Nearest-Neighbor based Wavelet Entropy Rate Measures for Intrapartum Fetal Heart Rate Variability

J. Spilka[1], S. G. Roux[1], N. B. Garnier[1], P. Abry[1], P. Goncalves[2], and M. Doret[3]

***Abstract*— The interpretation and analysis of intrapartum fetal heart rate (FHR), enabling early detection of fetal acidosis, remains a challenging signal processing task. The ability of entropy rate measures, amongst other tools, to characterize temporal dynamics of FHR variability and to discriminate non-healthy fetuses has already been massively investigated. The present contribution aims first at illustrating that a k-nearest neighbor procedure yields estimates for entropy rates that are robust and well-suited to FHR variability (compared to the more commonly used correlation-integral algorithm). Second, it investigates how entropy rates measured on multiresolution wavelet and approximation coefficients permit to improve classification performance. To that end, a supervised learning procedure is used, that selects the time scales at which entropy rates contribute to discrimination. Significant conclusions are obtained from a high quality scalp electrode database of nearly two thousands subjects collected in a French public university hospital.**

I. Introduction

Intrapartum fetal heart rate monitoring. Intrapartum fetal heart rate monitoring is mostly performed by cardiotocography, simultaneous recording of fetal heart rate and uterine contractions. Fetal heart rate variability (F-HRV) serves as a direct source of information about fetal behavior and reaction to the stress induced during delivery. Monitoring intrapartum F-HRV allows clinicians to timely and appropriately intervene to prevent adverse long term sequels caused by intrapartum asphyxia, such as neuro-development disability, cerebral palsy, and neonatal encephalopathy [1]. Intervention decision essentially relies on the visual inspection of FIGO-defined features, such as base rate and variability [2].

Related works. F-HRV complex temporal dynamics have been analyzed using numerous tools already involved in adult HRV analysis, such as spectrum estimation [3], nonlinear analysis [4], fractal analysis [5], [6]. Amongst those tools, the various declinations of the concepts related to complexities and entropies, and their abilities to discriminate between healthy and non healthy subjects, have received significant, continuous, and on-going efforts [7], [8], [9]. Complexities and entropies can be defined either from the (deterministic) dynamical system framework or from the (random) stochastic processes one. The former context led to the construction of tools that have become standard references in HRV analysis, both for adults and fetuses: *Approximate Entropy* (AE) [7], *Sample Entropy* (SE) [10], regarded as practical approximations to Kolmogorov-Sinai or Eckmann-Ruelle complexity measures (cf. e.g., [11] for a review). Further, they were refined in the so-called *Multiscale Entropy* (ME), that evaluates Sample Entropy for various low-pass-downgraded version of the original data [12]. The stochastic process framework lead to the definitions of Shannon and Rényi entropies and entropy rates. While gaps remain to be bridged, several relations connecting both worlds were obtained (cf. e.g., [11], [13] for reviews). The practical implementations of AE, SE and ME essentially rely on *Correlation-Integral* based algorithm (CI) [7], [12], [14] while that of Shannon entropy rates instead is usually based on a k*-nearest neighbor* (k-NN) algorithm [15], which was recently shown to bring robustness and improved performance to entropy estimation [16].

Goals, contributions and outline. This contribution aims at evaluating the extent to which an approximation of Shannon Entropy rate measured on normalized wavelet and approximation coefficients at various time scales lead to efficient discrimination between the temporal dynamics of F-HRV of healthy fetuses against fetuses suffering from acidosis. To that end, entropies and wavelet decompositions (recalled in Section III) are applied to a high quality database of almost two thousands of intrapartum F-HRV time series, collected in a French academic Hospital (described in Section II). The benefits of using the k-NN algorithm compared to the CI ones are discussed. Further, a supervised learning is used to investigate at which time scales wavelet and approximation coefficient based entropy rates permit to improve discrimination between acidotic and non acidotic fetuses (cf. Section IV).

II. Database

The database, collected at the public academic French Hospital Femme-Mere-Enfant, from 2000 to 2010, consists of 3049 intrapartum CTG signals, acquired using scalp electrode by STAN S21 or S31 system, with 12bit resolution, 500 Hz sampling rate (STAN, Neoventa Medical, Sweden). Clinical information for each woman and neonate were systematically collected by obstetricians in charge, cf. [17]. For the present study, CTG signals matching the *quality* criteria detailed in [4], [6] were selected, yielding 1816 records for analysis, amongst which 60 fetuses were suffering from metabolic acidemia (umbilical artery pH $\leq$ 7.05). We further refer to this later group as *abnormal*, as opposed to *normal* for the remainders.

Analysis is performed directly (without resampling) on the interarrival times time series $X = \{x_1, \ldots, x_n = t_{n+1} -$

Work supported by ANR BLANC 2010 FETUSES 18535

[1] CNRS, Physics Department, ENS de Lyon, France
jiri.spilka@ens-lyon.fr, stephane.roux@ens-lyon.fr
[2] INRIA, Comp. Science Department, ENS de Lyon, France
paulo.goncalves@ens-lyon.fr
[3] Femme-Mère-Enfant Hospital, Lyon, France
muriel.doret@chu-lyon.fr

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

$t_n, \dots x_N\}$, where t_n is the occurrence time of the n-th R peaks. Analysis is conducted systematically on the 40 minutes directly preceding the active stage of labour.

III. Methodology

Entropy rate. The temporal dynamics of the stochastic process underlying observation X can be characterized using Shannon Entropy rate $h = \lim_{m\to\infty} h_m$, where $h = H_{m+1} - H_m$ with H_m Shannon Entropy for the joint distribution p_m of the m-dimensional vector $X^{(m)}$ built from X: $X_i^{(m)} = (x_{i-m+1}, \dots, x_i)$: $H_m = -\mathbb{E}[\log p_m] = -\int \dots \int p_m(X^{(m)}) \log p_m(X^{(m)}) dx_1 \dots dx_m$, cf. e.g., [11]. Although m should be large, in practice observation finite size jeopardizes estimation performance for large m, so one usually uses $m = 3$ as a trade-off. Theoretical connections between an Information Theory quantity, $h_3 = H_3 - H_2$, and different dynamical system complexity measures, AE and SE, are detailed in e.g., [11], [13].

Estimation. Estimation of h_3 relies on our own implementation of the k-NN algorithm, described in [15]. The full description of that algorithm is beyond the scope of the present contribution, that instead intends to emphasize the central difference with the CI algorithm (commonly used to compute AE and SE). Both algorithms essentially aim at measuring the density of the distribution of p_m around $X_i^{(m)}$ and both rely on a distance d (usually the L^∞ norm distance) to evaluate how close $X_i^{(m)}$ is from other vectors $X_j^{(m)}$. The CI algorithm counts the number $B_i^m(\epsilon)$ of neighbors of a central point $X_i^{(m)}$ within a ball of radius ϵ which is set *a priori* (counting or not the central point itself), while the k-NN algorithm reverses the perspective and computes the distance ϵ needed to embed the k nearest neighbors of the central point, where k is set *a priori*. AE – resp. SE – is then computed from the ensemble average (over the set of vectors $X_i^{(m)}$ along the time series) of the logarithms of the (normalized) numbers B_i^m and B_i^{m+1} [7] – resp. the logarithms of the ensemble averages of B_i^m and B_i^{m+1} [10]. The benefits of the k-NN algorithm in terms of performance are discussed and illustrated in Section IV.

In the sequel, all quantities are computed on standardized time series X to avoid differences only due to variances which are irrelevant for temporal dynamics comparisons. Moreover, the theoretical $\log(2\epsilon)$ correction is added to AE and SE to account for the arbitrary choice of ϵ [13]. We choose $k = 10$ and $\epsilon = 0.2$, unless otherwise stated. Section IV-A compares these different entropy rate estimates: AE, SE measured from CI against h_3 measured using the k-NN algorithm.

Wavelet and Approximation coefficient entropy rate. In [12], *Multiscale Entropy* (ME) was defined, that evaluates SE on successive low-pass versions of data. Expanding on that idea, it is proposed here to investigate the benefits of computing h_3 on the (standardized) wavelet and approximations coefficients stemming from a Discrete Wavelet Transform (DWT). Let ϕ and ψ denote the scaling function and mother wavelet associated to 1D Multiresolution Analysis [18]. The wavelet and approximation coefficients X are defined respectively as $D_X(j,k) = \langle \psi_{j,k} | X \rangle$ and $A_X(j,k) = \langle \phi_{j,k} | X \rangle$, with $\{\psi_{j,k}(t) = 2^{-j}\psi(2^{-j}t - k)\}_{(j,k)\in\mathbb{N}^2}$ and $\{\phi_{j,k}(t) = 2^{-j}\phi(2^{-j}t - k)\}_{(j,k)\in\mathbb{N}^2}$.

The collections of h_3 computed, for each scale $a = 2^j$, from $A_X(j,k)$ and $D_X(j,k)$, referred to as $h_{A(j)}$ and $h_{D(j)}$, differ from ME in several ways. First, while $h_{A(j)}$ are measured on low pass degraded versions of X (as ME), the $h_{D(j)}$ are measured on band pass (around scale $a = 2^j$) filtered versions of X. They thus measure another form of multiscale entropy that focuses on temporal dynamics at scales around $a = 2^j$ only. Second, in ME, successive approximations are computed for linearly spaced scales $a = 1, 2, 3, \dots, \dots$ producing highly redundant and correlated low pass versions of X, whereas for $A_X(j,k)$ and $D_X(j,k)$ scales are logarithmically spaced $a = 2^j$, thus producing far less redundant low-pass and band-pass versions of X. The resulting $h_{A(j)}$ and $h_{D(j)}$ are thus probing the temporal dynamics of X in a wider range of scales. Third, smooth wavelet with large number of vanishing moments can be used, diminishing the impact of the windowing, compare to a crude direct uniform window average. Daubechies wavelets with 3 vanishing moments are used here. By convention, $A_X(0,\cdot) \equiv X$ and $h_{A(0)} \equiv h_3$ measured on X.

Feature selection and classification procedure. The goal now is to find, amongst a vector of features $\mathbf{x} = \{h_{A(0)}, \dots, h_{A(6)}, h_{D(1)}, \dots, h_{D(6)}\}$, the minimal subset (of size $l \ll L$) that is used for training a classifier, which generalizes well when applied to unseen data. To avoid selection bias and data overfitting, the Wrapper method for feature selection is used in the cross-validation (CV) procedure [19]. The wrapper involves a sequential forward selection (SFS) search procedure. It makes use of the Area Under Curve (AUC) of Receiver Operational Characteristics (ROC) curves for the performance criterion, and of Fisher linear discriminant (FLD) analysis [20] for the classifier. Equivalent results and conclusions were obtained using a SVM classifier (not reported here for space reasons). The CV procedure consist of a testing-training splitting inner-outer loop nested design, cf. [19, p. 85]: The inner-loop permits unbiased feature subset selection using the Wrapper approach, while the outer loop serves for decision threshold selection and performance evaluation. The procedure is repeated 50 times to reduce sampling bias. The number of best features is varied from $l = 1$ to 5, to evaluate the benefits of adding features.

IV. Results and Discussions

A. Comparing k-NN to CI algorithms

Synthetic data. Because it has been chosen to study 40 minutes of data (before labour), analyzed time series may have very different sample size (ranging from 3000 to 7000 beats, with mean at 5430 $\pm$ 594). To evaluate the impact this may have on entropy rate estimation, 1816 series of i.i.d., $\mathcal{N}(0,1)$, with sample sizes matching the empirical data

size distribution, are drawn at random. Fig. 1(a) shows that estimated means of SE, AE and h_3 clearly differ, that can be compared against theoretical values (cf. [13]). Fig. 1 thus confirms the commonly accepted result that the standard CI-based estimation of AE should not be used in practice as it shows large Mean Square Error (MSE) performance (large bias and variance), cf. Table I. It also shows both departure from Gaussianity and strong dependence on the choice of ϵ (Fig. 1(b)) and on the analysis scale $a = 2^j$ (Fig. 1(c)), because of its strong dependance on sample size. Further, although SE compares well against the k-NN based estimation of h_3, it shows a slightly larger sensitivity to the choice of ϵ (Fig. 1(b)) and more importantly a much larger variance when scales $a = 2^j$ increases, *i.e.*, when sample size decreases (Fig. 1(c)).

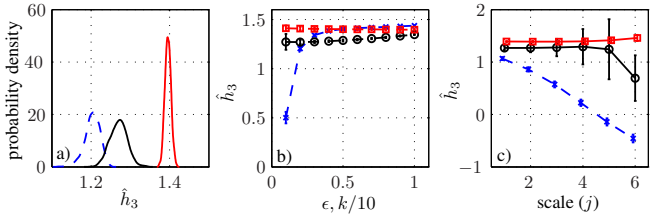

Fig. 1. **Comparisons on synthetic data.** i.i.d. $\mathcal{N}(0,1)$ with sample sizes matching data size distribution. (a): Empirical distributions ; (b): Dependence on ϵ and k ; (c): Dependence on (the log of the) scale $a = 2^j$. CI-AE: (×) ; CI-SE: (○) ; kNN-h_3: (□).

TABLE I

ESTIMATED ENTROPIES FOR I.I.D. $\mathcal{N}(0,1)$. SIZE DISTRIBUTION MATCH THAT OF REAL DATA.

Algorithm	Bias	SD	MSE	skewness
CI-AE	-0.2156	0.0195	0.2165	-0.6180
CI-SE	0.0043	0.0218	0.0222	0.0195
k-NN h_3	-0.0232	0.0076	0.0244	0.0247

Comparisons on real data. Applied to the real F-HRV data, the 3 estimates are observed to show equivalent performance in discriminating normal from abnormal fetuses (measured either in terms of p-value of the two sample Kolmogorov-Smirnov (KS) test of equality of the distributions or AUC), cf. Figs. 2(a) and (b). However, a careful examination of the distributions of the estimates yields interesting observations. First, Figs. 2(a) and (b) clearly show that real data are actually characterized by negative entropies and also that distributions of estimates for AE and SE are strongly skewed, a very negative characteristic for empirical estimates obtained from a real-world population, where central limit theorem is expected to yield Gaussian estimates. This is a direct consequence of the CI procedure: The number of *neighbors* obtained with embedding m, in an ϵ vicinity, are necessarily included in those obtained with embedding $m+1$, so that that the outputs of the CI procedure is bound to positive values, hence skewed. AE and SE are thus taking negative values only because the $\log 2\epsilon$ theoretical correction produces a negative shift. The k-NN procedure does not suffer from such a limitation. Thus by construction the CI-based estimates of AE and SE do not capture well the entropies corresponding to fetus R-peak interarrival distributions, likely because of their specific shapes induced by decelerations. Fig. 2(c) also clearly shows that CI-based estimates of AE and SE present (despite the theoretical correction $\log 2\epsilon$) a much stronger dependence on the arbitrary choice of ϵ, while the k-NN based estimate of h_3 reveals a mild dependence on the choice of k. This thus constitutes a very important property of the latter estimate that yields significant robustness against the arbitrariness of the choice of the vicinity parameter.

The results obtained from synthetic data as well as those observed on real fetuses data, very consistent with those reported in [16], strongly motivate us to prefer the k-NN based estimate of h_3, which is used in the sequel for multiscale feature selection.

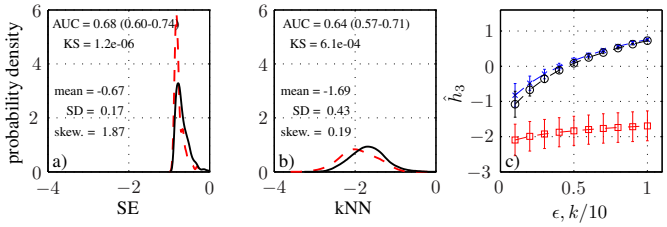

Fig. 2. **Comparisons on real data.** Empirical distributions of normal (solid) and abnormal (dashed) fetuses for SE (a), and KNN (b). Dependence of CI-AE (×), CI-SE (○), kNN-h_3 (□) on ϵ or k (right).

B. Multiscale entropy and feature selection

Independent evaluation of feature discriminative power. Table IV-B quantifies (with AUC) the discrimination abilities of h_3 computed on wavelet and approximation coefficients at different scales. It clearly shows that the $h_{D(j)}$ are individually poorly discriminative. For the $h_{A(j)}$, it reveals interestingly that $h_{A(j=1)}$ shows the best discriminative power and notably performs better than $h_{A(j=0)}$ (i.e., h_3 computed on X itself). Further, Fig. 3, reporting feature pairwise correlations, reveals that $h_{D(j=1)}$ shows low correlation with $h_{D(j\neq 1)}$ or $h_{A(j)}$ at all other scales, indicating that $D_X(j=1)$ has different temporal dynamics compared to all other $D_X(j)$ and $A_X(j=1)$. These two observations yield the first key result of the present contribution: The time series $X = A_X(j=0)$ can be read as the sum of a lower approximation $A_X(j=1)$, essentially consisting of a denoised version of X obtained from a low-pass average of two successive R-peak interarrival times, and an uncorrelated non informative noise $D_X(j=1)$. This high frequency noise (at scale of 2 beats) might originate either from imprecise location of R peaks in fetal electrocardiogram (FECG) or from quantization (rounding distance between successive R-peaks to an integer value). The denoised $A_X(j=1)$ thus provides a better description of F-HRV, with larger discriminative power between normal and abnormal fetuses.

Joint evaluation of feature discriminative power. The outcomes of the CV procedure described in Section III, for $l = 2$ and $l = 3$, are reported in Table III. It shows first that the CV procedure confirms the independent feature evaluation: $h_{A(j=1)}$ is systematically chosen as the first and most discriminative feature. In addition, the CV procedure

TABLE II

AUC FOR ENTROPY ON WAVELET AND APPROXIMATION COEFFICIENTS.

scale (2^j)	0	1	2	3	4	5	6
$A_X(j)$	0.64	0.69	0.67	0.66	0.63	0.60	0.55
$D_X(j)$	N/A	0.59	0.57	0.54	0.51	0.51	0.51

yields two new key results. Second, adding $h_{D(4)}$ to $h_{A(j=1)}$, as the second most discriminative feature, leads to substantial performance improvement. For the CV folds where $h_{D(4)}$ was not chosen, $h_{D(3)}$ or $h_{D(5)}$ are selected as second most discriminative feature, essentially yielding the same conclusion: A band pass filtered version of X, around a scale of $a = 2^4 = 16$ beats (possibly ranging from $2^3 = 8$ to $2^5 = 32$ beats), contains relevant information related to F-HRV temporal dynamics contributing to improved discrimination between normal and abnormal fetuses. Interestingly, a scale of 10 beats is usually observed as discriminant in fractal analysis (as the scale used to define two different scaling exponents in Detrended Fluctuations Analysis of F-HRV [5]). Third, it shows that adding a third feature does not improve discrimination: The AUC performance are not improved and the choice of the third features varies considerably across the folds of the CV procedure. Also, it is worth mentioning that supervised learning without CV procedure would lead to the incorrect conclusion that increasing the number of features to $l = 3$ (cf. Table III), and even to higher l (not shown here), yield improved classification performance, when it actually rather indicates data overfitting. CV in training and testing is thus crucial for feature selection in classification.

C. Conclusions

This contribution illustrates the improved and robust performance of the k-NN procedure for entropy rate estimation. Further, it shows that high frequency fluctuations (at the scale of 2 beats) in intrapartum R-peak interarrival times series can be regarded as noise. Entropies measured on $A_X(j = 1)$ are then more characteristics of F-HRV temporal dynamics. Also, band pass filtered data at scale of $\simeq 16$ beats convey information relevant to F-HRV temporal dynamics and discrimination between normal and abnormal fetuses. The importance of CV procedure in supervised learning feature selection is also clearly illustrated.

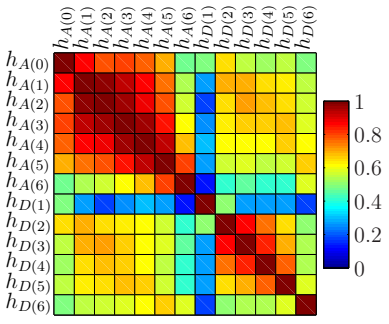

Fig. 3. **Pairwise correlations of the $h_{A(j)}$ and $h_{D(j)}$.**

TABLE III

FEATURE SELECTION WITH AND WITHOUT CV

feat. ($l = 2$)	selected [%]	rank	AUC (no-CV)	AUC (CV)
$h_{A(1)}$	100	1.00	0.69	0.69 (0.68-0.70)
$h_{D(4)}$	86	2.00	0.74	0.73 (0.72-0.74)
$h_{D(3)}$	8	2.00	x	x
$h_{D(5)}$	5	2.00	x	x
$h_{D(6)}$	1	2.00	x	x
feat. ($l = 3$)				
$h_{A(1)}$	100	1.00	0.69	0.69 (0.68-0.70)
$h_{D(4)}$	82	2.02	0.74	0.73 (0.72-0.74)
$h_{A(0)}$	51	2.98	0.76	0.72 (0.71-0.73)
$h_{D(5)}$	18	2.63	x	x
$h_{A(4)}$	15	2.97	x	x

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, no. 4, pp. 609–624, 2007.

[2] FIGO, "Guidelines for the Use of Fetal Monitoring," *International Journal of Gynecology & Obstetrics*, vol. 25, pp. 159–167, 1986.

[3] J. Van Laar et al., "Spectral analysis of fetal heart rate variability for fetal surveillance: Review of the literature," *Acta Obstetricia et Gynecologica Scandinavica*, vol. 87, no. 3, pp. 300–306, 2008.

[4] V. Chudáček et al., "Scattering transform for intrapartum fetal heart rate characterization and acidosis detection.," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2013, pp. 2898–2901, 2013.

[5] D. P. Francis et al., "Physiological basis of fractal complexity properties of heart rate variability in man.," *J Physiol*, vol. 542, no. Pt 2, pp. 619–629, Jul 2002.

[6] M. Doret et al., "Multifractal analysis of fetal heart rate variability in fetuses with and without severe acidosis during labor," *American Journal of Perinatology*, vol. 28, no. 4, pp. 259–266, 2011.

[7] S. M. Pincus and R. R. Viscarello, "Approximate entropy: a regularity measure for fetal heart rate analysis," *Obstet Gynecol*, vol. 79, no. 2, pp. 249–255, Feb 1992.

[8] S. Pincus, "Approximate entropy (ApEn) as a complexity measure," *Chaos*, vol. 5, no. 1, pp. 110–117, Mar. 1995.

[9] D. E. Lake et al., "Sample entropy analysis of neonatal heart rate variability.," *Am J Physiol Regul Integr Comp Physiol*, vol. 283, no. 3, pp. R789–R797, Sep 2002.

[10] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Am J Physiol Heart Circ Physiol*, vol. 278, no. 6, pp. H2039–H2049, Jun 2000.

[11] P. Grünwald and P. Vitányistard, "Kolmogorov complexity and information theory. with an interpretation in terms of questions and answers," *Journal of Logic, Language and Information*, vol. 12, no. 4, pp. 497–529, 2003.

[12] M. Costa et al., "Multiscale entropy analysis of complex physiologic time series.," *Phys Rev Lett*, vol. 89, no. 6, pp. 068102, Aug 2002.

[13] D. E. Lake, "Renyi entropy measures of heart rate gaussianity," *IEEE Trans Biomed Eng*, vol. 53, no. 1, pp. 21–27, Jan. 2006.

[14] P. Grassberger and I. Procaccia, "Estimation of the Kolmogorov-entropy from a chaotic signal," *Physical Review A*, vol. 28, no. 4, pp. 2591–2593, 1983.

[15] L. F. Kozachenko and N. N. Leonenko, "Sample estimate of the entropy of a random vector," *Problems of Information Transmission*, vol. 23, no. 1, pp. 95–101, 1987.

[16] A. Porta et al., "Entropy-based complexity of the cardiovascular control in Parkinson disease: Comparison between binning and k-nearest-neighbor approaches.," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2013, pp. 5045–5048, 2013.

[17] M. Doret et al., "Use of peripartum ST analysis of fetal electrocardiogram without blood sampling: a large prospective cohort study.," *Eur J Obstet Gynecol Reprod Biol*, vol. 156, no. 1, pp. 35–40, May 2011.

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

[19] I. Guyon et al., *Feature extraction: foundations and applications*, vol. 207, Springer, 2006.

[20] R. Duda et al., *Pattern Classification*, Wiley, November 2001.