# Improving cardiotocography monitoring: a memory-less stream learning approach
## *Position Paper*

Pedro Pereira Rodrigues[1,2,3], Raquel Sebastião[1,4], and Cristina Costa Santos[2,3]

[1] LIAAD - INESC Porto, L.A., Portugal
[2] Faculty of Medicine of the University of Porto, Portugal
[3] CINTESIS - University of Porto, Portugal
[4] Faculty of Sciences of the University of Porto, Portugal
{pprodrigues,csantos}@med.up.pt, raquel@liaad.up.pt

**Abstract.** Cardiotocography is widely used, all over the world, for fetal heart rate and uterine contractions monitoring before (antepartum) and during (intrapartum) labor, regarding the detection of fetuses in danger of death or permanent damage. However, analysis of cardiotocogram tracings remains a large and unsolved issue. State-of-the-art monitoring systems provide quantitative parameters that are difficult to assess by the human eye. These systems also trigger alerts for changes in the behavior of the signals. However, they usually take up to 10 min to detect these different behaviors. Previous work using machine learning for concept drift detection has successfully achieved faster results in the detection of such events. Our aim is to extend the monitoring system with memory-less fading statistics, which have been successfully applied in drift detection and statistical tests, to improve detection of alarming events.

## 1 Introduction

Cardiotocography is widely used, all over the world, for fetal heart rate (FHR) and uterine contractions (UC) monitoring before (antepartum) and during (intrapartum) labor, regarding the detection of fetuses in danger of death or permanent damage [1]. However, analysis of both FHR and UC tracings remains a large and unsolved issue [5].

### 1.1 Cardiotocography monitoring systems

State-of-the-art monitoring systems, like Omniview-SisPorto [4], provide quantitative parameters that are difficult to assess by the human eye. The system also triggers alerts for changes in the behavior of the signals. Moreover, in the *normal* stage of tracings, four different patterns may be considered [11]. However, they usually take up to 10 min to detect these different behaviors. All these features are associated with possible damage to the fetus, which is usually assessed by the Apgar score. The Apgar score is usually determined 1 and 5 minute after birth by evaluating the newborn and ranges from zero to ten.

### 1.2 Machine learning in healthcare

The application of data mining and machine learning techniques to medical knowledge discovery tasks is now a growing research area. These techniques vary widely and are based on data-driven conceptualizations, model-based definitions or on a combination of data-based knowledge with human-expert knowledge [14]. Also, the definition of clinical decision support systems is now a major topic since it may help the diagnosis, the prognosis of rate of mortality, the prognosis of quality of life, or even treatment selection. However, the complicated nature of real-world biomedical data has made it necessary to look beyond traditional biostatistics [13] without loosing the necessary formality. For example, naive Bayesian approaches are closely related to logistic regression [19]. Hence, those systems could be implemented applying methods of machine learning [14], since new computational techniques are better at detecting patterns hidden in biomedical data, and can better represent and manipulate uncertainties [19]. In cardiotocography monitoring, previous work using machine learning has successfully been applied, achieving faster results in the detection of behavioral changes [20], and successfully clustering fetal heart rate tracings [18].

### 1.3 Aim and outline

The aim of this work is to extend the cardiotocography monitoring system with memory-less fading statistics [17] in order to improve the detection of change of behavior and classifying apgar score. Specifically, we intend to:

- define fading statistics for fetal heart rate and uterine contractions;
- define fading statistics for association between the two tracings;
- assess the relevance of fading statistics evolution for detecting changes of behavior in tracings;
- assess the relevance of fading statistics evolution in the prediction of newborn outcome through the apgar score at 1 and 5 minutes.

The paper is organized as follows. Next section presents background knowledge on cardiotocography monitoring systems, learning from data streams and fading statistics. Section 3 ends the exposition with some expected impact.

## 2 Background

This work is related with three different areas of research: cardiotocography monitoring systems, learning from data streams, and fading statistics.

### 2.1 Computer-based systems for cardiotocography analysis

Cardiotocography is a technique for continuous recording of fetal heart rate (FHR) and uterine contractions that is widely used to reduce in birth asphyxia that results in death or permanent damage to the newborn. However, there are important inconsistences in interpretation by experts of cardiotocograms and subsequent clinical decision [2]. Computer analysis of cardiotocograms provides quantitative parameters that are difficult to assess by the human eye overcoming the observer variability in interpretation of cardiotocograms. A program for automated analysis of tracings, developed over the last 15 years in University of Porto, Omniview SisPorto, provides visual and sound alerts for non reassuring fetal state [4]. However, it usually takes up to 10 min to trigger these alerts and so forth new methods are needed to improve the detection of non reassuring fetal state. Omniview-SisPorto system also provides the following quantitative parameters useful to medical interpretation of cardiotocograms and to subsequent clinical decision: the FHR baseline, the number of accelerations, the percentage of tracing with abnormal short-term variability (STV) and long-term variability (LTV) and the average STV and LTV.

FHR baseline was defined using a complex algorithm developed to identify the mean FHR during stable segments, in the absence of fetal movements and uterine contractions. Accelerations are defined as increases in the FHR above the baseline, lasting 15-120 seconds and reaching a peak of at least 15 beats per minute (bpm). Abnormal STV is identified when the difference to adjacent FHR signals is less than 1 bpm and abnormal LTV is identified whenever the difference between maximum and minimum FHR values of a sliding 60 seconds window centered on them, does not exceed 5 bpm [4]. Figure 1 presents plots of the usual monitoring features extrated from cardiotocography, for two different cases (one normal and one abnormal).

### 2.2 Machine learning from data streams

What distinguishes current data from earlier one are automatic data feeds. We do not just have people who are entering information into a computer. Instead, we have computers entering data into each other [15]. Thus, there are applications in which the data is modeled best not as persistent tables but rather as transient data streams.

A data stream is an ordered sequence of instances that can be read only once or a small number of times using limited computing and storage capabilities. The data elements in the stream arrive online, being potentially unbounded in size. Once an element from a data stream has been processed it is discarded or archived. It cannot be retrieved easily unless it is explicitly stored in memory, which is small relative to the size of the data streams. These sources of data are characterized by being open-ended, flowing at high-speed, and generated by non stationary distributions [8,9]. Learning techniques which operate through fixed training sets and generate static models are obsolete in these contexts. Faster answers are usually required, keeping an anytime data model and enabling better decisions, possibly forgetting older information.

The sequences of data points are not independent, and are not generated by stationary distributions. We need dynamic models that evolve over time and are able to adapt to changes in the distribution generating examples [8]. If the process is not strictly stationary (as most of real-world applications), the target

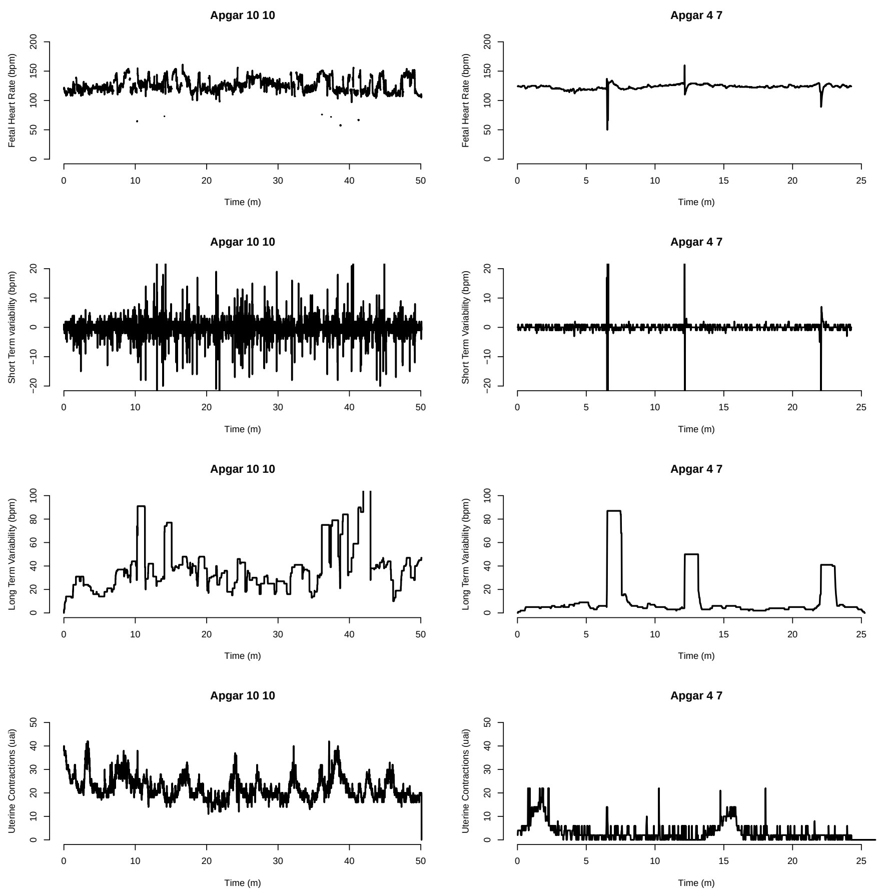


**Fig. 1.** Plots of monitoring features usual extrated from cardiotocography, for a normal patient (left) with Apgar scores equal to 10 (both at 1 and 5 minutes after birth) and an abnormal patient (right) with Apgar scores equal to 4 (1 minute after birth) and 7 (5 minutes after birth). Plots present the fetal heart rate (top), the STV (top middle), the LTV (top bottom) and the uterine contractions (bottom).

concept may gradually change over time. Hence data stream mining is an incremental task that requires incremental learning algorithms that take drift into account [7]. Previous work using stream learning for concept drift detection has successfully been applied to cardiotocography monitoring and achieved faster results in the detection of behavior changes [20].

### 2.3 Stream summarization using window models

In most streaming applications, recent data is the most relevant one [8]. To target this subset of data, a popular approach consists of defining a time window covering the most recent data. Actually, time windows are a commonly used approach to solve queries in open-ended data streams. Instead of computing an answer over the whole data stream, the query (or operator) is computed, eventually several times, over a finite subset of tuples. In this model, a time stamp is associated with each tuple. The time stamp defines when a specific tuple is valid (e.g. inside the window) or not. Several window models have been used in the literature. The most relevant are: *landmark*, *sliding* and *time-biased* windows.

Landmark windows [10] identify relevant points (the landmarks) in the data stream and the aggregate operator uses all record seen so far after the landmark. Successive windows share some initial points and are of growing size. In some applications, the landmarks have a natural semantic. For example, in daily basis aggregates the beginning of the day is a landmark.

Most of the time, we are only interested in computing statistics in the *strictly recent* past. The simplest approach are *sliding windows* of fixed size $w$. These type of windows are similar to *first in, first out* data structures: whenever an element $x_i$ is observed and inserted in the window, another element $x_{i-w}$ is forgotten. This is probably the most common approach to algorithms focusing on evolving recent data. However, due to the need to forget old observations, we need to maintain in memory all the observations inside the window. A recent work showed that, when dealing with evaluation of stream learning algorithms, the window size does not matter too much: the prequential error estimated over a sliding-window always converges fast to the holdout estimate, being on the other hand better suited for data streams [9].

Previous windows models use a catastrophic forget, that is, any past observation either is in the window or it is not inside the window. Usually in streaming settings, the concept generating data evolves smoothly, so old data is less but still important [7]. A smoother approach are *tilted time windows*, where time scale is compressed. The most recent data are stored inside the window at the finest detail (granularity). Oldest information is stored at a coarser detail, in an aggregated way, with the level of granularity depending on the application.

Even within a sliding window, the most recent data point is usually more important than the last one which is about to be discarded. This way, a simple approach could consider giving weights to data points depending on their age within the sliding window. Given its particular characteristics, a good approach for data streams uses an exponential approach, where the weight of a data point decreases exponentially with time: $\alpha$*-weighted window* [17]. The main advantages of this window model are two-fold. First, compared to traditional sliding windows, more importance is given to recent data points, as the weight of each observation decreases exponentially with time. Second, compared to other weighting approaches, it can be maintained on the fly. The main feature of the weighted sliding window model is the use of smooth forgetting. Hence, the computation of statistics over weighted windows raises several advantages when compared to statistics computed over simple sliding windows [17].

To avoid keeping all data in the window when computing statistics which are based on sums of the data points, and in order to include a smooth forgetting of information, the previous approach can be applied to achieve an approximated value for the elementary statistics on a data stream. Using the exponential weights introduced in the weighted window model, but applying then to all data points seen so far, the $\alpha$*-fading window* model is created, and similar statistics can be computed, to which we call *fading statistics*. A recent work showed that it is possible to use fading statistics as a error-bounded estimate of statistics computed over a weighted window [17].

The application of fading factors (which approximate the $\alpha$*-fading window* model) has been used in recent works. For example, given the fact that the prequential error [6] is based on the sum of errors along the stream, fading factors can be applied to achieve a memory-less approach to its computation over a sliding window. In a recent work, the authors have shown that the fading prequential error converges to the holdout estimate and is equivalent to the prequential error on a sliding window [9]. Also, on the same recent work, the authors also embedded fading factor techniques on statistical tests for comparing stream classification problems and change detection. Overall, the authors reported that the use of fading factors on the McNemar and Page-Hinkley tests gave results similar to the use of sliding windows [9].

## 3 Computing cardiotocography fading statistics

Any statistic that can be computed based on sums and counts (which are sums of variables taking values on $\{0, 1\}$) can be computed as a fading statistic, with the corresponding exponential bias towards recent examples [17]. The precise domain of cardiotocography monitoring possess characteristics that direct the search for relevant statistics.

### 3.1 $\alpha$-fading statistics

For a single continuous variable, two simple statistics are the average and standard deviation, but we can also compute a histogram to approximate the variables distribution. In streaming settings, these should take into account the age of the data points, so we should compute moving statistics in order to be adapted to the most recent data. Using the $\alpha$-fading window model, with $0 < \alpha < 1$, these statistics need to be defined slightly differently. Previous work [17] has presented definitions for $\alpha$-fading statistics including increments, sums and averages. Considering $i$ the number of current observations of a given variable $X$, when possible, we will use the recursive forms to illustrate its applicability to online systems:

- The **$\alpha$-fading increment** is a weighted count of observations, defined as

$$N_\alpha(i) = \begin{cases} 1, & i = 1 \\ 1 + \alpha \times N_\alpha(i-1), & i > 1 \end{cases} \tag{1}$$

  with $\lim\limits_{i \to \infty} N_\alpha(i) = \frac{1}{(1-\alpha)}$ (proof in [17]).
- The **$\alpha$-fading sum** is a weighted sum of the observations, where

$$S_{x,\alpha}(i) = \begin{cases} x_1, & i = 1 \\ x_i + \alpha \times S_{x,\alpha}(i-1), & i > 1 \end{cases} \tag{2}$$

  with the $\alpha$-fading increment being the total amount of weight given to observations in the $\alpha$-fading sum (proof in [17]).
- The **$\alpha$-fading average** is a weighted average of observations, where

$$M_{x,\alpha}(i) = \frac{S_{x,\alpha}(i)}{N_\alpha(i)}, \tag{3}$$

  with the $\alpha$-fading average approximating the $\alpha$-weighted average with a maximum error of $2\varepsilon R$, where $\varepsilon$ is the allowed proportion of weight given to observations outside the weighted window, and $R$ the range of the variable [17].

Similar approaches can be made for $\alpha$-fading variance (hence, standard deviation) and $\alpha$-fading correlation, where the sufficient statistics needed to compute the final measure are kept as $\alpha$-fading sums.

- The **$\alpha$-fading variance** is computed as (the fading factor can make the second term higher than the first one, hence the need for the absolute value):

$$V_{x,\alpha}(i) = \left\| \frac{S_{x^2,\alpha}(i)}{N_\alpha(i)} - \frac{S_{x,\alpha}(i)^2}{N_\alpha(i)^2} \right\|. \tag{4}$$

- Given its possible computation as an algebraic operation of sums, we can define the **$\alpha$-fading correlation** coefficient as

$$C_{x,y,\alpha}(i) = \frac{S_{xy,\alpha}(i) - \frac{S_{x,\alpha}(i)S_{y,\alpha}(i)}{N_\alpha(i)}}{\sqrt{\left\| S_{x^2,\alpha}(i) - \frac{S_{x,\alpha}(i)^2}{N_\alpha(i)} \right\|}\sqrt{\left\| S_{y^2,\alpha}(i) - \frac{S_{y,\alpha}(i)^2}{N_\alpha(i)} \right\|}}. \tag{5}$$

Another frequently used summary are online histograms. The histogram is defined by a set of $k$ non-overlapping intervals $I_1, \ldots, I_k$ in the range of the random variable, and a set of frequency counts $F_1(i), \ldots, F_k(i)$. For each observation $i$ of a given variable $X$, the online histogram counts are updated by making $F_l(i) = F_l(i-1) + 1$, if $x_i \in I_l$, or $F_l(i) = F_l(i-1)$ otherwise, with $l = 1..k$. In the fading window model, each **$\alpha$-fading frequency** is computed as

$$F_{\alpha,l}(i) = c_{li} + \alpha \times F_{\alpha,l}(i-1), \tag{6}$$

where $c_{li}$ is 1 if $x_i \in I_l$, and 0 otherwise. Hence, the $\alpha$-fading frequency is a $\alpha$-fading sum of a variable taking only values in $\{0, 1\}$. The collection of $\alpha$-fading frequencies creates the **$\alpha$-fading histogram**.

### 3.2 Summarizing the cardiotocogram

It is known that a low fetal heart rate baseline or variability is an indicator of problems to the fetus [3]. Hence, we shall monitor the $\alpha$-fading average $M_{h,\alpha}(i)$ and $\alpha$-fading variance $V_{h,\alpha}(i)$ of the fetal heart rate signal ($h$).

When analysing both fetal heart rate ($h$) and uterine contractions ($u$) signals, an increase of uterine contraction conjugated with a persistant decrease in fetal heart rate is also a sign of potencial damage [3]. Hence, negative correlations between the signals are alerting, so we shall monitor the $\alpha$-fading correlation $C_{h,u,\alpha}(i)$.

Empirically, changes in the distance between the two distributions could also alarm for problems so we also plan to monitor the evolution of the distance between the two $\alpha$-fading histograms, using well-known metrics, such as the Kullback-Leibler divergence [12].

Figure 2 presents plots of the usual monitoring features extrated from cardiotocography, for two different cases (one normal and one abnormal), but using fading statistics to improve visualization and analysis.

## 4 Concept drift detection using fading statistics

Several tests for change detection have been presented in the literature [7]. In previous work, sliding-window-based detection has already been applied on cardiotocography with good results [20]. Given the fact that the fading statistics already define a window model, we could apply simple statistic-based approaches to our problem.

One of the most referred is the Page-Hinkley Test (PHT), a sequential analysis technique typically used for monitoring change detection in signal processing [16]. It allows efficient detection of changes in the normal behavior of a process which is established by a model. This test maintains a cumulative variable $m_T$, defined as the cumulated difference between the observed values and their mean till the current moment:

$$m_T = \sum_{t=1}^{T} (x_t - \bar{x}_T - \delta), \tag{7}$$

where $\bar{x}_T = \frac{1}{T} \sum_{t=1}^{t} x_t$ and $\delta$ corresponds to the magnitude of changes that are allowed. The minimum value of this variable is also computed: $M_T = min(m_t, t =$

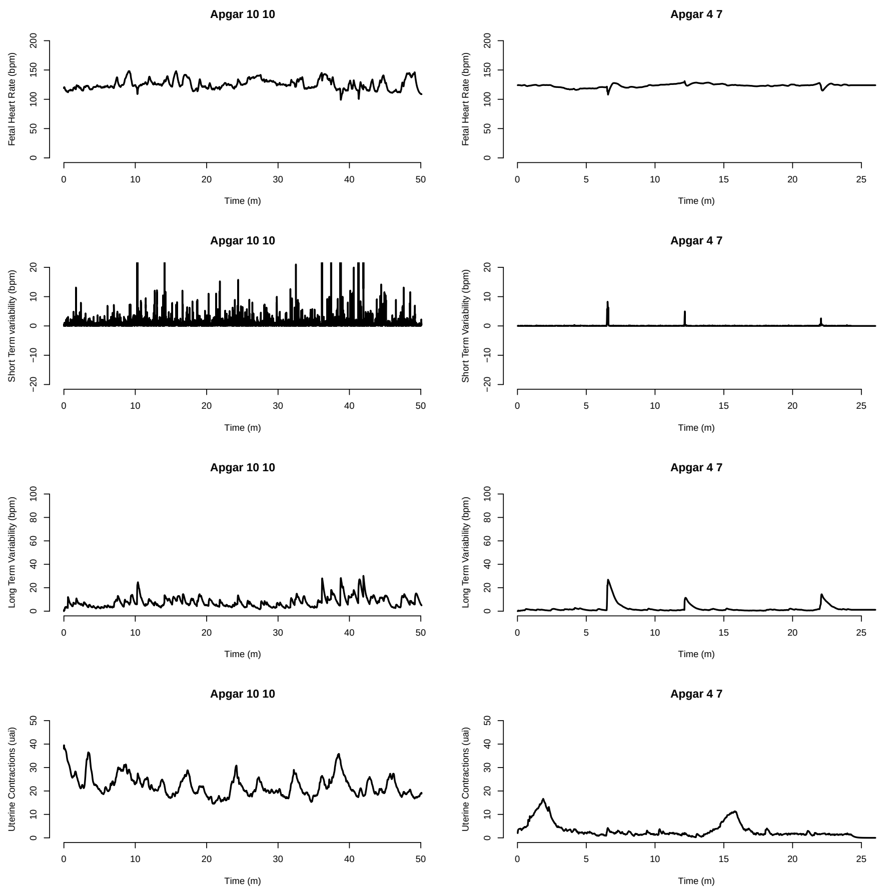


**Fig. 2.** Plots of monitoring features usual extrated from cardiotocography, for a normal patient (left) with Apgar scores equal to 10 (both at 1 and 5 minutes after birth) and an abnormal patient (right) with Apgar scores equal to 4 (1 minute after birth) and 7 (5 minutes after birth). Plots present the fetal heart rate $\alpha$-fading average (top, $\alpha = 0.980$ approximating a window of 1 minute with 1% error), the fetal heart rate $\alpha$-fading stardard deviation (top middle, $\alpha = 0.316$ approximating a window of 1 second with 1% error), the fetal heart rate $\alpha$-fading stardard deviation (top bottom, $\alpha = 0.980$ approximating a window of 1 minute with 1% error) and the uterine contractions $\alpha$-fading average (bottom, $\alpha = 0.980$ approximating a window of 1 minute with 1% error).

$1 \ldots T$). As a final step, the test monitors the difference

$$PH_T = m_T - M_T. \tag{8}$$

When this difference is greater than a given threshold ($\lambda$) we alarm a change in the distribution. The threshold $\lambda$ depends on the admissible false alarm rate. Increasing $\lambda$ will entail fewer false alarms, but might miss or delay some changes. To detect decreases, a similar test can be conducted. Previous work as shown that this test could also be adapted to fading statistics [9], so we plan to use it as concept drift detector for all the summaries defined in the previous section.

## 5 Available data and expected impact

A total of 31 antepartum FHR tracings obtained in a previously reported study are available. These tracings were acquired in four hospitals located in Portugal, Switzerland, Germany and Australia, in the context of a multicentre observational study [3]. Tracings were acquired using Hewlett-Packard M1350 fetal monitors at a 4 Hz sampling rate in three hospitals, and with a Sonicaid 8000 fetal monitor in the remaining hospital, using true beat to beat intervals. All tracings were acquired in singleton pregnancies with no fetal malformations, and had at least 30 minutes of duration and less than 15% signal loss. In ten tracings the newborn outcome was bad (Apgar at first minute after born was less than 7).

As traditionally targeted, we expect to identify in cardiotocografic tracings the non reassuring fetal state, predicting the newborn outcome, which is considered a bad newborn outcome when Apgar score measured 1 minute after the birth is under 7. Moreover, given the application of fading statistics as low-pass filter, we believe that the sole visualization of fading statistics evolution might in fact improve the physicians accuracy and agreement in the diagnosis.

**Acknowledgments**

This work is supported by Portuguese Foundation for Science and Technology (FCT) under projects KDUDS (PTDC/EIA-EIA/98355/2008) and CSI2 (PTDC/EIA-CCO/099951/2008).

## References

1. American College of Obstetricians and Gynecologists: ACOG practice bulletin. clinical management guidelines for obstetrician-gynecologists, number 70. "Intrapartum fetal heart rate monitoring". Obstetrics and Gynecology 106(6), 1453–1460 (2005)
2. Ayres-de-Campos, D., Bernardes, J., Costa-Pereira, A., Pereira-Leite, L.: Inconsistencies in classification by experts of cardiotocograms and subsequent clinical decision. British Journal of Obstetrics and Gynecology 106(12), 1307–1310 (1999)

3. Ayres-de-Campos, D., Costa-Santos, C., Bernardes, J.: Prediction of neonatal state by computer analysis of fetal heart rate tracings: the antepartum arm of the sisporto multicentre validation study. European Journal of Obstetrics and Gynecoly and Reproductive Biology 118(1), 52–60 (2005)
4. Ayres-de-Campos, D., Sousa, P., Costa, A., Bernardes, J.: Omniview-sisporto 3.5 - a central fetal monitoring station with online alerts based on computerized cardiotocogram+ST event analysis. Journal of Perinatal Medicine 36(3), 260–264 (2008)
5. Bernardes, J., Pereira, A.C., de Campos, D.A., van Geijn, H., Leite, P.: Evaluation of interobserver agreement of cardiotocograms. International Journal of Gynecology and Obstetrics 177, 1385–1390 (1997)
6. Dawid, A.P.: Statistical theory: The prequential approach. Journal of the Royal Statistical Society 147, 278–292 (1984)
7. Gama, J., Medas, P., Castillo, G., Rodrigues, P.P.: Learning with drift detection. In: Bazzan, A.L.C., Labidi, S. (eds.) Proceedings of the 17th Brazilian Symposium on Artificial Intelligence (SBIA 2004). Lecture Notes in Artificial Intelligence, vol. 3171, pp. 286–295. Springer Verlag, São Luiz, Maranhão, Brazil (October 2004)
8. Gama, J., Rodrigues, P.P.: Data stream processing. In: Gama, J., Gaber, M.M. (eds.) Learning from Data Streams - Processing Techniques in Sensor Networks, chap. 3, pp. 25–39. Springer Verlag (2007)
9. Gama, J., Sebastião, R., Rodrigues, P.P.: Issues in evaluation of stream learning algorithms. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2009). pp. 329–337. ACM Press, Paris, France (2009)
10. Gehrke, J., Korn, F., Srivastava, D.: On computing correlated aggregates over continual data streams. In: SIGMOD Conference. pp. 13–24 (2001)
11. Gonçalves, H., Bernardes, J., Rocha, A.P., Ayres-de-Campos, D.: Linear and non-linear analysis of heart rate patterns associated with fetal behavioral states in the antepartum period. Early Human Development 83(9), 585–591 (2007)
12. Johnson, D.H., Sinanovic, S.: Symmetrizing the kullback-leibler distance. Tech. rep., Computer and Information Technology Institute, Texas (March 2001)
13. Lucas, P.: Bayesian analysis, pattern analysis, and data mining in health care. Current Opinion in Critical Care 10, 399–403 (2004)
14. Mitchell, T.M.: Machine Learning. McGraw-Hill, international edn. (1997)
15. Muthukrishnan, S.: Data Streams: Algorithms and Applications. Now Publishers Inc, New York, NY (2005)
16. Page, E.S.: Continuous inspection schemes. Biometrika 41((1/2)), 100–115 (1954)
17. Rodrigues, P.P., Gama, J., Sebastião, R.: Memoryless fading windows in ubiquitous settings. In: Proceedings of the First Ubiquitous Data Mining Workshop. pp. 23–27. Lisboa, Portugal (August 2010)
18. Santos, C.C., Bernardes, J., Vitányi, P., Antunes, L.: Clustering fetal heart rate tracings by compression. In: 19th IEEE International Symposium on Computer-Based Medical Systems (CBMS 2006). pp. 685–690. IEEE Computer Society (2006)
19. Schurink, C., Lucas, P., Hoepelman, I., Bonten, M.: Computer-assisted decision support for the diagnosis and treatment of infectious diseases in intensive care units. Lancet Infectious Diseases 5, 305–312 (2005)
20. Sebastião, R., Gama, J., Rodrigues, P.P., Bernardes, J.: Monitoring incremental histogram distribution for change detection in data streams. In: Gaber, M.M., Vatsavai, R.R., Omitaomu, O.A., Gama, J., Chawla, N.V., Ganguly, A.R. (eds.) Knowledge Discovery from Sensor Data. Lecture Notes on Computer Science, vol. 5840, pp. 25–42. Springer Verlag (2010)