

Intrapartum Fetal Heart Rate Classification from Trajectory in Sparse SVM feature space

J. Spilka^{1,2}, J. Frecon¹, R. Leonarduzzi¹, N. Pustelnik¹, P. Abry¹, M. Doret³

Abstract—Intrapartum fetal heart rate (FHR) constitutes a prominent source of information for the assessment of fetal reactions to stress events during delivery. Yet, early detection of fetal acidosis remains a challenging signal processing task. The originality of the present contribution are three-fold: multiscale representations and wavelet leader based multifractal analysis are used to quantify FHR variability ; Supervised classification is achieved by means of Sparse-SVM that aim jointly to achieve optimal detection performance and to select relevant features in a multivariate setting ; Trajectories in the feature space accounting for the evolution along time of features while labor progresses are involved in the construction of indices quantifying fetal health. The classification performance permitted by this combination of tools are quantified on a intrapartum FHR large database ($\simeq 1250$ subjects) collected at a French academic public hospital.

I. INTRODUCTION

1) *Intrapartum fetal heart rate monitoring*: Fetal surveillance and cardiotocogram (CTG) monitoring during labor is current clinical practice aiming to assess the health status and well-being of the fetus. It notably helps in early detection of fetal acidosis, an important stake with respect to subsequent fetal and neonatal mortality and morbidity [1]. Fetal heart rate (FHR) is mostly evaluated by obstetricians using FIGO type guidelines, which are essentially based on 3 types of features: baseline evolution, acceleration and deceleration shape and occurrence rates, and variability. Suspicion of fetal acidosis leads to a decision of operative deliveries. While the strict use of FIGO criteria leads to a high sensitivity (correct detection of acidotic fetuses), it also comes with a low specificity (incorrect decision of operative delivery for a normal fetus). Therefore, significant research efforts gathering both medical and academic teams are devoted to better characterizing FHR, with an aim of improving specificity while maintaining high sensitivity (cf. e.g. [2]).

2) *Related works*: Beyond computerized versions of FIGO features, intrapartum FHR has been analyzed with a large variety of features. Linear features, computed from spectral analysis, have been massively used (cf. e.g., [3]). A variety of non linear features have also been tested [4], ranging from entropy rates (cf. e.g., [5]) to PRSA [6]. Multiscale representations (and multifractal analysis) computing jointly linear and non linear features have also been shown to yield satisfactory classification performance [7]–[9].

Supported by ANR grants AMATIS #112432, FETUSES #18535, and MZCR grant #NT11124-6/2010

¹ CNRS, Physics Department, ENS de Lyon, France

patrice.abry@ens-lyon.fr, firstname.lastname@ens-lyon.fr

² Czech Institute of Informatics, Robotics, and Cybernetics, Czech Technical University in Prague jiri.spilka@ciirc.cvut.cz

³ Femme-Mère-Enfant Hospital, Lyon, France

Supervised classification is common practice to determine the combination of linear and non linear features leading to acidosis performance (cf. e.g. [2], [10], [11]). Yet, machine learning strategies often concentrate on a single time window before delivery and thus do not fully exploit the long term evolution of FHR as labor progresses (see, a contrario, [8], [12]–[14] for preliminary attempts).

3) *Goals, contributions and outline*: In this context, the global goal of the present contribution is to use FHR trajectory in a feature space to assess the fetus health status. To that end, classical (improved and automated) FIGO features for baseline and deceleration characterization are complemented with wavelet leader based multifractal linear and non linear features providing practitioners with a detailed analysis of FHR variability. Features are defined and detailed in Section III-A. Further, supervised classification is conducted using Sparse Support Vector Machine (Sparse-SVM) aiming jointly to define the optimal performance frontier of the healthy domain and to perform feature selection (cf. Section III-B). Finally, FHR trajectories in feature space are used to quantify the fetus health status (cf. Section IV-B). This FHR analysis strategy is illustrated at work on a large (over 1000 subjects) database of FHR records collected in a French public academic hospital, which is described in Section II. Results are presented and discussed in Section IV.

II. DATABASE

1) *Data*: Recordings were collected at the public academic French Hospital Femme-Mère-Enfant, from 2000 to 2010, as daily routine cardiotocogram (CTG) monitoring, using STAN S21 or S23 devices with 12-bit resolution and 500 Hz sampling rate. The database consists of 3049 recordings together with clinical information for each woman and neonate, systematically collected by obstetricians in charge of delivery.

In the present study, analysis was performed on the first stage of labor. It was required that the end of FHR in the first stage was less than 20 minutes away from actual delivery, i.e. to the umbilical artery pH value measurement. This led to the selection of 1288 records, from which 37 were considered as *acidotic* ($\text{pH} \leq 7.05$) and 1251 as *normal* ($\text{pH} > 7.05$).

Table I reports, as an acidosis detection clinical benchmark, the confusion matrix corresponding to the clinicians' decision to perform an operative delivery because of fetal hypoxia. This clinical decision is related to $\text{pH} \leq 7.05$ and shows good specificity with yet low sensitivity.

2) *Signals and analysis*: From recorded fetal ECGs, the RR intervals were extracted by the manufacturer (Neoventa

TABLE I

ACIDOSIS DETECTION CLINICAL BENCHMARK. CLINICAL DECISION OF OPERATIVE DELIVERY FOR SUSPECTED FETAL HYPOXIA.

	SE	SP	#TP	#FP
fetal hypoxia	.49	.78	18	274

Medical, Molndal, Sweden). Traditionally, in heart rate variability analysis, RR intervals are interpolated using cubic spline interpolation to form regularly sampled beat-per-minute time series $X(t)$. As FHR does not contain energy at frequencies higher than 3 Hz, a sampling frequency of 10 Hz was chosen. Supervised classification was conducted on the last 20 minutes before delivery. Trajectories were obtained from the last 120 minutes before delivery.

III. FEATURES AND CLASSIFICATION

A. Features

1) *FIGO enhanced and automated features*: Among the clinical FIGO criteria, it has been chosen to assess baseline and deceleration overall impact. The floating baseline estimation $B(t)$, defined and detailed in [15], [16], is computed from $X(t)$ to characterize the short-time evolution of FHR: The local baseline within the 20-min window is then estimated as linear regression $\beta_0 + \beta_1 B(t)$, whose intercept β_0 is used as first feature. The impact of decelerations is quantified by the median absolute deviation (MAD) of FHR, after baseline subtraction

$$\text{MAD}_{\text{dtrd}} = \text{MAD}(X(t) - B(t)). \quad (1)$$

2) *Variability from multiscale analysis*: FHR variability is considered a key element of fetal well-being evaluation [17] and is classically quantified using the notions of long- or short-term variabilities. In the present work, and following our previous contributions [7], [9], [16], the concepts of multiscale representations and multifractal analysis for performing a refined analysis of FHR variability is used.

Multiscale analysis classically relies on wavelet coefficients, $d_X(j, k) = \langle \psi_{j,k} | X \rangle$, obtained from a collection $\{\psi_{j,k}(t) = 2^{-j} \psi(2^{-j}t - k)\}_{(j,k) \in \mathbb{N}^2}$ of dilated (to scale 2^j) and translated (to position $2^j k$) templates of the mother wavelet ψ , cf. e.g. [18]. Wavelet coefficients permit to measure the uniform Hölder regularity h_{\min} as [19]

$$\sup_k, \text{ for fixed } j |d_X(j, k)| \simeq K 2^{j h_{\min}}. \quad (2)$$

Further, for self-similar processes, such as fractional Brownian motion, commonly used to model HRV, the sample moments (of order $q > 0$) display power law behaviors as function of scale [19]

$$S(q, j) = \frac{1}{n_j} \sum_{k=1}^{n_j} |d_X(j, k)|^2 \simeq K_q 2^{j q H}, \quad (3)$$

with n_j the number of $d_X(j, k)$ available at scale 2^j .

It is now well accepted that self-similar models and parameter H alone cannot fully account for the richness of scaling encountered in real-world data and in HR notably (cf. e.g.,

[7], [20]). Multifractality enriches scaling characterization by permitting the scaling exponents in Eq. (3) to depart from the linear behavior in q , qH , and to be replaced by a concave function $\zeta(q)$. However, a relevant estimation of concave $\zeta(q)$ for all qs require the replacement of wavelet coefficients with wavelet leaders, consisting of multiscale quantities that better capture the fine scale fluctuations of regularity in data by scanning all details finer than the chosen analysis scale [19]. The *wavelet leaders* are defined as local suprema of (fractionally integrated) wavelet coefficients, taken within a narrow temporal neighborhood and for all finer scales

$$L_X^{(\gamma)}(j, k) := \sup_{\lambda' \subset 3\lambda} 2^{j'\gamma} |d_X(\lambda')|. \quad (4)$$

with $\lambda_{j,k} = [k2^j, (k+1)2^j)$ and $3\lambda_{j,k} = \bigcup_{m \in \{-1, 0, 1\}} \lambda_{j, k+m}$ [19]. The fractional integration parameter $\gamma \geq 0$ is chosen to ensure minimal regularity (cf. [19] and references therein for theoretical developments on multifractal analysis).

Multifractal properties are efficiently characterized by representations based on the log-leaders: $\ln L_X^{(\gamma)}(j, \cdot)$

$$C_p^{(\gamma)}(j) \equiv \text{Cum}_p \ln L_X^{(\gamma)}(j) \simeq c_p^0 + c_p \ln 2^j, \quad (5)$$

with the c_p related to $\zeta(q)$ (and hence to the multifractal spectrum) as $\zeta(q) \equiv \sum_{p \geq 1} c_p q^p / p!$. This polynomial expansion shows that the leading coefficient c_1 is closely related to H [7], [19]. c_2, c_3, c_4 associated to the evolution along scales of respectively the variance, skewness and kurtosis of $\ln L^{(\gamma)}(j)$, provide information beyond correlation and are thus non linear features.

B. Supervised classification

1) *Sparse Support Vector Machines*: Sparse-SVM have recently been proposed to achieve jointly state of the art classification and feature selection (cf. e.g., [21]). Estimation of the weights in the function, $D(\mathbf{x}) = \text{sign}(\mathbf{w}^\top \mathbf{x} + b)$, relies on the replacement of the classical L^2 -norm regularization by L_1 -norm soft regularization, leading to the following optimization problem

$$(\hat{\mathbf{w}}, \hat{b}) \in \arg \min_{\mathbf{w} \in \mathbb{R}^p, b \in \mathbb{R}} C \sum_{i=1}^N \max(0, 1 - y_i(\mathbf{w}^\top \mathbf{x}_i + b))^2 + \|\mathbf{w}\|_1,$$

where $y \in \{-1, 1\}$ is the class label (normal, acidotic status) and C is the regularization constant controlling the trade-off between sparsity and data fidelity, formulated with the square hinge loss function: Small C promotes sparsity. This is solved using a Forward-Backward Splitting Algorithm involving proximity operators (see e.g. [22]).

Performance is quantified with the traditional measures of sensitivity (SE), related to the number of true positive $\#TP$, specificity (SP), related to the number of false positives $\#FP$, and the area under receiver-operation-characteristic (ROC) curve (AUC). AUC is estimated by varying the offset b . SP is reported for an arbitrarily targeted SE of 0.7 for eased comparisons amongst the various classification settings. Performance is estimated for regularization parameter $C \in 2^{\{-10, -9, \dots, -4\}}$ using a classical cross validation leave-one-minority-out scheme (in the current settings 37-fold stratified cross-validation).

TABLE II
UNIVARIATE PERFORMANCE.

Feature	AUC	SE	SP	#TP	#FP
β_0	.65	.70	.54	26	580
MAD_{dtrd}	.74	.70	.60	26	496
H	.72	.70	.68	26	435
h_{min}	.69	.70	.51	26	615
c_1	.72	.70	.57	26	543
c_2	.62	.70	.45	26	694
c_3	.54	.70	.35	26	809
c_4	.56	.70	.34	26	825

For robustness, outliers are removed by replacing feature values that fall outside the interval $[Q_1 - 3 \cdot IQR, Q_3 + 3 \cdot IQR]$, where $IQR = Q_3 - Q_1$ and Q_1 and Q_3 denote the first and third quantiles. Further, features are individually standardized before their use in Sparse-SVM.

IV. RESULTS AND DISCUSSIONS

A. Univariate analysis

The feature vector used here reads: $\mathbf{x} = (\beta_0, MAD_{\text{dtrd}}, H, h_{\text{min}}, c_1, c_2, c_3, c_4)$. For the benchmark and comparisons, univariate analysis is conducted applying Sparse-SVM to each single independently. Table II shows that individual use of features MAD_{dtrd} , H , or c_1 provides satisfactory AUC, above 0.7. Thus, decelerations and variability, as measured by multiscale features, contribute to fetal acidosis detection. Yet, individually, each feature produces very low specificity for a targeted sensitivity (here 0.7), thus prompting for multivariate analysis.

B. Multivariate analysis and feature space trajectory

1) *Sparse-SVM on last 20 min*: Sparse-SVM is applied to the last 20 minute window only. Table III reports classification performance for different values of the regularization parameter C , with corresponding weights in Fig. 1. They show that, for the lowest C , the sole parameter MAD_{dtrd} is selected, thus confirming the predominance of decelerations in acidosis detection. They also illustrate that, as expected, increasing C decreases sparsity and also increases performance. Optimal performance is achieved for $C = 2^{-8}$, with features MAD_{dtrd} , H and β_0 selected. Interestingly, it indicates that the three typical patterns used in clinical practice, decelerations, variability, and baseline are useful jointly for acidosis detection. Also, it shows that c_1 with better univariate performance than β_0 , is not used in Sparse-SVM multivariate classification, likely because of strong correlation to H [19]. Fig. 2 displays feature space projections and decision boundary (with $C = 2^{-8}$) for the three best features. It shows that the larger values of each feature are associated with acidotic subjects, thus corresponding to local baseline increase, deep decelerations, and low variability.

2) *Trajectory in feature space*: The decision boundary (for $C = 2^{-8}$) defining the healthy domain learned from the last 20 minutes window is used to assess the evolution of features along time. Features are computed for 120 minutes prior to delivery in 20-minute-long windows with 5 min overlap. For each subject i and each time window t_k ,

TABLE III
SPARSE SVM PERFORMANCE FOR LAST 20 MINUTES.

$\log_2 C$	AUC	SE	SP	#TP	#FP
-10	0.74	0.70	0.63	26	462
-9	0.75	0.70	0.72	26	353
-8	0.78	0.70	0.75	26	315
-7	0.78	0.70	0.73	26	340
-6	0.79	0.70	0.72	26	349
-5	0.78	0.70	0.71	26	364
-4	0.78	0.70	0.70	26	376

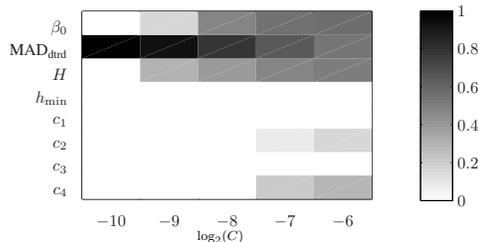


Fig. 1. Weight repartition as a function of C .

distance to the decision hyperplane is computed as $d_i(t_k) = y_i(\mathbf{w}^T \mathbf{x}_i(t_k) + b)$ and the cumulated distance $r_i = \sum_t d_i(t_k)$ quantifies the amount of time spent outside the healthy domain. Fig. 3 displays FHR trajectories in feature space and corresponding $d_i(t_k)$ and $r_i(t_k)$ for healthy and acidotic subjects. It illustrates that normal fetuses spent most of time in the healthy domain with sporadic excursion into the acidotic domain, while acidotic subjects leave the healthy domain at some time and then essentially remain outside.

Trajectory based performance is computed by comparing $r_i(0)$, i.e. at the end of the trajectory, against a threshold chosen to achieve the targeted 0.7 SE. Performance is reported in Table IV for each different sparsity level C . It confirms that use of the FHR time evolution yields further significant improvement in acidosis detection, compared to using only the last 20 minutes, with a specificity increased to 0.77 (283 FPs) for the targeted specificity 0.70 (26 TPs). Better performance is achieved for lower sparsity $C = 2^{-6}$, compared to $C = 2^{-8}$ for the last 20-minute window, cf. Table III. Interestingly, this involves the contribution of nonlinear features c_2 and c_4 to characterize variability. Though c_2 and c_4 do not show interesting univariate performance, they do help in time evolution characterization: In other words, nonlinear information in FHR, tracked over time, contributes to the discrimination between healthy and acidotic fetuses.

TABLE IV
SPARSE SVM PERFORMANCE FROM TRAJECTORY.

$\log_2 C$	AUC	SE	SP	#TP	#FP
-10	0.76	0.70	0.70	26	370
-9	0.77	0.70	0.68	26	404
-8	0.79	0.70	0.73	26	338
-7	0.78	0.70	0.76	26	303
-6	0.78	0.70	0.77	26	283
-5	0.77	0.70	0.77	26	287
-4	0.77	0.70	0.75	26	312

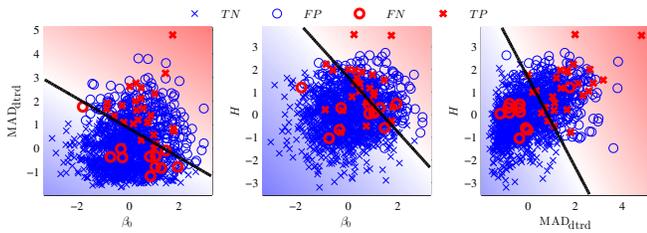


Fig. 2. Feature space projections with decision boundary.

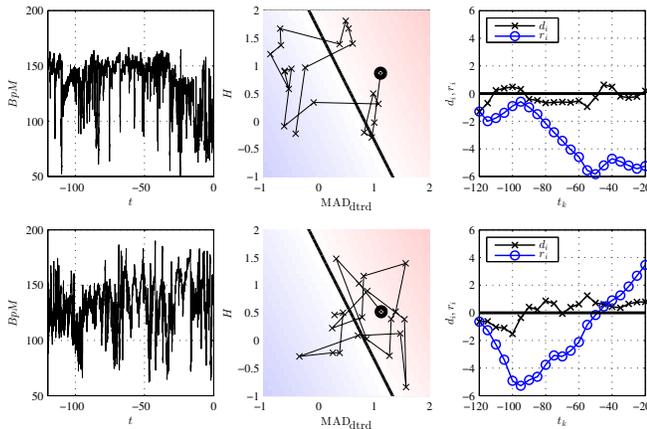


Fig. 3. Trajectories in feature space for healthy (top) and acidotic (bottom) subjects. From left to right: FHR, feature space with decision boundary, $d_i(t_k)$ and $r_i(t_k)$.

V. CONCLUSIONS

The contributions of this work are three-fold: multiscale representations and multifractal analysis for FHR variability characterization, Sparse-SVM for joint feature selection and good performance, and trajectories in the feature space to assess fetal health. Our results show remarkable performance compared to current clinical practice (cf. Tables I and IV: SP for Sparse-SVM compares to that achieved by clinicians while SE is increased from 0.49 to 0.70).

The notion of feature space trajectory takes advantage of the particularities of the powerful Sparse-SVM: It not only uses less features than classical SVM, removing redundant information and making the notion of trajectory easier to handle, but also constructs solutions in the primal space, thus permitting interpretations of the relative importance of features. Instead, the classical SVM, usually uses a nonlinear mapping (via the kernel trick) that makes the interpretation of involved features and trajectories much more difficult.

This study also confirms that the three important concepts of FHR (basal heart rate, decelerations, and variability), used in daily clinical practice, jointly contribute to fetal acidosis detection performance. Non linear features c_2 and c_4 associated to refined variability characterization also contribute to improve classification performance when trajectories are used.

REFERENCES

[1] E. Chandraharan and S. Arulkumaran, "Prevention of birth asphyxia: responding appropriately to cardiotocograph (ctg) traces," *Best Pract.*

Res. Clin. Obstet. Gynaecol., vol. 21, pp. 609–624, 2007.

[2] B.D. Fulcher, A.E. Georgieva, C.W.G. Redman, and N.S. Jones, "Highly comparative fetal heart rate analysis," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, 28 2012-sept. 1 2012, pp. 3135–3138.

[3] J. Van Laar, M. Porath, C. Peters, and S. Oei, "Spectral analysis of fetal heart rate variability for fetal surveillance: review of the literature," *Acta obstetrica et gynecologica Scandinavica*, vol. 87, no. 3, pp. 300–306, 2008.

[4] M. Signorini, G. Magenes, S. Cerutti, and D. Arduini, "Linear and nonlinear parameters for the analysis of fetal heart rate signal from cardiotocographic recordings," *IEEE Trans Biomed Eng.*, vol. 50, no. 3, pp. 365–374, Mar 2003.

[5] M. Costa, A. L. Goldberger, and C-K. Peng, "Multiscale entropy analysis of biological signals," *Phys Rev E Stat Nonlin Soft Matter Phys.*, vol. 71, no. 2 Pt 1, pp. 021906, Feb 2005.

[6] A. Georgieva, A. T. Papageorghiou, S. J. Payne, M. Moulden, and C W G. Redman, "Phase-rectified signal averaging for intrapartum electronic fetal heart rate monitoring is related to acidemia at birth," *BJOG*, vol. 121, no. 7, pp. 889–894, Jun 2014.

[7] M. Doret, H. Helgason, P. Abry, P. Goncalves, C. Gharib, and P. Gaucherand, "Multifractal analysis of fetal heart rate variability in fetuses with and without severe acidosis during labor," *Am. J. Perinatol.*, vol. 28, no. 4, pp. 259, 2011.

[8] V. Chudáček, J. Anden, S. Mallat, P. Abry, and M. Doret, "Scattering Transform for Intrapartum Fetal Heart Rate Variability Fractal Analysis: A Case-Control Study," *IEEE Trans. on Biomed. Eng.*, vol. 61, no. 4, pp. 1100–1108, April 2014.

[9] R. Leonarduzzi, J. Spilka, H. Wendt, S. Jaffard, M. Torres, P. Abry, and M. Doret, "p-leader based classification of first stage intrapartum fetal HRV," in *VI Congreso Latinoamericano de Ingeniería Biomédica (CLAIB)*, 2014, pp. 669–672.

[10] J. Spilka, V. Chudáček, M. Koucký, L. Lhotská, M. Huptych, P. Janků, G. Georgoulas, and C. Stylios, "Using nonlinear features for fetal heart rate classification," *Biomedical Signal Processing and Control*, vol. 7, no. 4, pp. 350–357, 2012.

[11] G. Georgoulas, C. D. Stylios, and P. P. Groumpos, "Predicting the risk of metabolic acidosis for newborns based on fetal heart rate signal classification using support vector machines," *IEEE Trans Biomed Eng.*, vol. 53, no. 5, pp. 875–884, May 2006.

[12] P.A. Warrick, E.F. Hamilton, D. Precup, and R.E. Kearney, "Classification of normal and hypoxic fetuses from systems modeling of intrapartum cardiotocography," *IEEE Trans. on Biomed. Eng.*, vol. 57, no. 4, pp. 771–779, 2010.

[13] H. Helgason, P. Abry, P. Goncalves, C. Gharib, P. Gaucherand, and M. Doret, "Adaptive Multiscale Complexity Analysis of Fetal Heart Rate," *IEEE Trans Biomed Eng.*, vol. 58, pp. 2186–2193, Mar 2011.

[14] S. Dash, J.G. Quirk, and P.M. Djuric, "Fetal heart rate classification using generative models," *IEEE Trans Biomed Eng.*, vol. 61, no. 11, pp. 2796–2805, Nov 2014.

[15] V. Chudáček, J. Spilka, P. Janků, M. Koucký, L. Lhotská, and M. Huptych, "Automatic evaluation of intrapartum fetal heart rate recordings: A comprehensive analysis of useful features," *Physiological Measurement*, vol. 32, pp. 1347–1360, 2011.

[16] P. Abry, S. Roux, V. Chudáček, P. Borgnat, P. Goncalves, and M. Doret, "Hurst Exponent and IntraPartum Fetal Heart Rate: Impact of Decelerations," in *26th Int. Symp. on Computer-Based Medical Systems*, 2013, pp. 1–6.

[17] A. Ugwumadu, "Understanding cardiotocographic patterns associated with intrapartum fetal hypoxia and neurologic injury," *Best Pract Res Clin Obstet Gynaecol*, vol. 27, no. 4, pp. 509–536, Aug 2013.

[18] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, 1998.

[19] H. Wendt, P. Abry, and S. Jaffard, "Bootstrap for empirical multifractal analysis," *IEEE Signal Proc. Mag.*, vol. 24, no. 4, pp. 38–48, 2007.

[20] L. Ivanov, P. Amaral, A. Goldberger, S. Havlin, M. Rosenblum, Z. Struzik, and H. Stanley, "Multifractality in human heartbeat dynamics," *Nature*, vol. 399, no. 6735, pp. 461–465, 1999.

[21] L. Laporte, R. Flamary, S. Canu, S. Déjean, and J. Mothe, "Nonconvex regularizations for feature selection in ranking with sparse svm," *IEEE Trans. Neural Networks and Learning Syst.*, vol. 25, no. 6, pp. 1118–1130, 2014.

[22] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.