Data in brief 29 (2020) 105164

Contents lists available at ScienceDirect

Data in brief

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

Data Article

Dataset on linear and non-linear indices for discriminating healthy and IUGR fetuses

Maria G. Signorini [a, *], Nicolò Pini [a], Alberto Malovini [b], Riccardo Bellazzi [c], Giovanni Magenes [c]

[a] Department of Electronics, Information and Bioengineering (DEIB), Politecnico Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy
[b] IRCCS Fondazione S. Maugeri, Via Maugeri 10, 27100 Pavia, Italy
[c] Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Via Ferrata 5, 27100 Pavia, Italy

ARTICLE INFO

Article history:
Received 21 December 2019
Received in revised form 13 January 2020
Accepted 14 January 2020
Available online 29 January 2020

Keywords:
Fetal heart rate monitoring
Cardiotocography
Intra uterine growth restricted
Physiology-based features
Multivariate analysis
Predictive analytics

ABSTRACT

The presented collection of data comprises of a set of 12 linear and nonlinear indices computed at different time scales and extracted from Fetal Heart Rate (FHR) traces acquired through Hewlett Packard CTG fetal monitors (series 1351A), connected to a PC. The sampling frequency of the recorded FHR signal is equal 2 Hz. The recorded populations consist of two groups of fetuses: 60 healthy and 60 Intra Uterine Growth Restricted (IUGR) fetuses. IUGR condition is a fetal condition defined as the abnormal rate of fetal growth. In clinical practice, diagnosis is confirmed at birth and may only be suspected during pregnancy. The pathology is a documented cause of fetal and neonatal morbidity and mortality. The described database was employed in a set of machine learning approaches for the early detection of the IUGR condition: "Integrating machine learning techniques and physiology based heart rate features for antepartum fetal monitoring" [1]. The added value of the proposed indices is their interpretability and close connection to physiological and pathological aspect of FHR regulation. Additional information on data acquisition, feature extraction and potential relevance in clinical practice are discussed in [1].

DOI of original article: https://doi.org/10.1016/j.cmpb.2019.105015.
* Corresponding author.
E-mail address: mariagabriella.signorini@polimi.it (M.G. Signorini).

https://doi.org/10.1016/j.dib.2020.105164

Specifications Table

Subject	Medicine - Obstetrics, Gynecology and Women's Health
Specific subject area	Perinatal Medicine, Fetal Monitoring
Type of data	Table of Fetal Heart Rate (FHR) parameters
How data were acquired	A set of linear and nonlinear algorithms were computed by means of algorithms implemented in Matlab acquired using Hewlett Packard CTG fetal monitors (series 1351A), connected to a PC
Data format	Raw table of FHR indices
Parameters for data collection	60 normals and 60 IUGR CTG recordings, each lasting > 30 minutes were collected on pregnant women, at different weeks of gestation. Parameters were extracted by means of algorithms reported in the scientific literature, considering non overlapping time windows of different duration, depending on the parameter. Each index reported in the table represents the average of the relevant parameter computed over the whole recording.
Description of data collection	CTG traces were recorded using Hewlett Packard CTG fetal monitors (series 1351A), connected to a PC during standard non-stress tests on pregnant women in a clinical environment under supervision of trained personnel. FHR was sampled at 2 Hz (one value each 500 ms)
Data source location	Department of Obstetrical-Gynaecological and Urological Science and Reproductive Medicine, Federico II University, Naples, Italy
Data accessibility	Repository name: Mendeley Data Data identification number: 10.17632/2953f8fgcy.1 Direct URL to data: https://doi.org/10.17632/2953f8fgcy.1#file-693e38db-a819-4b36-960e-446634ce45e5
Related research article	Signorini, M. G., Pini, N., Malovini, A., Bellazzi, R., & Magenes, G. (2020). Integrating machine learning techniques and physiology based heart rate features for antepartum fetal monitoring. *Computer methods and programs in biomedicine, 185*, 105015. https://doi.org/10.1016/j.cmpb.2019.105015

Value of the Data

- The selected indices are representative of time domain-frequency linear analyses and of nonlinear methods described in the literature for the investigation of the neurological mechanisms controlling the fetal heart rate in physiological and pathophysiological conditions.
- Data can be used for a multiparametric comparison of IUGR and normal FHR in the prenatal period in order to predict adverse outcomes at the labor and later developmental outcomes.
- The table can be used by the scientific community for developing reliable classifiers based on a multi parametric approach.
- As a further development, it would be possible to compute additional indices given the availability (upon formal request) of the raw FHR signals for both the populations.

1. Data description

The data enclosed with this Data in Brief article contain a table of 12 Fetal Heart Rate indices extracted from CTG tracings collected from two equally numerous populations of fetuses during the antepartum period: 60 healthy fetuses and 60 Intra Uterine Growth Restricted (IUGR) fetuses [1].

The 120 CTG tracings were recorded at Department of Obstetrical-Gynaecological and Urological Science and Reproductive Medicine, Federico II University, Naples, Italy. These tracings cannot be made publicly available due to privacy and ethical reasons. The indices contained in the data table, were computed on CTG tracings as described in Experimental Design, Materials, and Methods section below. Each datum reported in the table represents the average of the relevant index for the whole duration of the recording. This condensed information represents the raw data that we employed in the related research article for testing several machine learning methods with the purpose of discriminating healthy versus IUGR fetuses. This dataset is fully available as reported in Data accessibility section.

Coming to the description of the available dataset, it comprises of two different excel sheets which contain the parameters of the two populations.

The data are organized as follows:

1. Each row corresponds to a single participant (fetus) and each column corresponds to a given index.
2. The first column after the numerical id of the patient (second column in the sheet) contains the week of gestation at which the CTG was recorded.
3. Each cell contains the average value of the index (column) for each patient (row) computed on the whole FHR tracing. The averaging procedure was performed since the total length of each recording contains both activity and quiet periods of the relevant fetus.

Missing values are coded as NaN.

2. Experimental Design, materials, and methods

The presented Supplementary Material provides a more in-depth description of the steps towards the estimation of parameters employed in this study starting from the FHR signal. The original raw FHR recordings were sampled at 2 Hz and they were measured in beats per minutes (bmp).

1.1. Time domain parameters

Each FHR value coming from Hewlett Packard CTG fetal monitors (series 1351A) was transformed in equivalent RR interval and expressed in milliseconds for the computation of time domain parameters.

***Short Term Variability (STV)* [ms]**: it quantifies FHR variability on a short time scale. Considering an interbeat sequence of 1-minute duration, STV is defined as:

$$STV = \text{mean}[|T(i+1) - T(i)|]_i = \frac{\sum_{i=1}^{23} |T(i+1) - T(i)|}{23}$$

where T is the average FHR computed by diving the FHR recording in nonoverlapping windows of 5 consecutive FHR values (2.5 s for a sampling frequency of 2 Hz). *STV* is computed in a window of 1-minute duration so that 24 T values are obtained for each window. The corresponding STV estimate is obtained by averaging the differences between adjacent $T(i)$ values, having accelerations and deceleration excluded [2].

Interval Index (II): it provides an estimated of short term variability scaled by STV, defined as:

$$II = \frac{\text{std}[|T(i+1) - T(i)|]_i}{STV}$$

where$i = 1, \ldots, 23$ are the total number of T FHR values recorded in 1 minute [2].

***Delta* [ms]**: considering a window in time of 1-minute duration, Delta is defined as:

$$Delta = \max(T(i)) - \min(T(i))$$

where $i = 1, \ldots, 24$ which are the total number of T FHR values recorded in 1 minute [2].

***Long Term Irregularity* (*LTI*) [ms]**: is defined as the interquartile range of the distribution $m(j)$ which is defined as:

$$m(j) = \sqrt{T^2(j+1) + T^2(j)}$$

where $j = 1, \ldots, 71$ which are the total number of TFHR values in a 3-minute window. LTI quantifies FHR variability on a longer time scale with respect to the previously reported time domain indices excluding accelerations and decelerations [2].

1.2. Frequency domain parameters

Each FHR value coming from Hewlett Packard CTG fetal monitors (series 1351A) was transformed in equivalent RR interval and expressed in milliseconds for the computation of frequency domain parameters.

PSD Power Spectral Density (PSD) is a widely employed tool for HRV frequency analysis as it can quantitatively measure the periodic oscillations related to neural control activity, namely autonomic nervous system (ANS) modulation over the cardiac system.

In the context of this analysis, the PSD estimation for FHR was performed based on autoregressive (AR) modelling (parametric spectral estimation). The AR model utilized to mimic FHR dynamic is defined as:

$$\widehat{FHR}(j) = \sum_{i=1}^{p} a_i \cdot FHR(n-i) + w_j$$

where $w_j \sim WGN(0,\ \sigma^2)$(White Gaussian Noise), p is the model order (from 8 to 12), and a_i are the model parameters. The modelled FHR windows ($\widehat{FHR}(j)$) are of duration equal to 3 minutes ($j = 1, \ldots, 360$).

Model parameters are calculated recursively by means of the Levinson-Durbin algorithm. Once the proper model order is defined, so that the model parameters are determined, PSD is defined as:

$$PSD(f) = \frac{\sigma^2\Delta}{\left|1 - \sum_{k=1}^{p} a_k e^{-j2\pi kf\Delta}\right|^2} = \frac{\sigma^2\Delta}{A\left(e^{j2\pi f\Delta}\right)A^*\left(e^{j2\pi f\Delta}\right)} = \left.\frac{\sigma^2\Delta}{A(z)A^*\left(\frac{1}{z^*}\right)}\right|_{z=e^{j2\pi f\Delta}}$$

where Δ is the mean value of$\widehat{FHR}(j)$ in seconds and $A(z)$ is the z-transform of the transfer function of the AR process previously defined. Through this parametric approach, FHR signal undergoes an automatic decomposition into a sum of sinusoidal contributions, themselves identified by their corresponding central frequencies and the associated power [3].

In the context of frequency analysis applied to FHR it is possible to identify three specific frequency bands of interest, namely LF [0.03–0.15] Hz; MF [0.15–0.5] Hz; HF [0.5–1 Hz]. The corresponding associated power is computed by integrating the PSD power in these intervals, thus obtaining ***LF_pow* [ms^2/Hz]; *MF_pow* [ms^2/Hz]; *HF_pow* [ms^2/Hz].** Lastly, ***LF/(MF + HF)*** is obtained as: *LF_pow/(MF_pow + HF_pow)* [3].

1.3. Nonlinear domain parameters

Each FHR value coming from Hewlett Packard CTG fetal monitors (series 1351A) was transformed in equivalent RR interval and expressed in milliseconds for the computation of Approximate Entropy and Lempel and Ziv complexity.

***Approximate Entropy (ApEn)* [bits]**: quantifies a signal regularity by assessing the occurrence rate of patterns by comparing the patters themselves to a reference one of length m. Pattern similarity is defined based on a threshold r [4].

Given a sequence of N data points $u(i)$, $i = 1, \ldots, N$ the algorithm creates sequences $x_m(i)$ (based on window length m) and it computes for each $i \leq N - m + 1$ the quantity expressed as:

$$C_i^m(r) = \frac{1}{N}\{\text{count of } j \leq N - m + 1 \mid \text{distance } [x_m(i), x_m(j)] \leq r\}$$

Approximate Entropy (ApEn) is defined as:

$$ApEn(m,r) = \lim_{N\to\infty}\left[\Phi^m(r) - \Phi^{m+1}(r)\right]$$

where $\Phi^m(r) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1}\ln(C_i^m(r))$.

In the context of finite time series of length N as for FHR, ApEn can be written as:

$$ApEn(m,r,N) = \Phi^m(r) - \Phi^{m+1}(r)$$

In this work, ApEn was estimated by considering non-overlapping windows of duration equal to 3 minutes, with the following parameter setting: $m = 1, r = 0.1, N = 360$ and named ***ApEn(1,0.1)*** [5,6].

Lempel and Ziv complexity (LZC) **[bits]**: the first step towards its formulation encompasses the definition of the quantity$c(n)$ which measures the number of different sub strings and the rate of their recurrence in a given time series. According to the Information Theory, in turn it assesses the minimum quantity of information needed to define a binary string. LZC quantifies the rate of new patterns arising as signal evolves [7].

Suppose the number of symbols in the alphabet A is α and the length of sequence is equal n. The upper bound for $c(n)$ is given by:

$$c(n) < \frac{n}{(1-\varepsilon_n)\cdot\log(n)}$$

where

$$\varepsilon_n = 2\frac{1+\log(\log(\alpha n))}{\log(n)}$$

and $\log(x)$ means the logarithm of x to the base α.

When n is large enough $(n\to\infty)$, $\varepsilon_n \to 0$ so as a result:

$$b(n) = \lim_{n\to\infty} c(n) = \frac{n}{\log(n)}$$

where $b(n)$is the asymptotic behavior of $c(n)$ for a random string.

The normalized complexity is thus defined as:

$$C(n) = \frac{c(n)}{b(n)}$$

In order to obtain LZC estimation for FHR time series, the latter requires to be transformed into a symbolic sequence according to a binary and/or a ternary coding procedure.

Binary coding: given a FHR series $x(N)$, the sequence $y(N)$ is built by assigning 1 when the condition $x(n+1) > x(n)$ is verified for $n = 1,\ldots,N$. On the opposite case of signal decrease, $y(n)$ is assigned to 0 when the condition $x(n+1) \le x(n)$ is met.

Ternary coding: given a FHR series $x(N)$, the sequence $y(N)$ is built as in the binary coding case with the additional condition of signal invariance which is defined as $x(n+1) = x(n)$ and coded with the symbol 2 [7].

Additionally, in the context of recorded time series a factor p is introduced to define the minimum quantization level for a symbol change in the coded string (e.g. $y(n) = 1$ if $x(n+1) > x(n) + x(n)\cdot p$.

In this work, LZC was estimated by considering non-overlapping windows of duration equal to 3 minutes, with the following parameter setting: binary coding and $p = 0$ and named ***LZC(2,0)***. The choice of $p = 0$ reflects the current value for the quantization level, which is actually $\pm$ 0.5 bpm [8].

Each FHR value coming from Hewlett Packard CTG fetal monitors (series 1351A) was expressed in bpm for the computation of Acceleration Phase Rectified Slope and Deceleration Phase Rectified Slope

in order to be concordant with the common definition of acceleration and deceleration in fetal heart rate monitoring.

Acceleration Phase Rectified Slope (APRS) **[bpm]**: the computation of Phase Rectified Signal Averaging (PRSA) curve (which APRS is extracted from) starts from considering a time series x_i of length N ($i = 1,\ldots,N$) as FHR in this work. The first step towards PRSA computation is the determination of the so-called anchor point (APs). In this context, APs are defined as the time series points x_i fulfilling the following inequality:

$$\frac{1}{M}\sum_{j=0}^{M-1} x_{i+j} > \frac{1}{M}\sum_{j=1}^{M} x_{i-j}$$

where the parameter M is employed as the upper frequency bound for the periodicities to be detected by PRSA methodology.

After APs being detected, windows of length $2L$ are built symmetrically with respect to each AP. Given the fact that the majority of APs are temporally close one each other, the resulting windows are effectively overlapping. An additional specification for the parameter L is that it should be larger than the period of slowest oscillation to be detected [9].

The PRSA curve X_i is obtained by averaging the derived windows synchronized in their APs. After obtaining the PRSA curve, it is useful to summarize its characteristics by extracting different parameters. An example of such is APRS defined as:

$$APRS = \frac{\partial X_i}{\partial i}\bigg|_{i=AP}$$

The parameter APRS is a descriptor of the average increase in FHR amplitude and the time span of such increase event.

In this work, the considered signals x_i is the whole available FHR recording, thus resulting in a single APRS value. The parameters M and L are equal 40 and 200 respectively [6,10].

Deceleration Phase Rectified Slope (DPRS) **[bpm]**: the computations are analogous of those previously reported for APRS apart for the definition of APs which are defined as the time series points x_i fulfilling [9]:

$$\frac{1}{M}\sum_{j=0}^{M-1} x_{i+j} < \frac{1}{M}\sum_{j=1}^{M} x_{i-j}$$

Acknowledgments

The authors are grateful to Dr. M. Campanile and her team, Department of Obstetrical-Gynaecological and Urological Science and Reproductive Medicine, Federico II University, Naples, Italy for the collection of CTG tracings of healthy and IUGR fetuses. The research presented in this paper was supported by PRIN 2017 MIUR, ICT4MOMs [2017RR5EW3].

Conflict of Interest

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dib.2020.105164

References

[1] M.G. Signorini, N. Pini, A. Malovini, R. Bellazzi, G. Magenes, Integrating Machine Learning Techniques and Physiology Based Heart Rate Features for Antepartum Fetal Monitoring, Comput. Methods Progr. Biomed. (2019) 105015.

[2] D. Arduini, G. Rizzo, A. Piana, P. Bonalumi, P. Brambilla, C. Romanini, Computerized analysis of fetal heart rate—Part I: description of the sys- tem (2CTG), J. Matern. Fetal Invest. 3 (1993) 159–164.

[3] M.G. Signorini, G. Magenes, S. Cerutti, D. Arduini, Linear and nonlinear parameters for the analysis of fetal heart rate signal from cardiotocographic recordings, IEEE Trans. Biomed. Eng. 50 (3) (2003) 365–374.

[4] S. Pincus, Approximate entropy (ApEn) as a complexity measure, Chaos 5 (1) (Mar. 1995) 110–117.

[5] S.M. Pincus, R.R. Viscarello, Approximate entropy: a regularity measure for fetal heart rate analysis, Obstet. Gynecol. 79 (2) (1992) 249–255.

[6] A. Fanelli, G. Magenes, M. Campanile, M.G. Signorini, Quantitative assessment of fetal well-being through ctg recordings: a new parameter based on phase-rectified signal average, IEEE J. Biomed. Health Inf. 17 (5) (2013) 959–966.

[7] A. Lempel, J. Ziv, On the complexity of finite sequences, IEEE Trans. Inf. Theor. 22 (1) (1976) 75–81.

[8] M. Ferrario, M.G. Signorini, G. Magenes, Complexity analysis of the fetal heart rate for the identification of pathology in fetuses, Comput. Cardiol. 32 (2005) 989–992.

[9] A. Bauer, et al., Phase-rectified signal averaging detects quasi-periodicities in non-stationary data, Phys. Stat. Mech. Appl. 364 (2006) 423–434.

[10] M.G. Signorini, A. Fanelli, G. Magenes, Monitoring fetal heart rate during pregnancy: contributions from advanced signal processing and wearable technology, Comput. Math. Methods Med. 2014 (2014).