate analysis or did not consider newborns with Sepsis.

Since the presented work is a retrospective study, it has some limitations: ECGs were recorded with a sampling frequency of 128Hz, which may be considered almost as the frequency limit for a reliable assessment of HRV parameters [54]. Therefore, future studies could acquire ECG signals with higher sampling rates (es. 500-1000Hz). Moreover, it is well known that HRV parameters may be influenced by several factors such as sleep stage, infant's position and the NICU environment, time after feeding, as well [55]. Therefore, future studies could include these factors to investigate their impact on HRV features and indirectly on predicting neurodevelopment outcomes.

Another limitation of the current study was the visual quality assessment of the ECG recordings and the R peaks

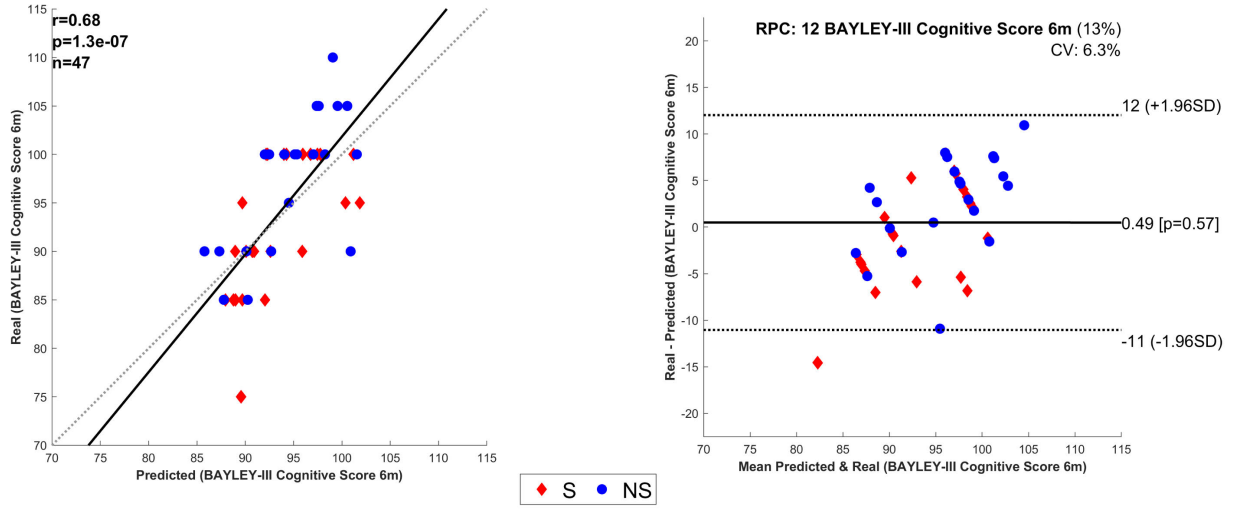


FIGURE 4. On the left: the correlation plot for the BAYLEY-III Cognitive score at 6 months. r is the Pearson correlation coefficient, p is the related p-value, and n is the number of observations. On the right is the corresponding Bland-Altam plot. The dotted line represents the limits of agreement. RPC is the reproducibility coefficient, SD is the standard deviation, and CV is the coefficient of variation. S=Sepsis, NS=No-Sepsis.

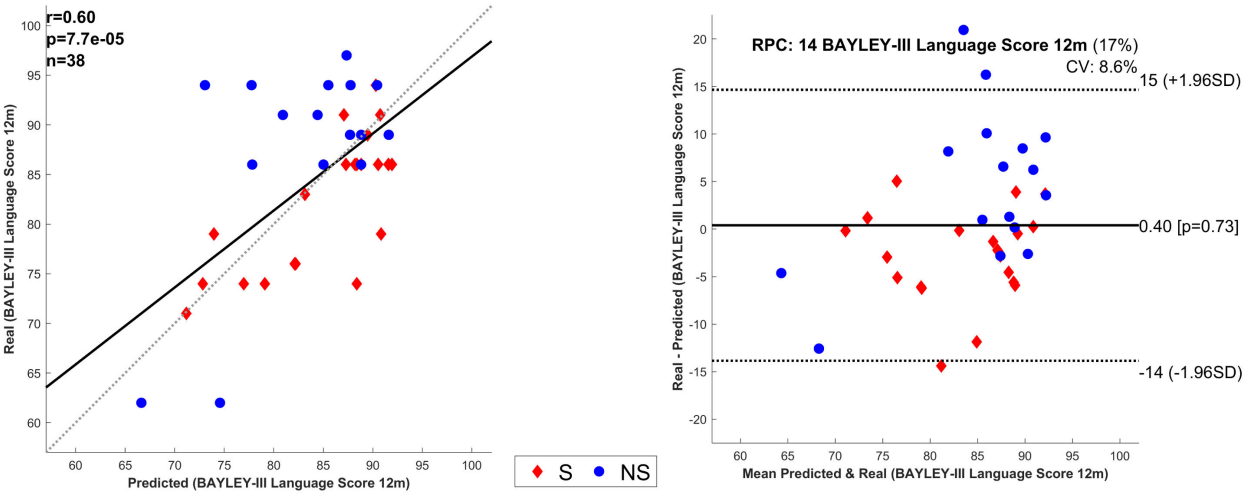


FIGURE 5. On the left: the correlation plot for the BAYLEY-III Language score at 12 months. r is the Pearson correlation coefficient, p is the related p-value, and n is the number of observations. On the right is the corresponding Bland-Altam plot. The dotted line represents the limits of agreement. RPC is the reproducibility coefficient, SD is the standard deviation, and CV is the coefficient of variation. S=Sepsis, NS=No-Sepsis.

detection before the HRV analysis. Future development of the current methodologies could include automatic tools for recognising contaminated windows and their correction in the pipeline before the HRV feature extraction. To the best of our knowledge, no studies on neonatal applications exist on this topic. However, previous works on adults have been investigated [56], and may be considered preliminary evaluations.

Furthermore, a sleep stage classifier may be included to split the observations based on belonging to a specific sleep phase [57]. Mainly, it requires EEG signals that are not part of the aims of the presented work. Sleep staging in newborns using HRV features instead of EEG may represent a remarkable development in this field. The analysis of newborns with Sepsis and the prediction of their BAYLEY-III scores by AI models required the analysis of almost 80 HRV features previously used in other neonatal applications without considering the statistical descriptors expansion [58], [59], [60], [61], [62].

Considering the models shown in Table 3, it appears that bagging models perform better on most of the BAYLEY-III scores than others. This might indicate that the data or the weak learners trained exhibited a high variance and low bias. However, this cannot be considered general because, for some BAYLEY-III scales (Motor 6m and Language 12m), LsBoost was the best choice, while SVR obtained the highest

TABLE 4. HRV features found statistically significant between S and NS groups. iqr=interquartile range, std=standard deviation (level of significance 0.05). (*) denotes the features with a p-value lower than 0.01.

Feature	Statistics Descriptors (Scale)
MFE	std(20)*, mean(2)*, skewness(11)*, std(9)*, mean(11), skewness(20), kurtosis(2), median (14), kurtosis (3), kurtosis (1), std (10), std(15), median (15).
MSE	kurtosis(2)*, skewness(17), iqr(2), skewness (10), std(20), iqr(1), mean(4).
MDE	skewness(20)*, kurtosis(3)*, median(3)*, mean(2), iqr(3), kurtosis (11), mean(3), mean (17), iqr(1), std(20), iqr(14), kurtosis (4), median (4).
Bubble Entropy	std.
CI MDE	std*, mean, kurtosis.
CI MSE	std, skewness, kurtosis.
CI MFE	std
pLF	median, skewness, std, mean

performance for 12-month follow-up on Cognitive scores. Therefore, it is not possible to assess that a model may perform better than others in general, varying according to the BAYLEY-III scale evaluated and the dataset.

However, future studies should evaluate other models to confirm these results, such as Gaussian Processes or Generalized Linear Models [61]. Moreover, FSR and ReliefF were evaluated as feature selection methods here, but other methodologies could be considered in future studies, such as LASSO or mRMR [61]. HRV multiscale entropy features were included in the validated models (MSE, MFE, and MDE at different scale levels). These findings confirmed that multiscale entropy indexes for HRV analysis could add helpful information in evaluating ANS activity in newborns [63]. HRV entropy features were already found to be linked to altered cardiovascular dynamics in newborns [31]. Furthermore, for the 12-month language scales, the low-frequency feature (LF, statistics descriptor: mean) was also included in the HRV-based regression model. In previous studies [55], LF changes were associated with ANS development. For clarity, it is well known that GA and the age of newborns at the time of ECG recording can impact HRV measures. However, as shown in Table 1, no statistical differences in terms of age between the S and NS groups were present. Therefore, future studies should check such differences in their cohort to verify possible effects on HRV parameters.

Regarding the statistical differences between S and NS on HRV parameters reported in Table 4, previous studies already showed possible impairments on ANS for infants with sepsis [64]. However, they considered only time- and frequency-domain features on a limited number of subjects (4 septic and 6 non-septic). Furthermore, the work presented here also evaluated HRV entropy-domain features on newborns with Sepsis, obtaining significant differences for most of them (as shown in Table 4). Recently, multi-scale entropy features were found helpful in characterizing ANS in newborns and their neurodevelopment [65], [66]. Moreover, in the literature, several works investigated if HRV features could predict the sepsis onset [67] or characterizing differences close to the sepsis event [68]. However, a direct comparison should be carefully made because our analysis was performed after the sepsis event and not directly related to it. In other words, our findings suggested that differences between S and NS cases regarding HRV values might exist after the sepsis event, and dedicated models could be investigated to improve prediction (one for S and another for NS cases). Unfortunately, the limited number of subjects in this work made it impossible to evaluate them. Moreover, our retrospective analysis did not have access to the exact time of the onset of the sepsis event. Therefore, we did not consider the delay between sepsis diagnosis and the time of the ECG recording as a possible confounding effect. Furthermore, we cannot evaluate if such differences may have disappeared during the infants' development. The only available information was that at 6 months and 12 months, infants with Sepsis obtained significantly lower scores on BAYLEY-III scales (as reported in Table 2). All these limits may explain the partial agreement between predicted and real BAYLEY-III outcomes as illustrated in the Bland-Altam plots in Figures 4 and 5. Therefore, future longitudinal ECG/HRV studies are needed to monitor the possible effects of Sepsis and determine if the infants may recover from it. It should be necessary to confirm that Sepsis in preterm infants may induce changes in the autonomic control of the heart, which can be measured at term equivalent age, which persist and are associated with neurodevelopmental outcomes.

To the best of our knowledge, there are currently no results on this specific topic in the literature. Thus, an optimal threshold for errors of regression models should be carefully defined to be reliably valuable in clinical practice. As the sensitivity of the BAYLEY-III scales is close to 5 points for our cohort, it is reasonable to consider a model promising when it provides an MAE (or MAE1/MAE0) value < 5.

Furthermore, most of the newborns considered in our study had BAYLEY-III scores higher than the threshold usually considered acceptable to exclude possible neurocognitive impairments (i.e., $\geq$ 90, Table 1 and [18]). Therefore, our methods should be evaluated on a more extensive set of infants with severe impairments, with BAYLEY-III scores <75. Significant differences for BAYLEY-III scores exist between S and NS groups for the cognitive scale at 6 months, as well as for the language scale at 12 months (Table 2), and these differences are detectable by the proposed

regression models. However, it is impossible to confirm whether the differences are due to preterm birth [21], Sepsis, or their combination. However, even if the variations were due only to preterm birth, the models could predict such alterations, allowing early detection of newborns with low BAYLEY-III scores at 6- and 12- months. Further studies are required to confirm that Sepsis affects the neurodevelopment of the involved infants and to evaluate if Sepsis might produce irreversible brain damage. In fact, it is well known that not all factors that contribute to neurodevelopmental outcomes are due to the cause of stay in the NICU. Although in the literature, several studies have confirmed that Sepsis may significantly alter neurodevelopment and proper brain functioning [69]. However, only a few works evaluate which physiological mechanisms might cause such an effect. Shah et al. [70] found that Sepsis may alter the white matter of newborns and that it may produce long-term effects. Thus, our results confirm that Sepsis may be active in neurodevelopment. However, further analysis is required to assess the localization or damaged entities in the Central and Autonomic Nervous Systems. Regression models must also be validated on the BAYLEY-III scores relative to 18-24 months of age to assess if and how HRV features might be able to predict scores of these scales and, thus, different periods of the infant's neurodevelopment. The proposed approach seems promising in predicting scores on the cognitive scale at 6 months and the language scale at 12 months. Neurodevelopmental tests, such as BAYLEY-III, are often operator-dependent and present several intrinsic sources of variability, limits, and pitfalls and tend to overestimate the infant's development [71]. Thus, obtaining an almost perfect correspondence between models' predictions and ground-truth scores might be challenging. A possible solution may be the combination of different evaluations made by several experts on the same subject. This analysis will be considered in future research.

V. CONCLUSION

This work exploits AI regression models to predict the neurodevelopmental scores of preterm newborns with and without Sepsis. The BAYLEY-III test computed the scores in three domains: cognitive, language, and motor. The quantitative HRV analysis was performed on ECG recordings recorded when the preterm infants reached the term age. Results are encouraging, giving an MAE lower than 5 points for the BAYLEY-III cognitive and language scales at 6- and 12-months. Our results confirmed that ANS development may be linked to CNS maturation and that HRV features could predict alterations affecting the correct neurodevelopment of newborns. If our results are confirmed, HRV-based regression models might also be used to support the clinical staff in the early detection of preterm newborns at risk of neurodevelopmental delays. Such models could be integrated into the clinical assessment as a pre-screening test before administering neurodevelopmental tests planned at different follow-up periods.

ACKNOWLEDGMENT

This work was carried out under the Ph.D. Course in Genetics, Oncology, and Clinical Medicine, GenOMeC, University of Siena (Italy).

REFERENCES

[1] J. J. González, S. Mañas, L. De Vera, L. D. Méndez, S. López, J. M. Garrido, and E. Pereda, "Assessment of electroencephalographic functional connectivity in term and preterm neonates," *Clin. Neurophysiol.*, vol. 122, no. 4, pp. 696–702, Apr. 2011.

[2] S. Beck, D. Wojdyla, L. Say, A. P. Bertran, M. Meraldi, J. H. Requejo, C. Rubens, R. Menon, and P. Van Look, "The worldwide incidence of preterm birth: A systematic review of maternal mortality and morbidity," *Bull. World Health Org.*, vol. 88, no. 1, pp. 31–38, Jan. 2010.

[3] 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," *Lancet*, vol. 379, no. 9832, pp. 2162–2172, Jun. 2012.

[4] N. Marlow, D. Wolke, M. A. Bracewell, and M. Samara, "Neurologic and developmental disability at six years of age after extremely preterm birth," *New England J. Med.*, vol. 352, no. 1, pp. 9–19, 2005.

[5] J. E. Lawn, H. Blencowe, S. Oza, D. You, A. C. Lee, P. Waiswa, M. Lalli, Z. Bhutta, A. J. Barros, P. Christian, C. Mathers, and S. N. Cousens, "Every newborn: Progress, priorities, and potential beyond survival," *Lancet*, vol. 384, no. 9938, pp. 189–205, Jul. 2014.

[6] *Born Too Soon: The Global Action Report on Preterm Birth*, World Health Org., Geneva, Switzerland, 2012.

[7] A. Zea-Vera and T. J. Ochoa, "Challenges in the diagnosis and management of neonatal sepsis," *J. Tropical Pediatrics*, vol. 61, no. 1, pp. 1–13, Feb. 2015.

[8] A. R. Rather and B. Kasana, "The third international consensus definitions for sepsis and septic shock (Sepsis-3)," *JMS SKIMS*, vol. 18, no. 2, pp. 162–164, Dec. 2015.

[9] M. Ershad, A. Mostafa, M. Dela Cruz, and D. Vearrier, "Neonatal sepsis," *Current Emergency Hospital Med. Rep.*, vol. 7, no. 3, pp. 83–90, 2019.

[10] E. J. Molloy, J. L. Wynn, J. Bliss, J. M. Koenig, F. M. Keij, M. McGovern, H. Kuester, M. A. Turner, E. Giannoni, J. Mazela, M. Degtyareva, T. Strunk, S. H. P. Simons, J. Janota, F. B. Plotz, A. van den Hoogen, W. de Boode, L. J. Schlapbach, and I. K. M. Reiss, "Neonatal sepsis: Need for consensus definition, collaboration and core outcomes," *Pediatric Res.*, vol. 88, no. 1, pp. 2–4, Jul. 2020.

[11] L. J. Schlapbach et al., "International consensus criteria for pediatric sepsis and septic shock," *JAMA*, vol. 331, no. 8, pp. 665–674, 2024.

[12] E. H. Chung, J. Chou, and K. A. Brown, "Neurodevelopmental outcomes of preterm infants: A recent literature review," *Translational Pediatrics*, vol. 9, no. S1, pp. S3–S8, Feb. 2020.

[13] S. Mukhopadhyay, K. M. Puopolo, N. I. Hansen, S. A. Lorch, S. B. DeMauro, R. G. Greenberg, C. M. Cotten, P. J. Sánchez, E. F. Bell, E. C. Eichenwald, and B. J. Stoll, "Impact of early-onset sepsis and antibiotic use on death or survival with neurodevelopmental impairment at 2 years of age among extremely preterm infants," *J. Pediatrics*, vol. 221, pp. 39–46.e5, Jun. 2020.

[14] B. Alshaikh, K. Yusuf, and R. Sauve, "Neurodevelopmental outcomes of very low birth weight infants with neonatal sepsis: Systematic review and meta-analysis," *J. Perinatol.*, vol. 33, no. 7, pp. 558–564, Jul. 2013.

[15] T. Ortgies, M. Rullmann, D. Ziegelhöfer, A. Bläser, and U. H. Thome, "The role of early-onset-sepsis in the neurodevelopment of very low birth weight infants," *BMC Pediatrics*, vol. 21, no. 1, p. 289, Dec. 2021.

[16] J. H. Pek, B. J. Yap, M. Y. Gan, S. T. T. Seethor, R. Greenberg, C. P. V. Hornik, B. Tan, J. H. Lee, and S.-L. Chong, "Neurocognitive impairment after neonatal sepsis: Protocol for a systematic review and meta-analysis," *BMJ Open*, vol. 10, no. 6, Jun. 2020, Art. no. e038816.

[17] M. S. Scher, "'The first thousand days' define a fetal/neonatal neurology program," *Frontiers Pediatrics*, vol. 9, Aug. 2021, Art. no. 683138.

[18] C. Del Rosario, M. Slevin, E. J. Molloy, J. Quigley, and E. Nixon, "How to use the Bayley scales of infant and toddler development," *Arch. disease childhood Educ. Pract. Ed.*, vol. 106, no. 2, pp. 108–112, Apr. 2021.

[19] A. Temko, O. Doyle, D. Murray, G. Lightbody, G. Boylan, and W. Marnane, "Multimodal predictor of neurodevelopmental outcome in newborns with hypoxic-ischaemic encephalopathy," *Comput. Biol. Med.*, vol. 63, pp. 169–177, Aug. 2015.

[20] N. Alotaibi, D. Bakheet, D. Konn, B. Vollmer, and K. Maharatna, "Cognitive outcome prediction in infants with neonatal hypoxic-ischemic encephalopathy based on functional connectivity and complexity of the electroencephalography signal," *Frontiers Human Neurosci.*, vol. 15, Jan. 2022, Art. no. 795006.

[21] S. R. Yiallourou, N. B. Witcombe, S. A. Sands, A. M. Walker, and R. S. C. Horne, "The development of autonomic cardiovascular control is altered by preterm birth," *Early Human Develop.*, vol. 89, no. 3, pp. 145–152, Mar. 2013.

[22] G. Thiriez, C. Mougey, D. Vermeylen, V. Wermenbol, J.-P. Lanquart, J. S. Lin, and P. Franco, "Altered autonomic control in preterm newborns with impaired neurological outcomes," *Clin. Autonomic Res.*, vol. 25, no. 4, pp. 233–242, Aug. 2015.

[23] T. F. Ask, S. Ranjitkar, M. Ulak, R. K. Chandyo, M. Hysing, T. A. Strand, I. Kvestad, L. Shrestha, M. Andreassen, R. G. Lugo, J. S. Shilpakar, M. Shrestha, and S. Sütterlin, "The association between heart rate variability and neurocognitive and socio-emotional development in Nepalese infants," *Frontiers Neurosci.*, vol. 13, p. 411, Apr. 2019.

[24] R. M. Goulding, N. J. Stevenson, D. M. Murray, V. Livingstone, P. M. Filan, and G. B. Boylan, "Heart rate variability in hypoxic ischemic encephalopathy: Correlation with EEG grade and 2-y neurodevelopmental outcome," *Pediatric Res.*, vol. 77, no. 5, pp. 681–687, May 2015.

[25] M. Lucchini, W. P. Fifer, R. Sahni, and M. G. Signorini, "Novel heart rate parameters for the assessment of autonomic nervous system function in premature infants," *Physiological Meas.*, vol. 37, no. 9, pp. 1436–1446, Sep. 2016.

[26] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985.

[27] H. Sedghamiz. (2014). *MATLAB Implementation of Pan Tompkins ECG QRS Detector*. [Online]. Available: https://it.mathworks.com/matlabcentral/fileexchange/45840-complete-pan-tompkins-implementation-ecg-qrs-detector

[28] L. Frassineti, A. Lanatà, and C. Mandredi, "HRV analysis: A non-invasive approach to discriminate between newborns with and without seizures," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 52–55.

[29] B. Olmi, C. Manfredi, L. Frassineti, C. Dani, S. Lori, G. Bertini, C. Cossu, M. Bastianelli, S. Gabbanini, and A. Lanatà, "Heart rate variability analysis for seizure detection in neonatal intensive care units," *Bioengineering*, vol. 9, no. 4, p. 165, Apr. 2022.

[30] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers Public Health*, vol. 5, p. 258, Sep. 2017.

[31] L. Frassineti, A. Lanatà, B. Olmi, and C. Manfredi, "Multiscale entropy analysis of heart rate variability in neonatal patients with and without seizures," *Bioengineering*, vol. 8, no. 9, p. 122, Sep. 2021.

[32] G. Manis, M. Bodini, M. W. Rivolta, and R. Sassi, "A two-steps-ahead estimator for bubble entropy," *Entropy*, vol. 23, no. 6, p. 761, Jun. 2021.

[33] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of biological signals," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 71, no. 2, Feb. 2005, Art. no. 021906.

[34] M. Vollmer, "HRVTool—An open-source MATLAB toolbox for analyzing heart rate variability," in *Proc. Comput. Cardiol. (CinC)*, Sep. 2019, p. 1.

[35] S. Latremouille, J. Lam, W. Shalish, and G. Sant'Anna, "Neonatal heart rate variability: A contemporary scoping review of analysis methods and clinical applications," *BMJ Open*, vol. 11, no. 12, Dec. 2021, Art. no. e055209.

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

[37] F. E. Vaccari, S. Diomedi, M. Filippini, C. Galletti, and P. Fattori, "A Poisson generalized linear model application to disentangle the effects of various parameters on neurophysiological discharges," *STAR Protocols*, vol. 2, no. 2, Jun. 2021, Art. no. 100413.

[38] E. Avots, K. Jermakovs, M. Bachmann, L. Päeske, C. Ozcinar, and G. Anbarjafari, "Ensemble approach for detection of depression using EEG features," *Entropy*, vol. 24, no. 2, p. 211, Jan. 2022.

[39] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, nos. 1–2, pp. 23–69, 2003.

[40] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning* (Springer Series in Statistics). New York, NY, USA: Springer, 2001.

[41] M. Awad and R. Khanna, "Support vector regression," in *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*. Cham, Switzerland: Springer, 2015, pp. 67–80.

[42] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004.

[43] J. Bi and K. P. Bennett, "A geometric approach to support vector regression," *Neurocomputing*, vol. 55, nos. 1–2, pp. 79–108, Sep. 2003.

[44] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.

[45] J. Mendes-Moreira, C. Soares, A. M. Jorge, and J. F. D. Sousa, "Ensemble approaches for regression: A survey," *ACM Comput. Surv.*, vol. 45, no. 1, pp. 1–40, Nov. 2012.

[46] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.

[47] A. Temko, E. Thomas, W. Marnane, G. Lightbody, and G. B. Boylan, "Performance assessment for EEG-based neonatal seizure detectors," *Clin. Neurophysiol.*, vol. 122, no. 3, pp. 474–482, Mar. 2011.

[48] K. V. Bunting, R. P. Steeds, K. Slater, J. K. Rogers, G. V. Gkoutos, and D. Kotecha, "A practical guide to assess the reproducibility of echocardiographic measurements," *J. Amer. Soc. Echocardiography*, vol. 32, no. 12, pp. 1505–1515, Dec. 2019.

[49] D. Giavarina, "Understanding bland Altman analysis," *Biochemia Medica*, vol. 25, no. 2, pp. 141–151, 2015.

[50] B. Y. H. Wee, J. H. Lee, Y. H. Mok, and S.-L. Chong, "A narrative review of heart rate and variability in sepsis," *Ann. Translational Med.*, vol. 8, no. 12, p. 768, Jun. 2020.

[51] M. Carrara, M. Ferrario, B. Bollen Pinto, and A. Herpain, "The autonomic nervous system in septic shock and its role as a future therapeutic target: A narrative review," *Ann. Intensive Care*, vol. 11, no. 1, p. 80, Dec. 2021.

[52] A. J. Spittle, M. M. Spencer-Smith, J. L. Cheong, A. L. Eeles, K. J. Lee, P. J. Anderson, and L. W. Doyle, "General movements in very preterm children and neurodevelopment at 2 and 4 years," *Pediatrics*, vol. 132, no. 2, pp. e452–e458, 2013.

[53] C. León, S. Cabon, H. Patural, G. Gascoin, C. Flamant, J.-M. Roué, G. Favrais, A. Beuchée, P. Pladys, and G. Carrault, "Evaluation of maturation in preterm infants through an ensemble machine learning algorithm using physiological signals," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 1, pp. 400–410, Jan. 2022.

[54] S. Laborde, E. Mosley, and J. F. Thayer, "Heart rate variability and cardiac vagal tone in psychophysiological research—Recommendations for experiment planning, data analysis, and data reporting," *Frontiers Psychol.*, vol. 8, p. 213, Feb. 2017.

[55] 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 Neurosci.*, vol. 14, p. 999, Sep. 2020.

[56] L. Citi, E. N. Brown, and R. Barbieri, "A real-time automated point-process method for the detection and correction of erroneous and ectopic heartbeats," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 10, pp. 2828–2837, Oct. 2012.

[57] K. Pillay, A. Dereymaeker, K. Jansen, G. Naulaers, S. Van Huffel, and M. De Vos, "Automated EEG sleep staging in the term-age baby using a generative modelling approach," *J. Neural Eng.*, vol. 15, no. 3, Jun. 2018, Art. no. 036004.

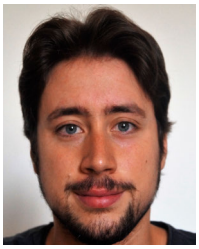
[58] A. H. T. Kong, M. M. Lai, S. Finnigan, R. S. Ware, R. N. Boyd, and P. B. Colditz, "Background EEG features and prediction of cognitive outcomes in very preterm infants: A systematic review," *Early Human Develop.*, vol. 127, pp. 74–84, Dec. 2018.

[59] O. De Wel, M. Lavanga, A. Dorado, K. Jansen, A. Dereymaeker, G. Naulaers, and S. Van Huffel, "Complexity analysis of neonatal EEG using multiscale entropy: Applications in brain maturation and sleep stage classification," *Entropy*, vol. 19, no. 10, p. 516, Sep. 2017.

[60] J. M. O'Toole, G. B. Boylan, R. O. Lloyd, R. M. Goulding, S. Vanhatalo, and N. J. Stevenson, "Detecting bursts in the EEG of very and extremely premature infants using a multi-feature approach," *Med. Eng. Phys.*, vol. 45, pp. 42–50, Jul. 2017.

[61] L. Frassineti, C. Manfredi, B. Olmi, and A. Lanatà, "A generalized linear model for an ECG-based neonatal seizure detector," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 471–474.

[62] L. Frassineti, A. Parente, and C. Manfredi, "Multiparametric EEG analysis of brain network dynamics during neonatal seizures," *J. Neurosci. Methods*, vol. 348, Jan. 2021, Art. no. 109003.
[63] R. Statello, L. Carnevali, A. Sgoifo, M. Miragoli, and F. Pisani, "Heart rate variability in neonatal seizures: Investigation and implications for management," *Neurophysiologie Clinique*, vol. 51, no. 6, pp. 483–492, Dec. 2021.
[64] A. A. Mrazek, M. T. Shabana, G. L. Radhakrishnan, G. C. Kramer, and R. S. Radhakrishnan, "Heart rate variability analysis is more sensitive at identifying neonatal sepsis than conventional vital signs," *J. Surgical Res.*, vol. 186, no. 2, p. 507, Feb. 2014.
[65] M. Lavanga, B. Bollen, A. Caicedo, A. Dereymaeker, K. Jansen, E. Ortibus, S. Van Huffel, and G. Naulaers, "The effect of early procedural pain in preterm infants on the maturation of electroencephalogram and heart rate variability," *Pain*, vol. 162, no. 5, pp. 1556–1566, 2021.
[66] M. Lavanga, E. Heremans, J. Moeyersons, B. Bollen, K. Jansen, E. Ortibus, G. Naulaers, S. Van Huffel, and A. Caicedo, "Maturation of the autonomic nervous system in premature infants: Estimating development based on heart-rate variability analysis," *Frontiers Physiol.*, vol. 11, Jan. 2021, Art. no. 581250.
[67] 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 J. Biomed. Health Informat.*, vol. 24, no. 3, pp. 681–692, Mar. 2020.
[68] B. A. Sullivan and K. D. Fairchild, "Vital signs as physiomarkers of neonatal sepsis," *Pediatric Res.*, vol. 91, no. 2, pp. 273–282, Jan. 2022.
[69] D. Rallis, P. Karagianni, E. Goutsiou, V. Soubasi-Griva, J. Banerjee, and C. Tsakalidis, "The association of the cerebral oxygenation during neonatal sepsis with the Bayley-III scale of infant and toddler development index scores at 18–24 months of age," *Early Human Develop.*, vol. 136, pp. 49–53, Sep. 2019.
[70] D. K. Shah, L. W. Doyle, P. J. Anderson, M. Bear, A. J. Daley, R. W. Hunt, and T. E. Inder, "Adverse neurodevelopment in preterm infants with postnatal sepsis or necrotizing enterocolitis is mediated by white matter abnormalities on magnetic resonance imaging at term," *J. Pediatrics*, vol. 153, no. 2, pp. 170–175, Aug. 2008.
[71] P. J. Anderson and A. Burnett, "Assessing developmental delay in early childhood—Concerns with the Bayley-III scales," *Clin. Neuropsychologist*, vol. 31, no. 2, pp. 371–381, Feb. 2017.



LORENZO FRASSINETI (Member, IEEE) received the B.S. degree in electronic and telecommunications engineering and the M.S. degree in biomedical engineering from the University of Florence, Italy, in 2015 and 2018, respectively, and the Ph.D. degree in genetics, oncology, and clinical medicine from the University of Siena, Italy, in 2023. His research interests include machine learning algorithms for biomedical applications, automatic detection of neurological disorders, paediatric bioengineering applications, and high-level signal processing. In 2023, he received the Research Grant from the University of Florence regarding the development of artificial models and algorithms for characterizing and recognizing the level of psychophysiological well-being in fragile subjects. He is an Associate Editor of *Biomedical Signal Processing and Control* (Elsevier).

GIOVANNA BERTINI was born in May 1966. She received the degrees (Hons.) in medicine and surgery, and specialization in pediatrics, in 1992 and 1997, respectively. She was a Neonatologist with a highly professional position at a university hospital in Florence, Italy. Her research interests include pediatric and neonatal neurophysiological field.

FEDERICO CALÀ (Member, IEEE) received the B.S. degree in mechanical engineering and the M.S. degree in biomedical engineering from the Università degli Studi di Firenze, Florence, Italy, in 2019 and 2022, respectively, where he is currently pursuing the Ph.D. degree in information engineering (curriculum: control, optimization, and complex systems). From 2021 to 2022, he was a Grant Holder with the Biomedical Engineering Laboratory, Università degli Studi di Firenze. His research interests include developing machine learning models to support the differential diagnosis of genetic syndromes and voice disorders through quantitative acoustical analysis. He won the 13th MAVEBA International Workshop Best Student Paper Award in 2023 from the Bioengineering Journal.

CARLO DANI was born in December 1963. He received the degree in medicine and surgery, in 1988, and the specialization degree (Hons.) in pediatrics and neonatology, in 1992 and 1995, respectively. He is currently a Full Professor of pediatrics with the Department of Surgical and Medical Critical Care, Section of Neonatology, Careggi University Hospital, Florence, Italy.

ANGELA PARENTE received the B.S. degree in electronic and telecommunications engineering and the M.S. degree in biomedical engineering from the Università degli Studi di Firenze, Florence, Italy, in 2019 and 2022, respectively. She is currently a Software Quality Assurance at a digital health company. She won the National Bioengineering Group's Laboratory of Biosignals, Bioimaging and Bioinformatics B3Lab Polytechnic University of Milan Graduation Award in 2023.

CARLOS ALBERTO REYES GARCÍA (Senior Member, IEEE) was born in Lagos de Moreno, Jalisco, Mexico. He received the Philosophy Doctor (Ph.D.) degree in computer science (artificial intelligence) from Florida State University, Tallahassee, FL, USA, in April 1994. He is currently a Full Time Researcher Level C of computer science with the Instituto Nacional de Astrofísicaa, Óptica y Electrónica (INAOE), Puebla, Mexico. His research interests include intelligent systems, fuzzy sets and logic, fuzzy neural networks, genetic algorithms, machine learning, knowledge representation, automatic speech and speaker recognition, and general pattern classification.

CLAUDIA MANFREDI has been an Associate Professor of bioengineering with National Scientific Qualification as a Full Professor with the Department of Information Engineering (DINFO), Università degli Studi di Firenze, Florence, Italy. She is the author and co-author of more than 200 articles and has Scopus H-index: 23. Her professional experience concerns the study of the human voice, the EEG signal, facial expressions, movements, and other physiological signals, with application to newborns and adults for detection and classification of neurological diseases. She is a member of the Editorial Board of *Biomedical Signal Processing and Control* (Elsevier Ltd.) and *Bioengineering* (MDPI). Since 1999, she organizes the series of biennial conferences: "International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications" (MAVEBA), Florence, sponsored by: IEEE-BME, Ente Cassa di Risparmio di Firenze, ISCA, AIIMB, INFM, AISV, and Elsevier, and she was an Editor of the Proceedings (Firenze University Press) and the Guest Editor of Special Issues on the journals: *Acta Acustica* (European Acoustics Association), *Medical Engineering and Physics* (Elsevier Ltd.), *Biomedical Signal Processing and Control* (Elsevier Ltd.), and *Bioengineering* (MDPI).

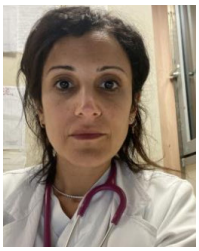
SILVIA LORI received the degree (Hons.) in medicine and surgery, with a specialization in neurology, and neurosurgery, in 1985, 1989, and 2004, respectively. She is a Doctor with a highly professional position at a university hospital in Florence, Italy. Her research interests include the pediatric and neonatal neurophysiological field.

SIMONETTA GABBANINI received the degree in health professions of technical sciences (diagnostic). She is currently a Neurophysiopathology Technician at public hospital. She is a SSN Professor with the Università degli Studi di Firenze, Florence, Italy. Her research interests include the neurophysiological instrumental diagnostics in neonatal intensive care units.

CLARA LUNARDI was born in March 1979. She received the degree in developmental psychology, with a specialization in relational systemic psychotherapy. She is a referent for the neuropsychological assessment in the high risk neonate and children follow up program. Her research interests include pediatric and neonatal neuropsychology.

CESARINA COSSU was born in February 1965. She received the degree in health professions of technical sciences (diagnostic). She is currently a Neurophysiopathology Technician in public hospital. Her research interests include the multimodal neurophysiological monitoring and diagnostics in infants.

MARIA BASTIANELLI received the degree in health professions of technical sciences (diagnostic). She was a Neurophysiopathology Technician in public hospital. She has been a SSN Professor with the Università degli Studi di Firenze, Florence, Italy. Her research interests include the analysis of evoked potentials in newborns and the neurophysiological assessment in the neonatal intensive care units.



CATERINA COVIELLO received the degree in medicine and surgery from the Università degli Studi di Siena, Siena, Italy, in 2007, with a specialization in pediatrics and neonatology. She is currently a full-time Staff Specialist Neonatologist with the Division of Neonatology, Careggi University Hospital, Florence, Italy.

ANTONIO LANATÀ (Member, IEEE) received the Ph.D. degree. He is currently an Associate Professor of bioengineering with the Department of Information Engineering, University of Florence, Italy. He is also the Head of the Biosignal and Bioinstrumentation Laboratory, University of Florence, and responsible for the Computational Physiology and Biomedical Systems (ComPBioS) Research Group. He is the author of more than 130 international scientific contributions. His research interests include designing and implementing wearable systems for physiological monitoring and statistical and nonlinear biomedical signal processing. Applications of his research include the assessment of the autonomic nervous system on affective computing, assessment of mood and mental/neurological disorders, and human/animal/robot interaction. He is a member of the IEEE Circuits and Systems Society, Engineering in Medicine and Biology, Information Theory, and Signal Processing Society. He is an Associate Editor of several international journals, such as IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, *Frontiers Bioelectronics*, *Bioengineering* (MDPI), *Bioelectronics*, *Biosensors*, *Algorithms*, *Electronics*, and *Animals*.

Open Access funding provided by 'Università degli Studi di Firenze' within the CRUI CARE Agreement