Sparse learning for Intrapartum fetal heart rate analysis

Patrice Abry, Jiří Spilka, R Leonarduzzi, V Chudáček, Nelly Pustelnik, M. Doret

Sparse learning for Intrapartum fetal heart rate analysis

Article · April 2018

DOI: 10.1088/2057-1976/aabc64

6 authors, including:

Patrice Abry
Ecole normale supérieure de Lyon
344 PUBLICATIONS **7,354** CITATIONS

SEE PROFILE

Jiří Spilka
Czech Technical University in Prague
36 PUBLICATIONS **457** CITATIONS

SEE PROFILE

Roberto Fabio Leonarduzzi
Ecole Normale Supérieure de Paris
29 PUBLICATIONS **128** CITATIONS

SEE PROFILE

Sparse learning for intrapartum fetal heart rate analysis

P. Abry[(1)], J. Spilka[(2)], R. Leonarduzzi[(1)], V. Chudáček[(2)], N. Pustelnik[(1)], M. Doret[(3)]

[(1)] Univ Lyon, Ens de Lyon, Univ Claude Bernard, CNRS, Laboratoire de Physique, Lyon, France

[(2)] CIIRC, Czech Technical University in Prague, Czech Republic

[(3)] Obstetric Dept., Femme-Mère-Enfant Hospital, Bron, France

E-mail: patrice.abry@ens-lyon.fr

Abstract. Fetal Heart Rate (FHR) monitoring is used during delivery for fetal well-being assessment. Classically based on the visual evaluation of FIGO criteria, FHR characterization remains a challenging task that continuously receives intensive research efforts. Intrapartum FHR analysis is further complicated by the two different stages of labor (dilation and active pushing). Research works aimed at devising automated acidosis prediction procedures are either based on designing new advanced signal processing analyses or on efficiently combining a large number of features proposed in the literature. Such multi-feature procedures either rely on a prior feature selection step or end up with decision rules involving long lists of features. This many-feature outcome rule does not permit to easily interpret the decision and is hence not well-suited for clinical practice. Machine-learning-based decision-rule assessment is often impaired by the use of different, proprietary and small databases, preventing meaningful comparisons of results reported in the literature. Here, sparse learning is promoted as a way to perform jointly feature selection and acidosis prediction, hence producing an optimal decision rule based on as few features as possible. Making use of a set of 20 features (gathering "FIGO-like" features, classical spectral features and recently proposed scale-free features), applied to two large-size (respectively $\simeq$ 1800 and $\simeq$ 500 subjects), well-documented databases, collected independently in French and Czech hospitals, the benefits of sparse learning are quantified in terms of: i) accounting for class imbalance (few acidotic subjects), ii) producing simple and interpretable decision rules, iii) evidences for differences between the temporal dynamics of active pushing and dilation stages, and iv) of validity/generalizability of decision rules learned on one database and applied to the other one.

Keywords: Intrapartum Fetal Heart Rate, Acidosis prediction, labor Stage, Sparse learning, scale-free, learning generalization.

1. Introduction

Intrapartum fetal monitoring. Worldwide, obstetricians routinely monitor, during delivery, fetal heart rate (FHR) to detect oxygen deprivation in fetuses. Their main objective is to predict fetal acidosis as early as possible, so as to make timely decisions for operative deliveries that prevent adverse outcomes (neural development disability, neonatal encephalopathy, and cerebral palsy) (Chandraharan & Arulkumaran 2007). Clinical routine for FHR evaluation remains based on visual inspection of FHR, relying on guidelines issued by the International Federation of Gynecology and Obstetrics (FIGO) (FIGO 1986, Ayres-de Campos et al. 2015) mostly focusing on decelerations (number, shape, depth and duration), variability, and baseline levels. This practice, however, is well known to suffer from substantial inter- and intra-observer variability (Hruban et al. 2015, Blackwell et al. 2011, Spilka, Chudáček, Janků, Hruban, Burša, Huptych, Zach & Lhotská 2014) and to induce unnecessary intervention or operative deliveries, increasing uselessly cesarean section rate, that are a posteriori found to have been avoidable (False Positive) (Alfirevic et al. 2006). Improving FHR analysis by having recourse to advanced statistical signal processing tools has thus been the topic of on-going research efforts.

Related works: feature design. Beyond attempts to automatize the computation of FIGO features–or generalized versions (cf., e.g., (Parer et al. 2006, Spilka et al. 2012, Nunes et al. 2017))–frequency-based features were used (Siira et al. 2005), following standard approaches for adult heart rate variability analysis (Akselrod et al. 1981). Such features are often referred to as *linear* since they essentially quantify autocorrelations in FHR (Gonçalves et al. 2006, Laar et al. 2008, Magenes et al. 2003, Siira et al. 2013). Further, advanced statistical signal processing was involved in FHR analysis. Notably, complexities in temporal dynamics were quantified using information theoretic quantities, such as entropy rates (Costa et al. 2002, Echeverria et al. 2004, Porta et al. 2013, Spilka, Roux, Garnier, Abry, Goncalves & Doret 2014, Granero-Belinchon et al. 2017), various nonlinear transforms (Magenes et al. 2000, Magenes et al. 2003, Chudáček, Anden, Mallat, Abry & Doret 2014, Georgieva et al. 2014), or the scale-free or (multi)fractal paradigm (Francis et al. 2002, Doret, Helgason, Abry, Gonçalvès, Gharib & Gaucherand 2011, Abry et al. 2013, Doret et al. 2015). Such features are termed *nonlinear* as they probe the dependencies in FHR temporal dynamics potentially beyond mere autocorrelations. For overviews, see e.g. (Spilka et al. 2012, Haritopoulos et al. 2016).

Related works: feature selection and machine learning. Faced with the modest acidosis-prediction performance achieved by standalone features, numerous works combined several features through supervised machine learning strategies (cf., e.g., (Bernardes et al. 1991, Costa et al. 2009, Warrick et al. 2010, Georgieva et al. 2013, Spilka et al. 2012, Czabanski et al. 2012, Warrick et al. 2010, Xu et al. 2014, Frasch et al. 2014, Dash et al. 2014, Spilka et al. 2017)). Most of these contributions acknowledge several problematic issues. First, the use of small-size databases (from

tens to hundreds of subjects) and of a large number of features results in a lack of robustness and generalizability (Frasch et al. 2014). Second, complicated decision rules involving too many features are not suited to clinical practice as clinicians often need to *conceptualize* what their decision is based on (Spilka et al. 2012, Xu et al. 2014)—feature selection is thus often performed prior to learning, but without guarantees on the joint optimality of both procedures. Third, the use on one database of decision rules learned from another does not always yield satisfactory performance.

Related works: labor stages. FHR analysis is further complicated by the existence of two distinct stages during labor. The dilation stage (stage I) consists in progressive cervical dilation and regular contractions. The active-pushing stage (stage II) portrays a fully dilated cervix combined with maternal pushing efforts. Analyses have been performed either globally, mixing both stages (Costa et al. 2009, Warrick et al. 2010), or focusing on stage I only (Spilka et al. 2017). Differences in the dynamics of each stage remain barely documented (see a contrario (Spilka, Abry, Goncalves & Doret 2014, Lim et al. 2014, Spilka et al. 2016*b*, Granero-Belinchon et al. 2017)).

Outline, goals, and contributions. The present contribution aims to assess the potential benefits, for acidosis prediction, of *sparse learning*—i.e. the combination of feature selection and classification into a single procedure. It also aims to show that decision rules learned for each stage need to be different and based on different features. Finally, it aims to validate that sparse learning yields decision rules that satisfactorily *generalize*—i.e. they remain valid on samples not used for training. To that end, the Sparse Support Vector Machine (S-SVM)–detailed in Section 3– is applied to FHR data from two independent large-size databases–described in Section 2. Results, discussed in Section 4, highlight the benefits of sparse classification rules, and enhance the evidence for differences between the temporal dynamics of both stages. Further, the generalization ability of S-SVM is explored through a cross-database evaluation.

The present work complements and strengthens preliminary attempts reported in (Spilka et al. 2016*a*, Spilka et al. 2016*b*, Spilka et al. 2017).

2. Data: databases and datasets

2.1. Databases

Two independent large-size databases are used in the present work. They were collected—with different technologies and constraints—at academic hospitals in Lyon, France (LDB), and Brno, Czech Republic (BDB). They share comparable clinical characteristics in terms of gestational age larger than 36 weeks, cf. Table 1 for details.

For LDB, FHR data were collected at the French public Hospital Femme-Mère-Enfant, from 2000 to 2010, during routine labor monitoring. The acquisition was performed using STAN S21 or S31 devices (Neoventa Medical, Molndal, Sweden) via internal fetal scalp electrodes (12-bit resolution and sampling rate 500 Hz). LDB contains 3049 recordings, and is documented with relevant clinical information such as

Table 1. Database and datasets. Clinical data, reported as mean (standard deviation), for the LDB and BDB databases, and S_I and S_{II} datasets. Only operative deliveries due to fetal distress only are included.

		LDB		BDB	
		Acidotic	Normal	Acidotic	Normal
S_I	# cases	29	1021	14	330
	Umb. cord art. pH	7.01 (0.03)	7.24 (0.07)	7.00 (0.05)	7.26 (0.08)
	Apgar score at 5 min	9.38 (0.90)	9.89 (0.53)	7.71 (1.77)	9.26 (0.89)
	Birth-weight (g)	3367 (435.31)	3328 (471)	3344 (543.18)	3367 (449)
	Male/female (n)	14 (48%)	547 (54%)	10 (71%)	170 (52%)
	t_{II} (min)	8.5 (5.2)	6.8 (5.1)	10.3 (4.8)	9.3 (3.9)
	# Operative delivery	13 (45%)	213 (21%)	–	–
S_{II}	# cases	27	735	12	116
	Umb. cord art. pH	7.01 (0.04)	7.22 (0.06)	6.99 (0.05)	7.22 (0.07)
	Apgar score at 5 min	9.56 (0.80)	9.90 (0.43)	7.67 (1.44)	8.91 (1.04)
	Birth-weight (g)	3469 (397)	3365 (444)	3265 (437)	3421.81 (405)
	Male/female (n)	13 (48%)	392 (53%)	5 (42%)	61 (53%)
	# $t_{II} \leq 15$ min (n)	27.9 (9.8)	27.6 (9.8)	25.0 (4.3)	23.7 (4.3)
	Operative delivery(n)	13 (48%)	152 (21%)	–	–

umbilical artery and venous pH after delivery and decision to intervene due to suspected fetal acidosis, cf. (Doret, Massoud, Constans & Gaucherand 2011) for details.

For BDB, 552 FHR recordings were acquired at the Hospital in Brno, from 2010 to 2012, using STAN S21 or S31 scalp electrodes, or Avalon FM40 or FM50 Doppler-based devices (Phillips Healthcare, Andover, MA), via either ultrasound probes or scalp electrodes (12 bit resolution, sampling rate 4 Hz). BDB has been made an open-access database (Chudáček, Spilka, Burša, Janků, Hruban, Huptych & Lhotská 2014).

Tracings were included in the present study according to clinical and data-quality criteria (Doret, Massoud, Constans & Gaucherand 2011, Spilka et al. 2017): gestational age ≥ 37 weeks, maternal age ≥ 18, tracing ending less than 20 minutes before delivery, after-delivery pH measurement available, less than 50% of missing data in either stage of the delivery process.

2.2. Datasets

For each database, tracings are split into two datasets, depending on the duration of the second stage (t_{II}): Set S_I is defined as $t_{II} \leq 15$ min and corresponds to births during stage I or in the early phase of the stage II. Set S_{II} corresponds to $t_{II} > 15$ min and thus to births during (a possibly long) stage II.

Assessment of the newborn's health status at delivery is based on pH, measured from an immediate post-birth umbilical-cord-artery blood sample, and used as a signature of fetal acidosis in the minutes before delivery (Amer-Wåhlin et al. 2001). Subjects were hence split into two classes: *acidotic* defined as pH ≤ 7.05 and *normal* with pH > 7.05. Clinical information on the resulting datastets is summarized in Table 1.

2.3. Data preprocessing, FHR time series, and analysis

Collected tracings consist of a list of R-Peak intervals (RRi) in ms. They were first corrected for outliers and missing data by a sliding median filter. RRi lists were resampled using linear spline interpolation to yield regularly-sampled beat-per-minute (bpm) time series $X(t)$. Sampling frequencies were set to $f_s = 10$ Hz for LDB and $f_s = 4$ Hz for BDB— due to the different sampling frequencies in both databases.

Because pH measured at delivery can only be reminiscent of the health status of fetuses in the last minutes of labor, FHR analysis was only conducted in the last 20 min of the first stage for S_I, and the last 20 min before delivery for S_{II}, as sketched in Fig. 1.

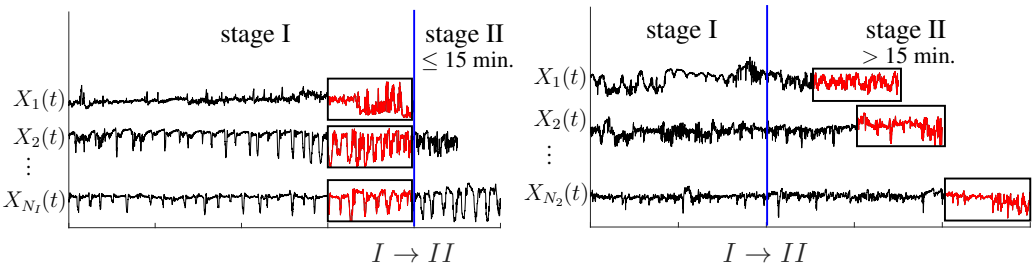

Figure 1. Stage splitting. Analyzed FHR data are marked by the time windows framed in rectangles boxes, corresponding to the last 20min of StageI (for StageI dataset, left) and the last 20 min before delivery for StageII dataset (right).

3. Methods: Feature Design and Sparse Learning

3.1. Feature Design

Following preliminary works reported in (Spilka et al. 2016*a*, Spilka et al. 2016*b*, Spilka et al. 2017), a set of 20 features is used, chosen amongst features either widely used in the literature (such as *automated FIGO-like* features), derived from adult heart rate analysis (such as *spectral* features), or more recently shown to be relevant for intrapartum fetal heart rate analysis (such as *scale-free* features). Table A1 in Appendix A summarizes the feature lists as well as other acronyms used throughout the manuscript.

3.1.1. Automated FIGO-like features (9). Nine FIGO-inspired features are used: $(\beta_0, \beta_1, LTV, STV, \#acc, \#dec, \mathrm{MAD}_{dt}, T_{stress}, A_{dec})$, as devised in (Chudáček et al. 2011, Spilka et al. 2012). They quantify: baseline $B(t)$ level and evolution as $B(t) = \beta_0 + \beta_1 t$; long- and short-term variabilities (LTV, STV) (Ayres-de Campos et al. 2015); accelerations/decelerations (Ayres-de Campos et al. 2015), via their numbers ($\#acc$ and $\#dec$), average depth (MAD_{dt}), average duration (T_{stress}) and average area (A_{dec}).

3.1.2. Spectral features (5). Spectral estimation is conducted over using the Welch periodogram (Manolakis et al. 2005). For adults, energies in well-defined frequency bands are known to be associated with the sympathetic/parasympathetic balance. Because no widely-accepted equivalent definitions are available for fetuses (Siira et al. 2013, Doret et al. 2015), the same bands as those for adults are used: very low frequency E_{VLF} ($[0.003, 0.04]$ Hz), low frequency E_{LF} ($[0.04, 0.15]$ Hz), and high frequency E_{HF} ($[0.15, 0.40]$ Hz). Further, the ratio LF/HF of E_{LF} and E_{HF}, and the spectral index α (estimated over LF and HF), are computed.

3.1.3. Scale-free features (6). Scale-free and multifractal features were recently shown to offer relevant and robust alternatives to the classical measurements of long- and short-term variabilities (STV and LTV), cf. (Doret, Helgason, Abry, Gonçalvès, Gharib & Gaucherand 2011, Abry et al. 2013, Doret et al. 2015). The Hurst parameter H (cf., e.g, (Samorodnitsky & Taqqu 1994) for a definition) is a linear feature, as it describes the autocorrelation function or the Fourier spectrum, yet in a scale-free spirit (Abry et al. 2013, Doret et al. 2015). It has been shown that it can be efficiently and robustly computed from discrete wavelet transforms (Abry & Veitch 1998, Abry & Didier 2018). Multifractal parameters $(h_{min}, c_1, c_2, c_3, c_4)$ (cf., e.g., (Wendt et al. 2007) for a definition) produce an advanced scale-free characterization of a time series' variability. Indeed, they describe the fluctuations of its regularity along time, based on the full dependence structure—not just the autocovariance. Efficient and robust estimators of multifractal parameters are obtained from wavelet leaders, a nonlinear transform of wavelet coefficients (Wendt et al. 2007). Parameters h_{min} and c_1 remain mostly driven by the autocorrelation and are hence closely related to H. Conversely, parameters (c_2, c_3, c_4) convey information not encoded in the autocorrelation and thus complement H, c_1, h_{min}. They are hence nonlinear features, quantifying the variability, asymmetry and heavy-tailness in FHR's temporal dynamics (Wendt et al. 2007, Doret, Helgason, Abry, Gonçalvès, Gharib & Gaucherand 2011).

3.2. Sparse Learning

3.2.1. Support Vector Machine. A Support Vector Machine (SVM) is a classical machine learning procedure (Hastie et al. 2009), already used for acidosis classification, cf., e.g., (Warrick et al. 2010, Xu et al. 2014). Let $\mathbf{x}_n \in \mathbb{R}^P$ denote P-dimensional feature vectors for each of the N subjects, and let $y_n \in \{-1, 1\}$ be the corresponding class labels. SVM produces a decision rule $d = sgn(\mathbf{w}^T\mathbf{x}_n + b)$ to classify the subjects. The parameters $(\mathbf{w}, \widehat{b})$ are estimated by minimizing a functional that enforces classification performance through the hinge loss function $F_{\mathbf{w},b}(\mathbf{x}, y) = \max\big(0, 1 - y(\mathbf{w}^T\mathbf{x} + b)\big)$:

$$(\widehat{\mathbf{w}}, \widehat{b}) \in \underset{\mathbf{w}\in\mathbb{R}^P,\, b\in\mathbb{R}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{n=1}^{N} F_{\mathbf{w},b}(\mathbf{x}_n, y_n), \tag{1}$$

where the regularization hyper-parameter $C > 0$ controls *data sparsity*. Indeed, when

solving Eq. (1), only a few samples (the so-called *support vectors*, whose overall amount is indirectly induced by C) actually contribute to the design of the decision rule.

While (1) can be easily solved by gradient descent algorithms, SVM suffers from several documented limitations (cf., e.g., (Warrick et al. 2010)): i) *data sparsity* is not well suited to biomedical applications, where feature distributions are often largely intertwined for the two classes; ii) decision rules involve essentially all proposed features—$\widehat{w}_p \neq 0, \forall p$—thus making their interpretation difficult (a severe impairment for transfer towards clinical practice) ; iii) While SVM show some robustness to unbalanced class sizes thanks to potential support vector sparsity, the special context of fetal acidosis prediction is characterized with extremely unbalanced sizes of the acidotic and healthy classes, $N_+/N_- \ll 1$, which requires special attention.

3.2.2. Sparse Support Vector Machine. To overcome such issues, feature selection procedures were envisaged (Xu et al. 2014, Bron et al. 2015, Soguero-Ruiz et al. 2015). Yet, the concatenation of two phases (feature selection and classification) casts shadows on their joint optimality. Instead, Sparse Support Vector Machines (S-SVM) aim to perform both operations jointly (Hastie et al. 2009, Bach et al. 2012, Laporte et al. 2014, Chierchia et al. 2016). The condition of *feature sparsity*—involving as few features as possible—requires trading away the *data sparsity*. This can be achieved by imposing an ℓ_1-norm to the weights $\widehat{w}$, a classical method to enforce as many zeros in $\widehat{w}$ as can be (Blondel et al. 2013, Combettes & Wajs 2005). In turns, to ensure that the resulting functional can actually be minimized, it is needed to square the hinge loss function $F_{\mathbf{w},b}(\mathbf{x}, y)$. Theoretical alternative strategies are investigated in (Chierchia et al. 2016).

Further, to account for the severe imbalanced class sizes in the fetal acidosis prediction problem, we propose to modify the penalization term, consisting of a sum $\sum_{n=1}^{N}$ across all subjects in the database, by further splitting it into two weighted sums according to the class sizes N_+ and N_-, with the additional introduction of an hyperparameter λ. Like for hyper parameter C, its tuning (by cross-validation) is expected to permit to optimize prediction performance.

The functional to minimize to estimate the S-SVM decision rule then reads:

$$(\widehat{\mathbf{w}}, \widehat{b}) \in \underset{\mathbf{w}\in\mathbb{R}^P,\, b\in\mathbb{R}}{\operatorname{argmin}} \|\mathbf{w}\|_1 + C\frac{\lambda}{N_-}\sum_{n=1}^{N_-} F^2_{\mathbf{w},b}(\mathbf{x}_n, y_n) + C\frac{(1-\lambda)}{N_+}\sum_{n=1}^{N_+} F^2_{\mathbf{w},b}(\mathbf{x}_n, y_n) \qquad (2)$$

where C controls the tradeoff between *feature sparsity* and classification performance, and $\lambda \in (0, 1)$ controls the balance between False Positives and False Negatives. In the decision rule

$$d = sgn(\mathbf{w}^T\mathbf{x}_n + b) \qquad (3)$$

only few weights w_p are non zero at the price though of increased difficulties to obtain the minimum of (2). Indeed, the non differentiable nature of the ℓ_1-norm precludes the use of gradients and requires that of more involved proximity operators (Blondel et al. 2013, Combettes & Wajs 2005). For fixed hyper-parameters, (C, λ), Eq. 2 is

solved in the primal space using a Forward-Backward Splitting Algorithm developed by ourselves.

3.2.3. Hyper-parameter tuning: Cross validation. The selection of hyper-parameters C and λ, as well as the assessment of *generalization*‡ performance, are challenging issues. Hyper-parameter (C, λ) selection is here done using *single-loop* cross-validation (SLCV) (Hastie et al. 2009). Generalization performance is estimated using *double-loop* cross-validation (DLCV), nesting and repeating SLCV to obtain realistic and reliable estimates (Spilka et al. 2017, Jonathan et al. 2000).

Following (Warrick et al. 2010, Dash et al. 2014), the decision rule threshold b is adjusted to achieve the largest Specificity (SP) for a Sensitivity (SE) above 0.7, i.e., the smallest number of False Positives (FP) for at least 70% of True Positives (TP).

4. Results: Sparse Learning, labor stages, cross-database evaluation

Using LDB only, Sections 4.1 and 4.2 first describe the benefits of sparse learning and assess differences between S_I and S_{II}. Then, Section 4.3 quantifies the ability of the decision obtained from LDB to generalize using the BDB database.

4.1. Single-feature acidosis prediction

4.1.1. Acidosis prediction performance. Table 2 reports acidosis prediction performance, for each feature independently and for S_I and S_{II} separately, in terms of achieved SP for a targeted SE above 0.70, balanced error rate $BER = (SP+SE)/2$, TP, FP, and Area under Curve (AUC) computed from the Receiver-Operator-Characteristic (ROC) curves (Fawcett 2006).

For S_I, Table 2 shows that features quantifying decelerations achieve the best performance, notably A_{dec}, T_{stress} and MAD_{dt} (FIGO-like) and E_{VLF} (spectral). Interestingly, LTV and STV—classically used to assess FHR variability—yield poor performance, while scale-free features H and c_1—robust estimates of FHR variability (Abry et al. 2013, Doret et al. 2015)—achieve satisfactory performance. This validates both the importance of variability for acidosis prediction and the intuition that variability should not be constructed on specific short or long time-scales, but should rather be based on the scale-free paradigm. Baseline level β_0 also yields satisfactory performance, while baseline trend β_1 surprisingly does not. In addition, spectral features on the LF and HF bands have poor individual power for acidosis prediction—confirming results in (Doret et al. 2015). Moreover, nonlinear scale-free feature c_2 also has satisfactory individual power, while higher-order non linear features $\{c_3, c_4\}$ show much poorer performance.

For S_{II}, Table 2 yields essentially the same conclusions, yet showing that decelerations are better accounted for by T_{stress}—quantifying the percentage of

‡ Performance achieved using data that were never used neither to tune (C, λ) nor to estimate $(\mathbf{w}, b)$.

Table 2. Fetal acidosis prediction: univariate performance

	S_I						S_{II}					
Feature:	AUC	SE	SP	BER	TP	FP	AUC	SE	SP	BER	TP	FP
β_0	.65	.72	.53	.63	21	477	.51	.70	.18	.44	19	600
β_1	.54	.72	.26	.49	21	760	.55	.70	.29	.50	19	523
$\#acc$	.50	.72	.17	.45	21	844	.55	.70	.30	.50	19	513
$\#dec$	.60	.72	.41	.56	21	608	.50	.70	.29	.50	19	518
A_{dec}	.71	.72	.59	.66	21	420	.56	.70	.36	.53	19	468
MAD_{dt}	.76	.72	.64	.68	21	363	.63	.70	.49	.60	19	371
T_{stress}	.73	.72	.69	.71	21	315	.71	.70	.58	.64	19	306
LTV	.53	.72	.28	.50	21	737	.52	.70	.22	.46	19	570
STV	.51	.72	.25	.49	21	767	.51	.70	.30	.50	19	517
LF/HF	.59	.72	.49	.61	21	516	.50	.70	.30	.50	19	517
E_{VLF}	.75	.72	.68	.70	21	329	.62	.70	.51	.60	19	363
E_{LF}	.60	.72	.38	.55	21	633	.55	.70	.46	.58	19	394
E_{HF}	.50	.72	.28	.50	21	738	.58	.70	.47	.59	19	389
α	.58	.72	.45	.59	21	563	.51	.70	.28	.49	19	530
h_{min}	.67	.72	.53	.63	21	481	.62	.70	.47	.59	19	390
c_1	.69	.72	.47	.60	21	541	.71	.70	.68	.69	19	236
c_2	.66	.72	.54	.63	21	475	.66	.70	.50	.60	19	366
c_3	.53	.72	.28	.50	21	731	.50	.70	.22	.46	19	573
c_4	.50	.72	.10	.41	21	916	.64	.70	.51	.61	19	358
H	.69	.72	.60	.66	22	403	.67	.70	.60	.65	19	295

underwent contractions—suggesting that too many or too close contractions alter the fetus' well-being. Also, and interestingly, variability is well accounted for by the multifractal feature c_1, which clearly outperforms the classical LTV and STV FIGO-like features or any spectral feature.

In summary, Table 2 shows that single-feature performances are satisfactory for a few FIGO-like features quantifying decelerations, and for scale-free features measuring variability. Yet, none of the features, used as stand-alone decision rule, yield outstanding performance, which motivates the investigation of the performance that can be obtained from joint use of multiple features to construct acidosis prediction rules.

4.1.2. Feature correlations. As a preliminary step, Fig. 2 reports the correlations amongst features. For S_I, it shows that features quantifying decelerations—MAD_{dt}, A_{dec} and T_{stress}—are strongly correlated. Further, they are strongly correlated with E_{VLF}, indicating an association of very low frequencies and decelerations. Scale-free features quantifying variability—$H, c_1, \text{h}_{\text{min}}$—are highly correlated, as theoretically expected (H and h_{min} are hence removed from multiple-feature analysis). However, they are only weakly correlated to STV and LTV—thus confirming their different performance. As expected, energies in frequency bands and LF/HF ratio are correlated (Doret et al. 2015). Baseline features do not correlate with any other feature, thus clearly carrying different information. Nonlinear features c_2, c_3, c_4 are weakly correlated

to linear features.

For S_{II}, observations are essentially identical. Yet, overall correlation is observed to be lower during S_{II}, than during S_I, indicating that during S_{II} features measure different aspects of FHR dynamics, while they are mostly related in S_I.

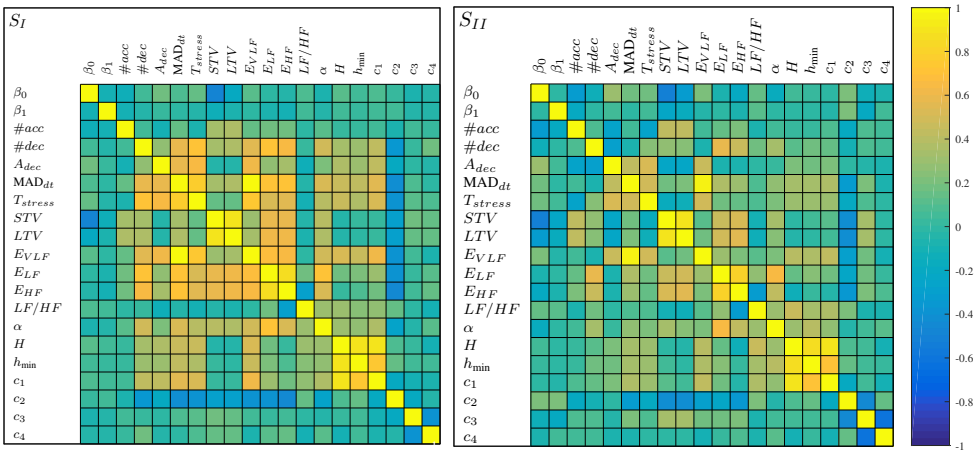

Figure 2. Feature correlation. Pairwise correlations for all pairs of features independently for S_I (left) and S_{II} (right).

4.2. Multiple-feature analysis and Sparse learning

4.2.1. Learning setup. Multiple-feature acidosis detection rules are investigated using the S-SVM methodology described in Section 3. Following preliminary results in (Spilka, Abry, Goncalves & Doret 2014, Spilka et al. 2016*b*), three decision rules are estimated for S_I only, for S_{II} only, and for S_I and S_{II} jointly. Hyper-parameter selection was done jointly for (C, λ); their optimal values are shown in Table 3.

Fig. 3 reports, for each setting, the estimated $\mathbf{w}$ as a function of the sparsity parameter $\log_2 C$. Hyper-parameter selection was achieved jointly for (C, λ), yet, for ease of exposition, Fig. 3 considers C only. The optimal λ are reported in Table 3 for the sake of completeness, and shown to be close to 0.5 in both cases, suggesting that the mere balance of miss-classifications by class sizes N_+/N_- is sufficient. Fig. 3 illustrates that the smaller C is, the smaller the number of selected features (i.e., the larger the number of $\mathbf{w}$ set to 0).

At the methodological level, Table 3, comparing SLCV and DLCV performances, shows that SLCV performances systematically overestimate those of DLCV, since the same data is used for training and parameter selection.

4.2.2. Stage I. For S_I, Fig. 3 (left) and Table 5 (top) show that optimal performance is achieved involving only a few features: 4 out of the 20 proposed ones. Interestingly, 3 of these are FIGO-like features: MAD_{dt}, T_{stress} and β_0—the first two associated with decelerations. Further, MAD_{dt}—characterizing deceleration duration and depth—plays a central role. Despite having low performance as a standalone feature, baseline level

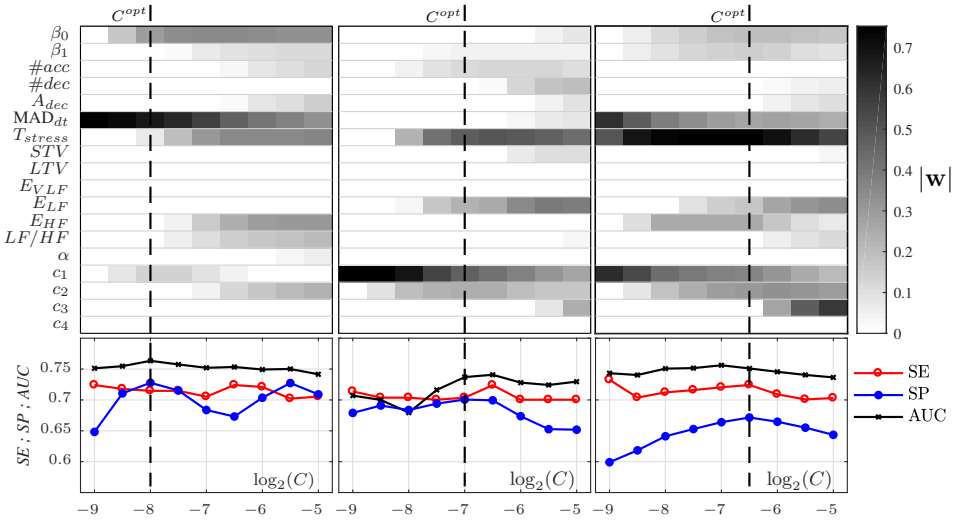

Figure 3. S-SVM performance. Feature selection (top row) and classification performance (bottom row) as function of the regularization parameter C, for each of the three sparse learning analyses applied to S_I only, S_{II} only and both stages jointly.

Table 3. Acidosis prediction performance – LDB Database. Performance are reported for DLCV (top) and SLCV (bottom) and different pairs of training / testing datasets.

Tr / Te	$\log_2 C$	λ	SE	SP	BER	#TP	#FN	#TN	#FP
S_I/S_I	-8	.49	.66	.74	.70	19	10	751	270
S_{II}/S_{II}	-6.5	.51	.59	.71	.65	16	11	520	215
S/S	-6	.47	.71	.66	.69	40	16	1160	596
S_I/S_I	-8	.48	.71	.73	.72	21	8	743	278
S_{II}/S_{II}	-7	.48	.72	.70	.71	20	7	514	221
S/S	-6.5	.47	.72	.67	.70	41	15	1179	577

β_0 is also involved, showing the relevance of this feature, whose nature is different, to complement the decision rule. In addition, variability is also involved in the multiple feature decision rule, yet using the scale-free feature c_1 rather than the LTV and STV ones. Table 5 indicates that acidotic fetuses, compared to healthy ones, are characterized by larger values of MAD_{dt} and T_{stress} (larger impact of decelerations), by a higher baseline β_0, and a larger c_1 (decrease in variability).

Table 3 (line 2) shows that this sparse and optimal decision permits to achieve a SE of 0.66 for a SP of 0.74 (hence $BER = 0.70$). Table 4 (top panel) reports clinical performance by obstetricians, showing satisfactory Specificity (SP) at the price of a low Sensitivity (SE), and is used as a benchmark. To ease comparisons, Table 4 (mid and bottom panels) also reports S-SVM performance, where the parameter λ has been tuned to match either the SE or the SP of the clinical benchmark, showing the nonnegligible benefit of S-SVM. However, this matched-tuning does not correspond to optimal sparse learning performance, as reported in Table 3. Further comparing Tables 3 and 4 clearly indicates that optimal Sparse Learning yields a substantial increase in the

Table 4. Fetal acidosis prediction. Performance for S_I and S_{II}, computed from clinical information (top panel, used as benchmark), and compared to those obtained from S-SVM classifier tuned to match the either the SE or the SP of the clinical benchmark (middle and bottom panels, respectively).

		SE	SP	BER	#TP	#FP	#FN	#TN
Clinical benchmark	S_I	.45	.79	.62	13	213	16	808
	S_{II}	.48	.79	.64	13	152	14	583
S-SVM, matched SE	S_I	.45	.89	.67	13	16	912	109
	S_{II}	.48	.86	.67	13	14	632	103
S-SVM, matched SP	S_I	.62	.79	.71	18	11	811	210
	S_{II}	.56	.79	.68	15	12	586	149

tradeoff between SE and SP for both S_I and S_{II}. Interestingly, other works based on different feature-selection procedures also concluded that a restricted number of well-selected features were to be preferred to decision rules involving too many features, cf., e.g., (Georgoulas et al. 2017).

In summary, classification in S_I strikingly selects a sparse decision rule with two (essentially one) features for decelerations, one for baseline, and one for variability—the two first being FIGO-like, the second being scale-free—and hence yielding a decision rule based on the three major pillars in FIGO criteria (baseline, variability, deceleration). Further and complementary comparisons are shared at `http://people.ciirc.cvut.cz/ spilkjir/Abry2018BPEXresults.html`.

4.2.3. Stage I vs. Stage II. Comparing Table 3 to Table 4 indicates that sparse learning improves on the clinical benchmark also for S_{II}. Further, Fig. 3 (middle) shows that optimal performance is obtained for a slightly lower level of sparsity (compared to S_I), hence involving 6 parameters, which further differ from those selected in S_I (cf. Table 5 (middle)). Decelerations are still involved, but now T_{stress} and E_{LF} become prominent: the frequency and closeness of decelerations become discriminative rather than their depth. A larger T_{stress} characterizes acidotic newborns, suggesting that too-frequent maternal pushing in Stage II may have negative consequences for the fetus well-being. Also, the number of accelerations $\#acc$ is selected and shows a decreased value for acidotic fetuses, confirming that accelerations in FHR remain a sign of good fetal health. Baseline almost does not contribute to the acidosis prediction in S_{II} (only marginally via β_1). Variability is also strongly selected, but it is now measured *both* through c_1 and c_2: An increased c_1 in acidotic fetuses suggests a lower overall variability, whereas an increased $|c_2|$ betrays sporadic burstiness and localized transient decreases of variability, within the overall decrease.

The fact that different features are selected for S_I and S_{II} indicates that temporal dynamics in each stage are different. Though lower than in S_I, acidosis prediction performance remains satisfactory (cf. Table 3). Interestingly, Table 3 also shows that

Table 5. Selected feature statistics. Median (maximum absolute deviation) per class for each selected feature in S_I only, S_{II} only, and both decision rules (S).

	name	**w**	Acidotic	Normal
S_I	MAD_{dt}	**.90**	20.3 (7.1)	10.1 (5.0)
	β_0	**.38**	156 (10)	147 (12)
	c_1	**.18**	0.65 (0.14)	0.54 (0.11)
	T_{stress}	**.08**	0.63 (0.13)	0.43 (0.20)
S_{II}	T_{stress}	**.63**	0.59 (0.09)	0.53 (0.13)
	c_1	**.62**	0.79 (0.12)	0.64 (0.13)
	$\lvert c_2\rvert$	**.32**	0.19 (0.09)	0.16 (0.07)
	E_{LF}	**.30**	36.2 (16.1)	39.7 (21.7)
	$\#acc$	**.14**	0.0 (0.6)	0.0 (0.9)
	$\beta_1(\cdot 1E5)$	**.05**	5.8 (94)	-7.5 (94)
S	T_{stress}	**.74**	0.61 (0.11)	0.49 (0.18)
	c_1	**.36**	0.71 (0.14)	0.58 (0.12)
	$\lvert c_2\rvert$	**.30**	0.15 (0.08)	0.10 (0.07)
	MAD_{dt}	**.27**	20.4 (6.7)	13.6 (5.8)
	E_{HF}	**.26**	5.14 (3.48)	5.20 (4.29)
	β_0	**.19**	153 (13)	147 (12)
	E_{LF}	**.17**	25.2 (15.9)	22.5 (19.7)
	$\beta_1(\cdot 1E5)$	**.15**	9.9 (74)	-2.1 (71)

the decrease in performance when comparing SLCV and DLCV is larger for StageII, thus suggesting that the learning stage is significantly more difficult, likely caused by a wider inter-individual variety in FHR temporal dynamics for StageII.

For the analysis of S_I and S_{II} jointly, Fig. 3 (right) shows that a much larger number of features is required, resembling a combination of those selected independently (cf. Table 5). This confirms that temporal dynamics in S_I and S_{II} are actually different, and that a decision rule that tries to be efficient on both stages somehow mixes them up, decreasing overall performance (cf. Table 3).

In summary, since clinicians very well know when the second stage started, there is no reason not to take advantage of that information by using individual, simpler decision rules that are tailored (*learned*) for each stage—thus providing a better interpretation.

4.3. Sparse Learning and cross-database evaluation

So far, generalization performance, that is performance that would be achieved if the learned decision rule was applied to data that were never used for the training, was evaluated following the classical, yet computationally intensive and practically demanding, DLCV procedure (Hastie et al. 2009), yielding the performance reported in Table 3. Access to two independent databases (LDB and BDB) permits here to evaluate generalization performance in an additional way, by *learning* the decision rules on the larger LDB and then applying it to the smaller BDB. Table 6 indicates that performance measurements thus achieved are comparable to those using DLCV

on the LDB alone. This indicates that the proposed decision rules have a very good generalization ability—a remarkable result since both databases differ in several respects (FHR recording technique, data quality, class imbalance, sampling frequency,... cf. Section 2). This suggests that S-SVM yields generalizable decision rules, and that such a robustness comes as a by-product of their sparsity (or simplicity). Table 6 further comforts the relevance of using independent decision rules for both stages: performance using independent rules is much better than for joint ones.

Table 6. Generalization performance. Learning on the LDB Database, performance evaluation on the BDB database.

Train/Eval	AUC	SE	SP	BER	TP	FN	TN	FP
S_I/S_I	.78	.64	.80	.72	9	5	263	67
S_{II}/S_{II}	.73	.58	.77	.68	7	5	89	27
S/S	.72	.42	.81	.62	11	15	361	85
S/S_I	.66	.21	.86	.54	3	11	283	47
S/S_{II}	.76	.67	.67	.67	8	4	78	38

5. Conclusions and perspectives

The present contribution has quantified the benefits and performance of sparse learning—as implemented by Sparse Support Vector Machines—for fetal acidosis prediction through intrapartum FHR analysis. It has shown that, despite the availability of a large number of features, decision rules only involving a few of them were favored.

Interestingly, decision rules essentially involved features associated to the three major groups of FHR characteristics underlying FIGO definitions: Baseline level, decelerations/accelerations, and variability. The present work also shows that FIGO-like features remain competitive compared to those devised from advanced statistical signal processing. However, as pointed out in previous works (Doret, Helgason, Abry, Gonçalvès, Gharib & Gaucherand 2011, Abry et al. 2013, Doret et al. 2015)), scale-free parameters H or c_1 (and to a lesser extent c_2) provide a far more robust assessment of FHR variability than FIGO's LTV or STV.

Sparse learning also yields decision rules that clearly differ for Stage I and Stage II—evidencing significant differences in their temporal dynamics. Given that the stage of delivery is a naturally available clinical information, there is a significant benefit in using it to design stage-specific decision rules.

Finally, it has been shown that sparse learning achieves very satisfactory generalized performance, in the sense that decision rules learned from one database can be satisfactorily applied to another, even if collected in a different hospital, with different acquisition devices and potentially slightly different clinical practice. This assessment is permitted by the use of several and really different databases, which is rarely reported in the scientific literature.

The satisfactory performances achieved by the proposed sparse learning decision rules, tailored to each stage and validated on an independent database, open the way toward prototype implementations aiming at clinical-practice experimentation.

Future works will also include the assessment of both feature automated computation and acidosis early detection with respect to data quality, notably with respect to the level of missing data in Stage II. This requires further work to devise routines that permits the robust computation of features and robust decision strategies when data quality is poor.

Acknowledgments

Work supported by Grant ANR-16-CE33-0020 MultiFracs.

Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum provided under the programme *Projects of Large Research, Development, and Innovations Infrastructures* (CESNET LM2015042) is greatly appreciated.

References

Abry P & Didier G 2018 *Bernoulli* **24**(2), 895–928.

Abry P, Roux S, Chudáček V, Borgnat P, Goncalves P & Doret M 2013 *in* '26th International Symposium on Computer-Based Medical Systems (CBMS)' pp. 1–6.

Abry P & Veitch D 1998 *IEEE Trans. on Info. Theory* **44**(1), 2–15.

Akselrod S, Gordon D, Ubel F A, Shannon D C, Berger A C & Cohen R J 1981 *Science* **213**(4504), 220–222.

Alfirevic Z, Devane D & Gyte G 2006 *Cochrane Database Syst Rev* **3**(3), CD006066.

Amer-Wåhlin I, Hellsten C, Norén H, Hagberg H, Herbst A, Kjellmer I, Lilja H, Lindoff C, Månsson M, Mårtensson L, Olofsson P, Sundström A & Maršál K 2001 *Lancet* **358**(9281), 534–538.

Ayres-de Campos D, Spong C Y, Chandraharan E & FIGO 2015 *Int J Gynaecol Obstet* **131**(1), 13–24.

Bach F, Jenatton R, Mairal J & Obozinski G 2012 *Foundations and Trends in Machine Learning* **4**(1), 1–106.

Bernardes J, Moura C, de Sa J P & Leite L P 1991 *J Perinat Med* **19**(1-2), 61–65.

Blackwell S, Grobman W A, Antoniewicz L, Hutchinson M & Gyamfi Bannerman C 2011 *Am J Obstet Gynecol* **205**(4), 378.e1–378.e5.

Blondel M, Seki K & Uehara K 2013 *J Mach Learn* **93**(1), 31–52.

Bron E et al. 2015 *Biomedical and Health Informatics, IEEE Journal of* **19**(5), 1617–1626.

Chandraharan E & Arulkumaran S 2007 *Best Pract Res Clin Obstet Gynaecol* **21**(4), 609–624.

Chierchia G, Pustelnik N, Pesquet J C & Pesquet-Popescu B 2016 *preprint* .

Chudáček V, Anden J, Mallat S, Abry P & Doret M 2014 *Biomedical Engineering, IEEE Transactions on* **61**(4), 1100–1108.

Chudáček V, Spilka J, Burša M, Janků P, Hruban L, Huptych M & Lhotská L 2014 *BMC Pregnancy Childbirth* **14**(1), 16.

Chudáček V, Spilka J, Janků P, Koucký M, Lhotská L & Huptych M 2011 *Physiological Measurement* **32**, 1347–1360.

Combettes P L & Wajs V R 2005 *Multiscale Model Simul* **4**(4), 1168–1200.

Costa A, Ayres-de Campos D, Costa F, Santos C & Bernardes J 2009 *Am J Obstet Gynecol* **201**(5), 464.e1–464.e6.

Costa M, Goldberger A L & Peng C K 2002 *Phys Rev Lett* **89**(6), 068102.

Czabanski R, Jezewski J, Matonia A & Jezewski M 2012 *Expert Systems with Applications* **39**(15), 11846–11860.

Dash S, Quirk J & Djuric P 2014 *IEEE Trans Biomed Eng* **61**(11), 2796–2805.

Doret M, Helgason H, Abry P, Gonçalvès P, Gharib C & Gaucherand P 2011 *Am J Perinatol* **28**(4), 259.

Doret M, Massoud M, Constans A & Gaucherand P 2011 *Eur J Obstet Gynecol Reprod Biol* **156**(1), 35–40.

Doret M, Spilka J, Chudáček V, Gonçalves P & Abry P 2015 *PLoS ONE* **10**(8), e0136661.

Echeverria J C, Hayes-Gill B R, Crowe J A, Woolfson M S & Croaker G D H 2004 *Physiol Meas* **25**(3), 763–774.

Fawcett T 2006 *Pattern Recognition Letters* **27**(8), 861–874.

FIGO 1986 *International Journal of Gynecology & Obstetrics* **25**, 159–167.

Francis D P, Wilson K, Georgiadou P, Wensel R, Davies, Ceri Davies L, Coats A & Piepoli M 2002 *J Physiol* **542**(Pt 2), 619–629.

Frasch M G, Xu Y, Stampalija T et al. 2014 *Physiol meas* **35**(12), L1.

Georgieva A, Papageorghiou A T, Payne S J, Moulden M & Redman C W G 2014 *BJOG* **121**(7), 889–894.

Georgieva A, Payne S J, Moulden M & Redman C W G 2013 *Neural Computing and Applications* **22**(1), 85–93.

Georgoulas G, Karvelis P, Spilka J, Chudáček V, Stylios C D & Lhotská L 2017 *Health and Technology* .
URL: *http://dx.doi.org/10.1007/s12553-017-0201-7*

Gonçalves H, Rocha A P, de Campos D A & Bernardes J 2006 *Med Biol Eng Comput* **44**(10), 847–855.

Granero-Belinchon C, Roux S G, Abry P, Doret M & Garnier N B 2017 *Entropy* **19**(12).

Haritopoulos M, Illanes A & Nandi A 2016 Springer International Publishing Cham pp. 1187–1192.

Hastie T, Tibshirani R & Friedman J 2009 *The Elements of Statistical Learning* Springer Series in Statistics 2nd ed edn Springer US.

Hruban L, Spilka J, Chudáček V, Janků P, Huptych M, Burša M, Hudec A, Kacerovský M, Koucký M, Procházka M, Korečko V, Seget'a J, Šimetka O, Měchurová A & Lhotská L 2015 *Journal of Evaluation in Clinical Practice* **21**(4), 694–702.

Jonathan P, Krzanowski W & McCarthy W 2000 *Stat Comput* **10**(3), 209–229.

Laar J V, Porath M M, Peters C H L & Oei S G 2008 *Acta Obstet Gynecol Scand* **87**(3), 300–306.

Laporte L, Flamary R, Canu S, Déjean S & Mothe J 2014 *IEEE Trans Neural Netw Learn Syst* **25**(6), 1118–1130.

Lim J, Kwon J Y, Song J, Choi H, Shin J C & Park I Y 2014 *Early Hum Dev* **90**(2), 81–85.

Magenes G, Signorini M G & Arduini D 2000 *in* 'Proc. IEEE-INNS-ENNS International Joint Conference on Neural Networks IJCNN 2000' Vol. 3 pp. 637–641.

Magenes G, Signorini M G, Ferrario M, Pedrinazzi L & Arduini D 2003 *in* 'Proc. 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society' Vol. 3 pp. 2295–2298 Vol.3.

Manolakis D, Igle V & Kogon S 2005 *Statistical and Adaptive Signal Processing* Artech House Boston.

Nunes I, de Campos D A, Ugwumadu A, Amin P, Banfield P, Nicoll A, Cunningham S, Costa-Santos P S P C & Bernardes J 2017 *Obstet Gynecol.* **129**(1), 83–90.

Parer J T, King T, Flanders S, Fox M & Kilpatrick S J 2006 *J Matern Fetal Neonatal Med* **19**(5), 289–294.

Porta A, Bari V, Bassani T, Marchi A, Tassin S, Canesi M, Barbic F & Furlan R 2013 *Conf Proc IEEE Eng Med Biol Soc* **2013**, 5045–5048.

Samorodnitsky G & Taqqu M 1994 *Stable non-Gaussian random processes* Chapman and Hall New York.

Siira S M et al. 2005 *BJOG* **112**(4), 418–423.

Siira S, Ojala T H, Vahlberg T J, Rosn K G & Ekholm E M 2013 *Early Hum Dev* **89**(9), 739–742.

Soguero-Ruiz C et al. 2015 *Biomedical and Health Informatics, IEEE Journal of* **PP**(99), 1–1.

Spilka J, Abry P, Goncalves P & Doret M 2014 *in* 'Computing in Cardiology Conference (CinC), 2014' pp. 777–780.

Spilka J, Chudáček V, Janků P, Hruban L, Burša M, Huptych M, Zach L & Lhotská L 2014 *Journal of Biomedical Informatics* **51**(0), 72–79.

Spilka J, Chudáček V, Koucký M, Lhotská L, Huptych M, Janků P, Georgoulas G & Stylios C 2012 *Biomedical Signal Processing and Control* **7**(4), 350–357.

Spilka J, Frecon J, Leonarduzzi R, Pustelnik N, Abry P & Doret M 2017 *IEEE J Biomed and Health Inform* **21**, 664–671.

Spilka J, Roux S, Garnier N, Abry P, Goncalves P & Doret M 2014 *in* 'Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE' pp. 2813–2816.

Spilka J et al. 2016*a* *in* 'XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016: MEDICON 2016'.

Spilka J et al. 2016*b* *in* 'Proc. Int. Workshop Biosignal Interpretation (BSI)' Osaka, Japan.

Warrick P, Hamilton E, Precup D & Kearney R 2010 *IEEE Trans Biomed Eng* **57**(4), 771–779.

Wendt H, Abry P & Jaffard S 2007 *IEEE Signal Proc. Mag.* **24**(4), 38–48.

Xu L, Redman C W, Payne S J & Georgieva A 2014 *Physiol meas* **35**(7), 1357–71.

Appendix A. Abbreviation and Symbol list

Table A1. Acronyms and features list

General		FHR	Fetal Heart Rate
		FIGO	International Federation of Gynecology and Obstetrics
		SVM	Support Vector Machine
		S-SVM	Sparse Support Vector Machine
		LDB	Lyon Database
		BDB	Brno Database
		t_I	Duration of first stage
		t_{II}	Duration of second stage
		S	Set of all records in a database
		S_I	Set of records with $t_{II} \leq 15$ min
		S_{II}	Set of records with $t_{II} > 15$ min
Features	FIGO	β_0, β_1	Intercept and slope of linear fit of baseline
		LTV	Long-term variability
		STV	Short-term variability
		$\#acc, \#dec$	Number of accelerations and decelerations
		MAD_{dt}	Average depth of decelerations
		T_{stress}	Average duration of decelerations
		A_{dec}	Average area of decelerations
	Spectr.	VLF, LF, HF	Very low-, low-, and high-frequency ranges of FHR
		E_{VLF}, E_{LF}, E_{HF}	Energy at each frequency range
		α	Spectral index
	MF	H	Hurst parameter
		$h_{min}, \{c_m\}_{m=1,\ldots,4}$	Multifractal features
Performance		SLCV	Single-Loop Cross-Validation
		DLCV	Double-Loop Cross-Validation
		SP	Specificity
		SE	Sensitivity
		FP, TP	Number of False and True Positives
		BER	Balanced Error Rate
		AUC	Area under ROC curve