

Predicting the Risk of Metabolic Acidosis for Newborns Based on Fetal Heart Rate Signal Classification Using Support Vector Machines

George Georgoulas, Chrysostomos D. Stylios*, *Member, IEEE*, and Peter P. Groumpos, *Senior Member, IEEE*

Abstract—Cardiotocography is the main method used for fetal assessment in every day clinical practice for the last 30 years. Many attempts have been made to increase the effectiveness of the evaluation of cardiotocographic recordings and minimize the variations of their interpretation utilizing technological advances. This research work proposes and focuses on an advanced method able to identify fetuses compromised and suspicious of developing metabolic acidosis. The core of the proposed method is the introduction of a support vector machine to “foresee” undesirable and risky situations for the fetus, based on features extracted from the fetal heart rate signal at the time and frequency domains along with some morphological features. This method has been tested successfully on a data set of intrapartum recordings, achieving better and balanced overall performance compared to other classification methods, constituting, therefore, a promising new automatic methodology for the prediction of metabolic acidosis.

Index Terms—Feature extraction, fetal heart rate (FHR), intrapartum monitoring, metabolic acidosis, support vector machines (SVMs).

I. INTRODUCTION

CARDIOTOCOGRAPHY was introduced into obstetrics practice in the seventies and since then it has been widely used for antepartum and intrapartum fetal surveillance. Cardiotocogram (CTG) consists of two distinct signals, i.e., the continuous recording of instantaneous fetal heart rate (FHR) and uterine activity (UA). These two biosignals are depicted in Fig. 1 with FHR at the upper part and UA at the lower part. FHR variability is believed to reflect the interactions between the sympathetic nervous system (SNS) and the parasympathetic nervous system (PSNS) of the fetus [1]. Stimulation of the PSNS results in a decrease in heart rate of the normal fetus while stimulation of the SNS results in an increase in heart rate. During stressful situations for the fetus, such as the uterine contractions at the time of delivery, the sympathetic nerves may act as a compensatory mechanism to improve the fetal heart pumping activity [1], which is reflected in the FHR signal variations.

Manuscript received November 12, 2004; revised October 15, 2005. Asterisk indicates corresponding author.

G. Georgoulas and P. P. Groumpos are with the Laboratory for Automation and Robotics, Department of Electrical and Computer Engineering, University of Patras, Rion 26500, Greece (e-mail: georgoul@ee.upatras.gr; groumpos@ee.upatras.gr).

*C. D. Stylios is with the Department of Communications, Informatics and Management, Technological Institute of Epirus, 47100 Artas, Greece (e-mail: stylios@teleinfom.teiep.gr).

Digital Object Identifier 10.1109/TBME.2006.872814

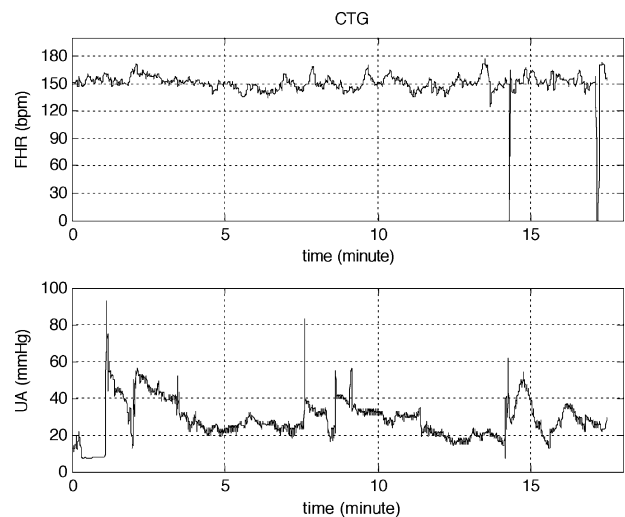


Fig. 1. A typical (digitized) CTG (FHR at the upper part and UA at the lower part).

During the critical period of labor, FHR—the subtlest component of a CTG—is utilized as an indication of the fetal condition and, primarily, as a warning of possible fetal and neonatal compromise, namely metabolic acidosis [2]. Severe hypoxic injury of the fetus can lead to neuro-developmental disability and cerebral palsy or even death. However, neuro-developmental disability and cerebral palsy are often diagnosed several years after birth. Therefore, the objective of monitoring and interpreting FHR patterns is to detect these fetuses that during delivery are at significant risk of developing metabolic acidosis and subsequently to alert the obstetricians to intervene before there is an irreversible damage to the fetus.

Nevertheless, there is controversy regarding the effectiveness of the use of cardiotocography and its consistency [3], especially when it is interpreted by eye inspection. Studies of FHR reliability have shown significant intra- and inter-observer variation in tracing interpretation [4], [5], indicating that even though specific guidelines have been published for its interpretation [6], [7], the different levels of expertise and experience have catalytic influence on the final judgment. These findings pointed out the need for developing automated techniques to reliably interpret the FHR signal and provide early estimations and warnings about the fetal condition.

On the other hand, it is strongly believed that FHR tracings truly convey much more information than what is actually interpreted by obstetricians [8]. During the last two decades, in the area of cardiotocography there have been proposed methods and

systems that range from simple feature extraction utilizing conventional programming techniques [8]–[16] and artificial neural networks [17], [18], to systems capable of performing various diagnostic tasks [19]–[32].

Professor Bernardes and his colleagues [19]–[21] developed a computerized system based on algorithmic manipulation of the guidelines given by the International Federation of Obstetrics and Gynaecology (FIGO) [6]. Magenes *et al.* [22], [23] used artificial neural networks to discriminate between normal and pathological fetal conditions. Kol and Thaler [24] also employed artificial neural networks to interpret nonstress tests. Chung *et al.* [25] developed an algorithm to analyze and predict acidosis. Salamalekis *et al.* [26] employed scale-dependent features extracted from the FHR along with information derived from pulse oximetry recordings and self-organizing maps to diagnose fetal hypoxia. Struzik and Wijngaarden [27] proposed a method based on the cumulative effective Hölder exponent for online monitoring of fetal condition during labor. Professor Ifeachor and his group developed a crisp expert system [28], which they subsequently transformed into a fuzzy system [29] in order to deal with the intrinsic uncertainty in FHR interpretation. Professor Alonso-Betanzos and her group developed and evolved an expert system called NST-EXPERT [30], [31] to create Computer Aided Foetal Evaluator [32], which integrates algorithms with artificial intelligence paradigms, merging FHR analysis and contextual analysis of all pathological and physiological aspects involved in fetal monitoring.

In general, most of the aforementioned approaches employ methods from the field of signal processing and incorporate the doctor's expertise, in order to reach a satisfactory level of reliability so as to act as decision support systems in obstetrics. Up to now, none of them has been adopted worldwide for everyday clinical practice and almost all methods still require an extensive validation, especially for the intrapartum period. Thus, this research effort aims to exploit the capabilities of electronic fetal monitoring by eliminating the difficulties in reading and interpreting the FHR tracings, and to develop an automated computerized system for alerting on possible metabolic acidosis.

In this research work the selection of an appropriate set of FHR features is investigated, which can be fed to a nonlinear classifier in order to “foresee” risky situations (potential compromise of the fetus) during the final stage of labor. The selected features are extracted automatically from the FHR signal in time and frequency domain—some of them have been successfully used for the antepartum case [22], [23]—along with some morphological features (such as baseline value, number of accelerations, etc., e.g., see [6]). Every FHR recording is labeled based on the pH value of the fetal umbilical artery blood samples (acquired right after delivery): the normal group, consisting of babies with pH > 7.20 and the “at risk” group, consisting of babies with pH < 7.10. Using this segregation, it is assumed that two well-defined classes of FHR signals with minimum overlapping were created. It must be mentioned that, as in most medical applications, the training set was imbalanced in the sense that the class containing the fetuses with the low umbilical artery pH value (“at risk” group) was underrepresented compared to the other class. Different combinations of features were examined exhaustively so as to define the set of features that gives the best results. The proposed methodology introduces the use

of a new powerful tool from the field of pattern recognition, the support vector machines (SVMs) [33]–[38], which can classify the FHR recordings achieving balanced performance in discriminating the two classes.

SVMs have been recently introduced in the framework of statistical learning theory [35] and have been used successfully for a number of applications in the field of pattern recognition [33]–[38]. Their experimental success and their ability to generalize well, even when the sample size is small [39], prompted us to select the SVMs as the classification tool in this research work.

This paper is structured as follows; Section II gives a brief introduction to SVMs. Section III describes the stages of the proposed methodology; it discusses the selection of different feature sets and how they are extracted from FHR signal. Section IV presents the data, the different scenarios used to test the proposed methodology, the metric used to evaluate the performance, and the experimental results. Finally, Section V concludes the paper discussing the results and giving hints for further future research.

II. SUPPORT VECTOR MACHINES

SVMs are learning systems that are trained using an algorithm based on optimization theory [33]–[36]. For real life problems, given l observations $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^l$, the SVM solution finds the hyperplane in feature space that keeps both the empirical error small and maximizes the margin between the hyperplane and the instances closest to it. This can be done by minimizing

$$\begin{aligned} & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i \\ & \text{subject to :} \\ & y_i (\langle \mathbf{w} \cdot \boldsymbol{\varphi}(\mathbf{x}_i) \rangle + b) \geq 1 - \xi_i \text{ with } \xi_i \geq 0, \\ & i = 1, 2, \dots, l \end{aligned} \quad (1)$$

where ξ_i are slack variables, which are introduced to allow the margin constraints to be violated, and $\boldsymbol{\varphi}(\cdot)$ is the nonlinear mapping from the input space to the feature space. Parameter C controls the tradeoff between maximizing the margin and minimizing the error and it is usually determined through a cross-validation scheme [36], [40].

The class prediction for an instance \mathbf{x} is given by

$$f(\mathbf{x}) = \text{sign} \left(\sum_{i=1}^l y_i \alpha_i \langle \boldsymbol{\varphi}(\mathbf{x}_i) \cdot \boldsymbol{\varphi}(\mathbf{x}) \rangle + b \right) \quad (2)$$

where the coefficients α_i are calculated by maximizing the Lagrangian

$$\begin{aligned} & \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j \langle \boldsymbol{\varphi}(\mathbf{x}_i) \cdot \boldsymbol{\varphi}(\mathbf{x}_j) \rangle \\ & \text{Subject to : } \sum_{i=1}^l y_i \alpha_i = 0 \text{ and } 0 \leq \alpha_i \leq C, \\ & i = 1, 2, \dots, l. \end{aligned} \quad (3)$$

The points for which $a_i > 0$, are called support vectors and are the points lying closest to the hyperplane. If the nonlinear mapping function is chosen properly, the inner product in the feature space can be written in the following form:

$$\langle \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j) \rangle = K(\mathbf{x}_i, \mathbf{x}_j) \quad (4)$$

where K is called the inner-product kernel [33]–[36].

The above formulation is inappropriate for the case of imbalanced distributions and another approach is required. The simplest one is to use different error weights C^+ and C^- in order to penalize more heavily the undesired type of error, and/or the errors related to the class with the smallest population [41], [42]. Therefore, the optimization problem is modified as follows:

$$\begin{aligned} & \text{Minimize } \frac{1}{2} \|\mathbf{w}\|^2 + C^- \sum_{i:y_i=-1} \xi_i + C^+ \sum_{i:y_i=+1} \xi_i \\ & y_i (\langle \mathbf{w} \cdot \varphi(\mathbf{x}_i) \rangle + b) \geq 1 - \xi_i \\ & \text{with } \xi_i \geq 0, i = 1, 2, \dots, l. \end{aligned} \quad (5)$$

By using higher penalty value for the class with the smallest population, which in most medical applications is the class that needs to be correctly identified, we induce a boundary that is more distant from that class.

III. NOVEL METHODOLOGY FOR FHR PROCESSING AND CLASSIFICATION

The proposed overall procedure is depicted in Fig. 2; it consists of six stages. The sixth stage of the proposed methodology determines the configuration of the SVM classifier that yields the best results.

A. The Fetal Heart Rate Signal

FHR can be either obtained by Doppler ultrasound (the most common method employed during the antepartum period) or directly from the fetal electrocardiogram via scalp electrodes (during the intrapartum period and after the rupture of the membranes) [43]. The cardiac events are easily recognized and the time intervals between them (in seconds) are transformed into an instantaneous rate for each interval between cardiac events. This instantaneous rate is sent to the output of the cardiotocograph. Most devices sample the output of the cardiotocograph at fixed sampling intervals T_s (depending on the manufacturer). A new value for FHR is not assigned to the output until the next sampling instance. Therefore, the FHR in most cardiotocographs is measured in beats/min (bpm) as shown in Fig. 1 and it is a discrete signal, which is not perfectly aligned with the cardiac events.

The experienced clinician can distinguish various patterns on these tracings. The clinician assesses fetal condition by eye inspection of the “morphological” characteristics of FHR, as those are described in the guidelines given by FIGO or the National Institute of Child Health and Human Development Research Planning Workshop [6], [7]. As already pointed out, this approach is highly subjective and of limited reproducibility [3]. Targeting a more objective and reproducible approach, computerized systems, such as those mentioned in the introductory section, have

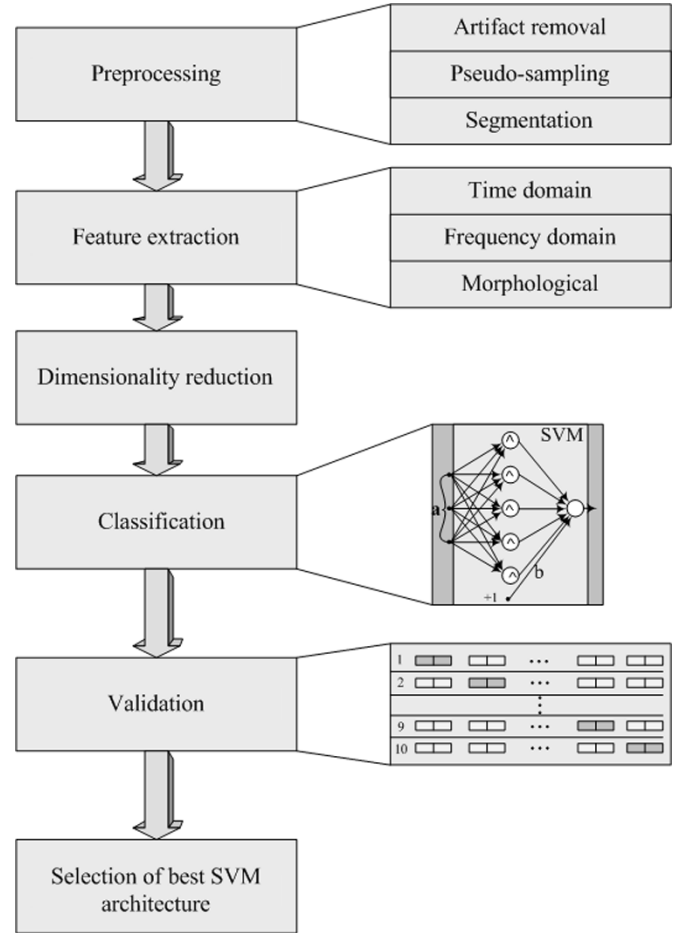


Fig. 2. The overall proposed methodology for FHR classification.

been employed lately to describe and interpret FHR tracings. Different signal processing and pattern recognition techniques have been used to analyze FHR not only in the time domain but also in the frequency domain with very promising results [8], [22], [23], [44].

B. Preprocessing-Artifact Removal Stage

The FHR is a noisy signal, usually containing “spiky” artifacts, which occur mainly due to fetal movements or displacement of the transducer. This becomes more apparent during the final stage of labor, which is a very stressful period, both for the mother and the fetus, and it is reported that missing data in fetal heartbeat records can amount to about 20%–40% of total data [27]. In order to extract a representative set of features from FHR, it is necessary to remove those artifacts by using a preprocessing stage. The employed preprocessing method was introduced in [19]. Whenever a difference between adjacent samples higher than 25 bpm is found, a linear interpolation is applied between the first of these two samples and the first sample of a new stable FHR segment, where a stable FHR segment is defined as a group of five adjacent samples with the difference (in bpm) between them less than 10 bpm.

In addition to the spiky segments, FHR includes segments of missing values. In this case we also applied linear interpolation. Especially, just before the delivery the artifacts were so many

that the artifact removal stage resulted in a completely unnatural flat FHR trace. Those final 1-2 min of the recordings were manually removed. It must be mentioned that no special action was taken when the artifacts or the missing segments were part of a fetal pattern (acceleration or deceleration). After applying the aforementioned preprocessing procedures to the whole data set, all but one of the initial signals, had signal quality above 80% [mean value: 91.3; standard deviation (SD): 6.5] where signal quality is defined as the percentage of FHR values that were not interpolated using the aforementioned procedure [21].

C. Feature Extraction Stage

After the preprocessing stage, the stage of feature extraction follows (Fig. 2). The possible extracted features can be divided into three categories:

- features extracted in time domain;
- features extracted in frequency domain;
- “morphological” features.

The selection of the first two sets of features comes naturally since we are dealing with a time series signal and, as a result, time series analysis tools and other methods from time and frequency domain can act supplementary and, as a matter of fact, have been used in many biomedical applications [45]. The third set of features is motivated from the everyday interpretation of the FHR signal by obstetricians over the past 30 years [6], [7].

1) *Time Domain Features*: There are a number of methods that can evaluate variations in heart rate [44]. The features/indices in time domain that we employed have already been used with reasonable success in the antepartum case [22], [23] and, therefore, we decided to test them in the intrapartum case too. The features employed were the following.

- Mean value of FHR signal $\text{mean} = (1/N) \sum_{i=1}^N \text{FHR}(i)$ where N is the total number of samples of the recording.
- Standard deviation of FHR signal:

$$\text{SD} = \sqrt{(1/(N-1)) \sum_{i=1}^N (\text{FHR}(i) - \text{mean})^2}.$$

- The Delta value is shown in the equation at bottom of page, where m is the time duration of the recording in minutes.
- Short-term variability:

$$\text{STV} = \frac{\sum_{i=1}^{24} |s\text{FHR}(i+1) - s\text{FHR}(i)|}{24}$$

where $s\text{FHR}(i)$ is the value of the signal $\text{FHR}(i)$ taken every 2.5 s (i.e., once every ten samples $s\text{FHR}(i) = \text{FHR}(10 \times (i-1) + 1)$).

- Interval index:

$$\Pi = \frac{\text{STV}}{\text{std}[s\text{FHR}(i)]}.$$

- Long term irregularity (LTI), defined as the interquartile range $[1/4, 3/4]$ of the distribution $m(j)$ with $m(i) = \sqrt{\text{FHR}^2(i) + \text{FHR}^2(i+1)}$, which is a means to evaluate long term variability [9].
- Total value of the Delta:

$$\text{Delta}_{\text{total}} = \max_{i \in [1, N]} (\text{FHR}(i)) - \min_{i \in [1, N]} (\text{FHR}(i)).$$

Therefore, the feature set selected in the time domain was

$$T\text{dset} : \{\text{mean}, \text{std}, \text{Delta}, \text{STV}, \Pi, \text{LTI}, \text{Delta}_{\text{total}}\}.$$

2) *Frequency Domain Features*: Various spectral methods have been used for the analysis of adults' heart rate [44]. However, in the case of FHR, there is no standardized use of frequency bands. In this paper, we experimented with two different sets of features, extracted using slightly different partitioning of the frequency bands. We divided the frequency range into three bands [46] and we calculated the energy of the signal contained in each one of them. The three bands were: a) the very low-frequency (VLF) 0–0.05 Hz; b) the low-frequency (LF) 0.05–0.15 Hz (referred to as Mayers' waves [46]); c) the high-frequency (HF) 0.15–0.5 Hz, which “corresponds to fetal movements” [46].

As a fourth feature we used the ratio of energies in the bands LF, HF.

LF/HF is a standard measure in adults and it is thought to express the balanced behavior of the two branches of the autonomic nervous system [44].

Therefore the first frequency feature set is

$$F\text{dset}1 : \{\text{VLF}, \text{LF}, \text{HF}, \text{LF}/\text{HF}\}.$$

The other alternative frequency feature selection was chosen following suggestions of [8]. In this case, we partitioned the frequency range into four bands: a) VLF 0–0.03 Hz “related to long period and nonlinear contributions” [8]; b) the LF 0.03–0.15 Hz “mainly correlated with neural sympathetic activity” [8]; c) the movement frequency (MF) 0.15–0.5 Hz, which “depends on fetal movements and maternal breathing” [8]; d) the HF 0.5–1 Hz, which “marks the presence of fetal breathing” [8].

As a fifth feature for this feature, set we used the ratio

$$\text{Delta} = \frac{\sum_{k=1}^m \left[\max_{i \in [(k-1) \cdot 240 + 1, k \cdot 240]} (\text{FHR}(i)) - \min_{i \in [(k-1) \cdot 240 + 1, k \cdot 240]} (\text{FHR}(i)) \right]}{m}$$

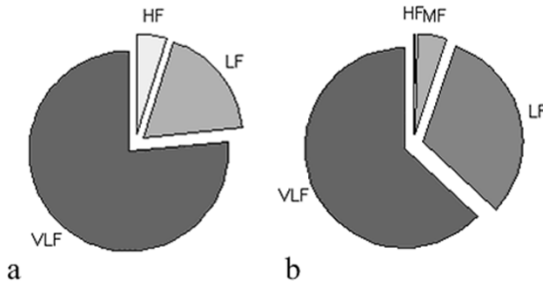


Fig. 3. Energy contribution of the different frequency bands for a normal FHR trace (a) using the 3-band division [46] and (b) using the 4-band division [8].

$LF/(HF+MF)$, which “quantifies the autonomic balance between neural control mechanism from different origin (in accordance with the LF/HF ratio normally calculated in adults)” [8].

Therefore the second feature set $Fdset2$ is defined as

$$Fdset2 : \{VLF2, LF2, MF, HF, LF/(HF + MF)\}.$$

Fig. 3 shows the relative contribution of the frequency bands for the FHR using the two distinct partitions. The above partitioning of the frequency band is not the only one found in the literature [25]. However, we restricted ourselves only to those two partitions for the needs of this paper.

3) “Morphological” Features: Conventional interpretation of FHR is based upon certain morphological characteristics, according to the guidelines given in [6] and [7]. In this paper, we examined two sets of morphological parameters. The first set consisted of only four parameters.

- Baseline—(“the mean level of fetal heart rate when this is stable; accelerations and decelerations being absent” [6]).
- Number of accelerations—(“acceleration is defined as the transient increase in heart rate above the baseline by 15 bpm or more, lasting 15 s or more” [6]).
- Number of decelerations—(“deceleration is defined as the transient episode of slowing fetal heart rate below the baseline level by more than 15 bpm and lasting 10 s or more” [6]).
- The percentage of the time occupied by decelerations.

Therefore the first morphological feature set is $Mset1 : \{\text{Baseline, Number of accelerations, Number of decelerations, Percentage duration of decelerations}\}$.

We additionally examined a second morphological feature set, which also uses the above parameters, but additionally distinguishes the decelerations into three types.

- Mild decelerations, if they do not exceed 120 s [21].
- Prolonged decelerations, if they last 120–300 s [21].
- Severe decelerations, if they exceed 300 s [21].

Therefore, the second morphological feature set is $Mset2 : \{\text{Baseline, Number of accelerations, Number of mild decelerations, Number of prolonged decelerations, Number of severe decelerations, Percentage duration of decelerations (mild, prolonged, severe)}\}$

For the calculation of the baseline we employed the algorithm proposed in [13], which is based on an iterative process

easy to implement. It consists of the following steps: removal of components of the FHR signal associated with accelerations and decelerations; linear interpolation across the gaps, and low pass filtering (this procedure is repeated three times [13]). In this paper, we did not take into account any information contained in the UA signal; thus, we did not further characterize the detected decelerations as early or late [6], [7]. Moreover, some of the time domain features such as the standard deviation of the FHR and STV are actually used by obstetricians in everyday practice. Therefore, these could also be included in the morphological feature set. However, following Magenes *et al.* [22], [23], we adopted the segregation and definitions they have proposed.

Having extracted these features/indexes, we then proceeded to the next stage. We could proceed directly to the classification of the recordings, but instead we introduced a dimensionality reduction stage seeking for improvement in the generalization performance of our classifier.

D. Dimensionality Reduction Stage

In pattern recognition tasks, by using fewer features than those available, usually potential improvement—better generalization—can be achieved. Actually, when we build a classifier we tend to extract several features, which may convey redundant information about the problem at hand. Therefore, it is worth trying to select the most suitable subset of the extracted features.

Principal component analysis (PCA), or Karhunen-Loeve transformation, is a way to perform dimensionality reduction by linear combination of the original features in such a way that preserves as much of the relevant information as possible [47]. This method computes eigenvalues of the correlation matrix of the input data vector and then projects the data orthogonally onto the subspace spanned by the eigenvectors (principal components) corresponding to the dominant eigenvalues. Even if the whole set of the eigenvectors is retained, this may also lead to an improvement of the classification performance, because the new set has features that are uncorrelated and this, in general, improves the classification capabilities of a classifier [47].

E. Classification Stage Using SVMs

After the dimensionality reduction stage we labeled each one of the feature vectors with $\{+1\}$ if it belonged to newborn with $ph < 7.1$ (20 cases)—“at risk” group—and $\{-1\}$ if it belonged to a newborn with $ph > 7.2$ (60 cases)—normal group.

Different SVMs with different nonlinear decision surfaces can be constructed, depending on the choice of the kernel function. Among others, the most popular are the polynomial learning machines, the radial basis function (RBF) networks and the two-layer perceptrons [47]. In our experimental procedure we employed two types of kernels:

- 1) RBF kernels:

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{x}_i\|^2\right) \quad (6)$$

where the width σ^2 is specified *a priori* by the user and is common for all the kernels

2) Polynomial kernels of degree d :

$$K(\mathbf{x}, \mathbf{x}_i) = (\langle \mathbf{x}\mathbf{x}_i \rangle + 1)^d. \quad (7)$$

As it will be explained in Section IV, when SVMs with the same penalty parameter for both classes were examined, this resulted in almost perfect specificity. However, the sensitivity was very poor because the classifiers tended to classify every trace in the normal class. Thus, a different approach was required and so, the scheme with the two penalty parameters C^+ and C^- was employed. As it was reported [48]–[50] and proved through exhaustive experimentation, the ratio C^+/C^- should be set to the inverse of the corresponding cardinalities of the classes.

In this research work, we experimented with various configurations of the learning machines varying the width σ for the RBF kernels (6) and the degree d for the polynomial kernels (7) (the cross validation approach to model selection [36]). For each configuration of the kernel we tested different values for the values C^+ and C^- keeping their ratio equal to three (the inverse ratio of the corresponding cardinalities 1/20/60). Following the medical terminology we refer to the group with the hypoxic babies as the positive cases (the cases that are *at risk*) and the normal group as the negative (the cases that are *not at risk*).

IV. DATA SET AND EXPERIMENTAL RESULTS

A. Data Set Description

We tested the proposed method (Fig. 2) for FHR feature extraction and classification using 80 recordings. The recordings were collected in the context of the Research Project POSI/CPS/40 153/2001, funded by Fundação para a Ciência e Tecnologia, Portugal. The data recordings had various lengths, ranging from 20 min to more than 1 h. Regarding the recordings, 57 of them were acquired using an HP 1350 fetal monitor at a sampling frequency of 4 Hz, and 23 were acquired using a Toitu MT810B. In both cases, scalp electrodes were used for the acquisition, giving accurate recordings [51]. The latter recordings were irregularly sampled and had to be transformed into “pseudo-regularly” sampled signals. This was performed by copying the way HP 1350 operates. To be more specific, HP 1350 operates in the following way: if a new beat is detected during the sampling interval (sampling interval = 0.25 s), the cardiocograph assigns the computed FHR value to the next output value; otherwise, it assigns it to the previous output value. Thus, from the irregularly sampled FHR signal we reconstructed the sequence of the detected beats (R-peaks) and then we created a regularly sampled FHR sequence using a zero-order-hold procedure, exactly the same way HP 1350 would have done.

For simplicity and practical reasons we adopted the approach of imitating the operation of HP 1350 fetal monitor. However, more sophisticated methods exist, such as the one proposed by Berger *et al.* [52] to deal with the irregular signal produced by Toitu MT810B and the algorithm proposed by Bracale *et al.* [53] for the signal produced by the HP 1350. Because the recordings from the HP 1350 monitor had a lot of missing samples (as it

is usually the case [27]) an attempt to derive the original irregularly sampled sequence would be a very difficult, if not virtually impossible, task. As a result, we considered the outcome of the HP 1350 as our standard and we used linear interpolation to deal with the missing samples.

As aforementioned, the duration of the recordings ranged from 20 min to more than 1 h; thus, for homogeneity reasons we used segments of equal duration for each recording and performed the subsequent analysis on these segments only. We tried two different segmentations. We cropped, starting from the end of the recording (or as close to the end as possible), segments lasting 10 and 20 min (after the removal of any unnaturally flat segments as described in Section III-B).

The experimental data consisted of tracings belonging to newborns with umbilical artery pH either less than 7.10 or greater than 7.20. Newborns with umbilical artery pH value in the range (7.10, 7.20) were not included in the data set. Intuitively, by this division we were expecting to have two distinguished classes that could be separated. With these cut-off values, 20 newborns (ten from the HP 1350 set and ten from the Toitu MT810B set) were labeled as positive (“at risk”) and 60 newborns were labeled as negative (normal). Because of the restricted number of cases, we used the stratified tenfold cross-validation scheme [54], an extension of regular multifold cross validation [47] in order to evaluate the performance of the proposed methodology. Therefore, we divided the 80 cases into ten nonoverlapping groups containing eight cases each (six normal and two at risk). For each one of the ten subsets we created one training set comprising of the rest nine sets. Within the training set we used ninefold stratified cross validation to tune the parameters of the SVM. Once the parameters were fine-tuned the SVM was retrained for this set of parameters using the whole training set and its performance was evaluated using the corresponding subset that was originally left out. The aforementioned procedure was adopted in order not to use the same set both for tuning and estimating the performance of the proposed classifier [55].

B. Different Scenarios Examining the Proposed Methodology

We exhaustively tested the classification performance for the two segment durations (10 and 20 min) using various combinations of the features so as to identify the best feature set and the most appropriate duration for FHR classification.

- Only the four morphological features, $\{Mset1\}$.
- Only the six morphological features, $\{Mset2\}$.
- Only the seven time domain features, $\{Tdset\}$.
- Only the four frequency domain features, $\{Fdset1\}$.
- Both the time and frequency domain features, $\{Tdset, Fdset1\}$ (11 features).
- The 15 extracted features from all three domains, $\{Mset1, Tdset, Fdset1\}$.
- The 17 extracted features from all three domains, $\{Mset2, Tdset, Fdset1\}$.
- Only the five frequency domain features, $\{Fdset2\}$.
- Both the time and frequency domain features, $\{Tdset, Fdset2\}$ (12 features).
- The 16 extracted features from all three domains, $\{Mset1, Tdset, Fdset2\}$.

The 18 extracted features from all three domains, $\{Mset2, Tdset, Fdset2\}$.

For each combination, we also experimented retaining different number of principal components at the dimensionality reduction stage (ranging from 1 to the maximum number of features). For each scenario, SVM classifiers with different asymmetric penalty parameters were tested. For comparison reasons, experiments were also conducted using three conventional classifiers: the k-nearest neighbor, the linear and the quadratic classifiers [56].

C. Metrics to Evaluate Experimental Results

Due to the imbalanced nature of the data set, the selection of accuracy (overall classification rate) as a metric is not the best choice. For example a classifier that classifies everything as negative (all cases classified as normal) will be 75% accurate but it will be completely useless. A more appropriate metric is the geometric mean $g = \sqrt{a^+ \cdot a^-}$ [57], where a^+ is the accuracy, which is observed separately on positive examples (also known with the term sensitivity) and a^- is the accuracy observed separately on negative examples (also known with the term specificity).

Another approach to compare classifiers is by using their corresponding receiver operating characteristic (ROC) curves. The area under the ROC curve (AUC) is also another single scalar value that can be used for classifier comparison [58], [59]. The AUC of a classifier is the probability with which the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance. On the other hand, maximization of the geometric mean corresponds to fitting rectangles under the ROC curve and choosing the rectangle of greatest area.

For this work we used the geometric mean as the performance metric, which our classifier intended to maximize. In addition to this, we also calculated the AUC for each one of the optimal classifiers as a second performance measure.

D. Experimental Results

The overall procedure (Fig. 2) was repeated for all the 11 scenarios of feature set selection and for the two different SVM classifiers with polynomial and RBF kernels. The RBF kernel machines outperformed the polynomial machines and both of them outperformed the conventional methods of k-nearest neighbor linear and quadratic discriminant classifiers. Thus, we elaborate on the results of SVMs with RBF kernels, which performed better. The difficult task of predicting the risk of acidemia based on the intrapartum FHR trace, which was regarded with skepticism in the early 1990s [60], was made even more difficult because of the imbalanced nature of the data set. This can be easily verified by examining the performance of conventional classifiers such as the k-nearest neighbor, linear and quadratic classifiers. For this particular data set, all three conventional classifiers were biased heavily towards the normal cases. If we had used only the accuracy as a means to compare the four classifiers we would have mistakenly drawn the conclusion that they are all somehow of equal performance. Using the g-metric it can be seen that the SVM classifier outperforms the other three classifiers achieving balanced accuracy for both

TABLE I
CLASSIFICATION RESULTS FOR 20-min SEGMENTS
UTILIZING PCA FOR FEATURE REDUCTION

	k-nn (6 PCs) {Mset2, Tdset, Fdset2}	ldc (5 PCs) {Mset2, Tdset, Fdset2}	qdc (5 PCs) {Tdset}	SVM (3 PCs) {Tdset, Fdset2}	SVM (2 PCs) {Tdset, Fdset2, Mset1}
accuracy (%)	76.25	81.25	83.75	81.25	78.75
accuracy+ (%)	25	45	50	70	70
accuracy- (%)	93.33	93.33	95	85	81.68
g (%)	48.31	64.81	68.92	<u>77.14</u>	75.61
AUC	0.37	0.48	0.59	0.75	<u>0.78</u>

classes (Table I). The superiority of the SVM over the other classifiers is made even more apparent in terms of AUC. In fact, the conventional classifiers almost totally fail to rank correctly the unseen data, proving the complexity and difficulty of this particular problem.

For the 10-min segments the performance deteriorates for all the tested feature sets. The best results were achieved using the set $\{Mset2, Tdset, Fdset1\}$ and retaining only two principal components: $g = 70.83\%$, $a^+ = 70\%$, $a^- = 70.67\%$, $a = 71.25\%$, (error rate = 0.2875, standard deviation of error = 0.0506), AUC = 0.7.

Regarding the morphological features the results indicate that when they are used alone, they perform worse compared to the other sets of features. This shows that further investigation is needed since the obstetricians believe that the decelerations are of paramount importance for the outcome of the delivery. In that direction the morphological module would be redesigned using a refinement of the concepts of FHR baseline, variability, accelerations and decelerations proposed by the FIGO [6] and the National Institute of Child Health and Human Development Research Planning Workshop [7], as in [21] and [61], or using a neural network approach [17], [18] and would be further evaluated.

For all the test sets, the proposed method using the SVM classifier gave values for g and AUC higher than those achieved by conventional classifiers (with the exception of the g value for the morphological features alone). By adjusting the two penalty parameters C^+ and C^- we were able to achieve high accuracies separately for each one of the two classes. However, the limiting value for these particular test sets in terms of the g metric was the one using features both from the time and the frequency domain $\{Tdset, Fdset2\}$: $g = 77.14\%$, $a^+ = 70\%$, $a^- = 85\%$, $a = 81.25\%$, (error rate = 0.1975, standard deviation of error = 0.0445), AUC = 0.75, retaining only three principal components. Regarding the AUC the best performance was recorded when in the aforementioned features set we added the morphological features (with only one type of decelerations) $\{Mset1, Tdset, Fdset2\}$: $g = 75.61\%$, $a^+ = 70\%$, $a^- = 81.67\%$, $a = 78.75\%$, (error rate = 0.2125, standard deviation of error = 0.0457), AUC = 0.78, retaining only two principal components.

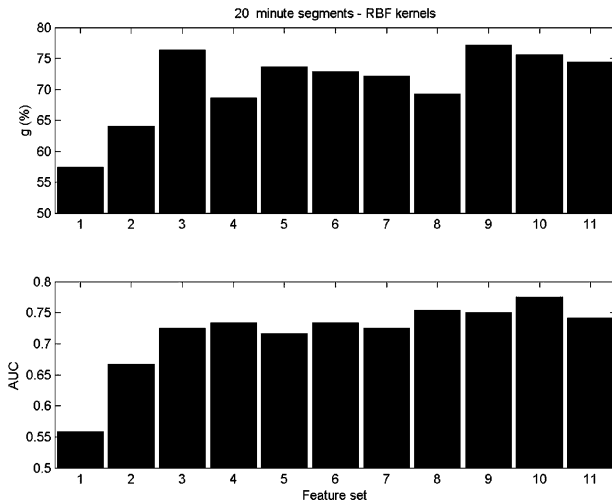


Fig. 4. Classification performance of the different 11 feature sets selection using RBF kernels for 20-min segments comparing them using (a) g-metric, (b) AUC.

Fig. 4 shows the best results for each one of the 11 sets for 20-min segments for RBF kernels and asymmetric penalty parameters. Table I presents the best performance achieved by the conventional classifiers for 20-min time segments. Fig. 4 and Table I shows that when g-mean is chosen as a measure of performance, the best choice of features will include features from the time and the frequency domain. Especially, the segregation for frequency domain proposed by Signorini *et al.* [8] seems to be a better choice over the one proposed in [46].

When the AUC is chosen as a measure of performance, all selections of feature sets, with the exception of the morphological features when used alone, exhibit similar performance with the feature set achieving the best score among them being the $\{Mset1, Tdset, Fdset2\}$.

In terms of accuracy, the quadratic discriminant classifier is the best one but we can hardly recommend its use since it lacks completely the ability to take into account the imbalanced nature of this particular problem. By adjusting the penalty parameters of the SVM it is possible to achieve high classification rates increasing dramatically the specificity at the expense of decreasing the sensitivity.

V. CONCLUSION

This research work proposes a novel integrated methodology to identify FHR signals of fetuses that are suspicious of developing metabolic acidosis. This is achieved by applying an analysis and processing stage for the FHR signal, which produces a set of features that are descriptive with respect to this specific problem. Then, SVMs are proposed and used to integrate the difficult task of classification and identification.

The design of the experimental process and the use of two distinct thresholds to separate the two classes does not allow us to make direct comparisons with other similar works concerning the prediction of metabolic acidosis during the second stage of labor [25], [26], [62]. If, however, we try to indirectly compare the proposed methodology with those developed by other researchers we can outline the following:

- 1) Comparing with the results reported by Chung *et al.* [25], their algorithm achieves better sensitivity 88% and worse specificity 75%. However, their algorithm has limitations [62] and, due to the small sample size used, further evaluation is needed.
- 2) Comparing with the results reported by Salamalekis *et al.* [26], their algorithm seems to perform better than the results reported here. However, the use of fetal pulse oximetry may account for the high values in the performance of their proposed methodology.

The results of our research are comparable but worse to our previous work, where we used features extracted from frequency and time domain along with Hidden Markov Models for the classification of FHR traces [63], and accuracy equal to 83% with balanced performance was achieved. However, in that work we had used equally balanced data from the two classes.

Another issue that we would also point out is that the pH value should be probably chosen lower for the “at risk” cases. A more justified threshold would be to choose threshold value of pH at 7, but this would compromise more the classification performance, since only two cases of our data would belong to the positive ones. It is obvious that with this partitioning, even the modified SVM classifier would have problems to cope with such distribution of cases. It is worth mentioning that only very low pH values (6.8) are related to neonatal death or major neurological damage [1].

Concluding, even though the achieved results are quite promising, we must be very careful before we definitely suggest the proposed methodology as the best choice for the identification of metabolic acidosis based solely on FHR traces. Therefore, the whole procedure has to be tested using more cases in order to allow “extrapolation” for any unseen case. However, this work is a positive step toward a more objective analysis of the FHR signal for the prediction of metabolic acidosis using SVMs. In future work, we will also consider the use of the Apgar score as another index component for the formation of the classes, something which was not used in this study [61].

To sum up, the far from optimum performance of the SVMs (and the other conventional classifiers) in terms of AUC, indicates that even though FHR trace conveys valuable information, when this is used independently, it is not enough to achieve very high performance. Therefore, for maximizing the performance it is proposed to incorporate the presented method of the analysis and interpretation of the FHR in a larger framework for fetal surveillance during labor, which will integrate clinical, biophysical and biochemical data both of the mother and the fetus.

ACKNOWLEDGMENT

The authors would like to thank Prof. J. Bernardes, Department of Gynecology and Obstetrics, Porto Faculty of Medicine, Porto, Portugal, and SisPorto Project, for providing the fetal heart data and critical comments, within the Research Project POSI/CPS/40 153/2001, from Fundação para a Ciência e Tecnologia, Portugal. They would also like to thank the anonymous reviewers for their valuable comments and suggestions that improved this paper.

REFERENCES

- [1] J. T. Parer, *Handbook of Fetal Heart Rate Monitoring*, 2nd ed. Philadelphia, PA: Saunders, 1997.
- [2] H. P. van Geijn, "Developments in CTG analysis," *Bailliere's Clin. Obstet. Gynaecol.*, vol. 10, no. 2, pp. 185–209, Jun. 1996.
- [3] D. MacDonald, A. Grant, M. Sheridan-Pereira, P. Boylan, and I. Chalmers, "The Dublin randomized controlled trial of intrapartum fetal heart rate monitoring," *Am. J. Obstet. Gynecol.*, vol. 152, no. 5, pp. 524–539, Jul. 1985.
- [4] J. Bernardes, A. Costa-Pereira, D. Ayres-de-Campos, H. P. van Geijn, and L. Pereira-Leite, "Evaluation of interobserver agreement of cardiotocograms," *Int. J. Gynecol. Obstet.*, vol. 57, no. 1, pp. 33–37, 1997.
- [5] D. Ayres-de-Campo, J. Bernardes, A. Costa-Pereira, and L. Pereira-Leite, "Inconsistencies in expert's classification of cardiotocograms and subsequent clinical decision," *Br. J. Obstet. Gynaecol.*, vol. 106, pp. 1307–1310, 1999.
- [6] G. Rooth, A. Huch, and R. Huch, "Guidelines for the use of fetal monitoring," *Int. J. Gynaecol. Obstet.*, vol. 25, pp. 159–167, 1987.
- [7] National Institute of Child Health and Human Development Research Planning Workshop, "Electronic fetal heart rate monitoring: research guidelines for interpretation," *Am. J. Obstet. Gynecol.*, vol. 177, no. 5, pp. 1385–1390, Dec. 1997.
- [8] M. G. 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.
- [9] D. Arduini, G. Rizzo, G. Piana, A. Bonalumi, P. Brambilla, and C. Romanini, "Computerized analysis of fetal heart rate: I. Description of the system (2CTG)," *J. Matern. Fetal Invest.*, vol. 3, pp. 159–163, 1993.
- [10] R. Mantel, H. P. van Geijn, F. J. M. Caron, J. M. Swartjes, E. E. van Woerden, and H. W. Jongsma, "Computer analysis of antepartum fetal heart rate: 1. Baseline determination," *Int. J. Biomed. Comput.*, vol. 25, no. 4, pp. 261–272, May 1990.
- [11] —, "Computer analysis of antepartum fetal heart rate: 2. Detection of accelerations and decelerations," *Int. J. Biomed. Comput.*, vol. 25, no. 4, pp. 273–286, May 1990.
- [12] M. Mongelli, R. Dawkins, T. Chung, D. Sahota, J. A. D. Spencer, and A. M. Z. Chang, "Computerized estimation of the baseline fetal heart rate in labour: the low frequency line," *Br. J. Obstet. Gynaecol.*, vol. 104, no. 10, pp. 1128–1133, Oct. 1997.
- [13] G. M. Taylor, G. J. Mires, E. W. Abel, S. Tsantis, T. Farrell, P. F. W. Chien, and Y. Liu, "The development and validation of an algorithm for real time computerized fetal heart rate monitoring in labour," *Br. J. Obstet. Gynaecol.*, vol. 107, pp. 1130–1137, Sep. 2000.
- [14] G. S. Dawes, M. Moulden, and C. W. Redman, "Computerized analysis of antepartum fetal heart rate," *Am. J. Obstet. Gynecol.*, vol. 173, no. 4, pp. 1353–1354, 1995.
- [15] J. Jezewski and J. Wrobel, "Foetal monitoring with automated analysis of cardiotocogram: the KOMPOR system," in *Proc. 15th Ann. Card. IEEE/EMBS*, San Diego, CA, 1993, pp. 638–639.
- [16] K. Maeda, "Computerized analysis of cardiotocograms and fetal movements," *Bailliere's Clin. Obstet. Gynaecol.*, vol. 4, no. 4, pp. 1797–1813, Dec. 1990.
- [17] C. Ulbricht, G. Dorffner, and A. Lee, "Neural networks for recognizing patterns in cardiotocograms," *Artif. Intell. Med.*, vol. 12, pp. 271–284, 1998.
- [18] O. Fontenla-Romero, A. Alonso-Betanzos, and B. Guijarro-Berdinas, "Adaptive pattern recognition in the analysis of cardiotocographic records," *IEEE Trans. Neural Netw.*, vol. 12, no. 5, pp. 1188–1195, Sep. 2001.
- [19] J. Bernardes, C. Moura, J. P. M. de Sa, and L. Pereira-Leite, "The Porto system for automated cardiotocographic signal analysis," *J. Perinat. Med.*, vol. 19, pp. 61–65, 1991.
- [20] J. Bernardes, C. Moura, J. P. M. de Sa, L. Pereira-Leite, and H. P. van Geijn, "The Porto system," in *A Critical Appraisal of Fetal Surveillance*, H. P. van Geijn and F. J. A. Copray, Eds. New York: Elsevier Science, 1994, pp. 315–324.
- [21] D. Ayres-de-Campos, J. Bernardes, A. Garrido, J. P. M. de Sa, and L. Pereira-Leite, "SisPorto 2.0: a program for automated analysis of cardiotocograms," *J. Matern. Fetal Med.*, vol. 9, pp. 311–318, 2000.
- [22] G. Magenes, M. G. Signorini, and D. Arduini, "Classification of cardiotocographic records by neural networks Neural Networks," in *Proc. IEEE-INNS-ENNS Int. Joint Conf. on Neural Networks (IJCNN'00)*, 2000, vol. 3, pp. 637–641.
- [23] —, "Multiparametric analysis of fetal heart rate: comparison of neural and statistical methods," in *Proc. Medicon 2001*, pp. 360–363.
- [24] S. Kol and I. Thaler, "Interpretation of nonstress tests by an artificial neural network," *Am. J. Obstet. Gynecol.*, vol. 172, no. 5, pp. 1372–1379, May 1995.
- [25] T. K. H. Chung, M. P. Mohajer, X. J. Yang, A. M. Z. Chang, and D. S. Sahota, "The prediction of fetal acidosis at birth by computerized analysis of intrapartum cardiotocography," *Br. J. Obstet. Gynaecol.*, vol. 102, pp. 454–460, Jun. 1995.
- [26] E. Salamalekis, P. Thomopoulos, D. Giannaris, I. Salloum, G. Vassios, A. Prentza, and D. Koutsouris, "Computerised intrapartum diagnosis of fetal hypoxia based on fetal heart rate monitoring and fetal pulse oximetry recordings utilising wavelet analysis and neural networks," *Br. J. Obstet. Gynaecol.*, vol. 109, no. 10, pp. 1137–1142, Oct. 2002.
- [27] Z. R. Struzik and W. J. van Wijngaarden, Cumulative Effective Hölder Exponent Based Indicator for Real Time Fetal Heartbeat Analysis During Labour Rep. INS-R0110, Nov. 2001.
- [28] E. C. Ifeachor, R. D. F. Keith, J. Westgate, and K. R. Greene, "An expert system to assist in the management of labour," in *Proc. World Congr. Expert Systems*, 1991, vol. 4, pp. 2615–2622.
- [29] J. F. Skinner, J. M. Garibaldi, and E. C. Ifeachor, "A fuzzy system for fetal heart rate assessment," in *Proc. 6th Fuzzy Days on Computational Intelligence*, Dortmund, Germany, 1999, pp. 20–29.
- [30] A. Alonso-Betanzos, V. Moret-Bonillo, L. D. Devoe, J. R. Searle, B. Baniyas, and E. Ramos, "Computerized antenatal assessment: the NST-EXPERT project," *Automedica*, vol. 14, pp. 3–22, 1992.
- [31] A. Alonso-Betanzos, B. Guijarro-Berdinas, V. Moret-Bonillo, and S. Lopez-Gonzalez, "The NST-EXPERT project: the need to evolve," *Artif. Intell. Med.*, vol. 7, no. 4, pp. 297–313, 1995.
- [32] B. Guijarro-Berdinas, A. Alonso-Betanzos, and O. Fontenla-Romero, "Intelligent analysis and pattern recognition in cardiotocographic signals using a tightly coupled hybrid system," *Artif. Intell.*, vol. 136, pp. 1–27, 2002.
- [33] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.
- [34] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [35] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
- [36] K. R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 2, no. 2, pp. 181–201, Mar. 2001.
- [37] H. Byun and S. W. Lee, "Application of support vector machines for pattern recognition: A survey," in *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 2002, vol. 2388, Proc. 1st Int. Workshop SVM 2002, pp. 213–236.
- [38] K. Veropoulos, N. Cristianini, and C. Campbell, "The application of support vector machines to medical decision support: a case study," in *Advanced Course in Artificial Intelligence (ACAI'99)*, Jul. 1999.
- [39] R. P. W. Duin, "Classifiers in almost empty spaces," in *Proc. 15th International Conference on Pattern Recognition (ICPR'00)*, Barcelona, Spain, Sep. 3–8, 2000, vol. 2, pp. 1–7.
- [40] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
- [41] E. E. Osuna, R. Freund, and F. Girosi, Support Vector Machines: Training and Applications MIT, A.I. Memo no. 1602, March 1997.
- [42] K. Veropoulos, C. Campbell, and N. Cristianini, "Controlling the sensitivity of support machines," in *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI99)*, Stockholm, Sweden, 1999, pp. 55–60.
- [43] M. C. Carter, "Present-day performance qualities of cardiotocographs," *Br. J. Obstet. Gynaecol.*, vol. 100, pp. 10–14, Mar. 1993, suppl. 9.
- [44] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart rate variability. Standards of measurements, physiological interpretation, and clinical use," *Eur. Heart J.*, vol. 17, no. 3, pp. 354–381, 1996.
- [45] E. N. Bruce, *Biomedical Signal Processing and Signal Modeling*. New York: Wiley, 2001.
- [46] O. Sibony, J. P. Fouillot, M. Benaudias, A. Benhalla, P. Blot, and C. Sureau, "Spectral analysis: a method for quantitating fetal heart rate variability," in *A Critical Appraisal of Fetal Surveillance*, H. P. van Geijn and F. J. A. Copray, Eds. New York: Elsevier Science, 1994, pp. 325–332.
- [47] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1999.
- [48] K. K. Lee, S. R. Gunn, C. J. Harris, and P. A. S. Reed, "Classification of imbalanced data with transparent kernels," in *Proc. INNS-IEEE Int. Joint Conf. Neural Networks*, 2001, pp. 2410–2415.

- [49] S. Perkins, N. R. Harvey, S. P. Brumby, and K. Lacker, "Support vector machines for broad area feature extraction in remotely sensed images," *Proc. SPIE*, vol. 4381, Apr. 2001.
- [50] R. Akbani, S. Kwek, and N. Japkowicz, "Applying support vector machines to imbalanced datasets," in *Proc. Eur. Conf. Machine Learning (ECML'2004)*, Pisa, 2004, pp. 39–50.
- [51] M. F. Docker, "Doppler ultrasound monitoring technology," *Br. J. Obstet. Gynaecol.*, vol. 100, pp. 454–460, Mar. 1993, suppl. 9.
- [52] R. D. Berger, S. Akselrod, D. Gordon, and R. J. Cohen, "An efficient algorithm for spectral analysis of heart rate variability," *IEEE Trans. Biomed. Eng.*, vol. 33, no. 9, pp. 900–904, Sep. 1986.
- [53] M. Bracale, M. Romano, M. Cesareli, P. Bifulco, and M. Sansone, "Cardiotocographic data pre-processing and AR modeling of fetal heart rate signals," presented at the World Congress on Medical Physics and Biomedical Engineering, Sydney, Australia, Aug. 24–29, 2003, Paper 3363, unpublished.
- [54] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Belmont, CA: Wadsworth Int. Group, 1984.
- [55] S. L. Salzberg, "On comparing classifiers: pitfalls to avoid and a recommended approach," *Data Mining Knowledge Discovery*, vol. 1, pp. 317–328, 1997.
- [56] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.
- [57] M. Kubat and S. Matwin, "Addressing the curse of imbalanced data sets: One-sided sampling," in *Proc. 14th Int. Conf. Machine Learning*, 1999, pp. 179–186.
- [58] F. Provost and T. Fawcett, "Analysis and visualization of classifier performance: comparison under imprecise class and cost distribution," in *Proc. 3rd Int. Conf. Knowledge Discovery and Data Mining*, 1997, pp. 43–48.
- [59] P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, 1997.
- [60] G. S. Dawes, "Computerized fetal heart rate analysis," in *A Critical Appraisal of Fetal Surveillance*, H. P. van Geijn and F. J. A. Copray, Eds. New York: Elsevier Science, 1994, pp. 311–314.
- [61] J. Bernardes, D. Ayres-de-Campos, A. Costa-Pereira, L. Pereira-Leite, and A. Garrido, "Objective computerized fetal heart rate analysis," *Int. J. Gynecol. Obstet.*, vol. 62, pp. 141–147, 1998.
- [62] B. K. Strachan, D. S. Sahota, W. J. van Wijngaarden, D. K. James, and A. M. Z. Chang, "Computerized analysis of the fetal heart rate and relation to acidemia at delivery," *Int. J. Obstet. Gynaecol.*, vol. 108, no. 8, pp. 848–852, August 2001.
- [63] G. Georgoulas, G. Nolas, C. Stylios, and P. P. Groumpos, "Classification of fetal heart rate during labour using Hidden Markov Models," in *Proc. IEEE Int. Joint Conf. on Neural Networks*, Budapest, Hungary, Jul. 25–29, 2004, vol. 3, pp. 2471–2476.



George Georgoulas received the Diploma degree in electrical engineering from the University of Patras, Patras, Greece, in 1999. He is currently working toward the Doctoral degree at the University of Patras.

His scientific interests include biomedical signal processing and machine knowledge.



Chrysostomos Stylios (M'96) received the Diploma degree in electrical engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 1992 and the Ph.D. degree from Department of Electrical and Computer Engineering, University of Patras, Patras, Greece in 1999.

From 2000 through 2004, he was with the faculty of the Computer Science Department, University of Ioannina, Ioannina, Greece, as adjunct Assistant Professor. Since 1999, he is a Senior Researcher with the Laboratory for Automation and Robotics, University of Patras. In February 2005 he was elected Assistant Professor at Department of Communications, Informatics, and Management, Technological Education Institute of Epirus, Greece. He has published over 60 journals and conference papers, book chapters, and technical reports. His research interests include computational intelligent techniques, modeling of complex systems, intelligent systems, decision support systems, hierarchical systems, supervisory control, and artificial intelligence techniques for medical applications. He is a member of the National Technical Chamber of Greece since 1992.



Peter P. Groumpos (S'73–M'78–SM'04) received the Ph.D. degree in electrical engineering from the State University of New York at Buffalo in 1978.

He is professor in the Department of Electrical and Computer Engineering at the University of Patras, Patras, Greece. He is president and CEO of Patras Science Park and Director of the Laboratory for Automation and Robotics. He was formerly on the faculty at Cleveland State University, Cleveland, OH, from 1979 through 1989. He was the director of the Communication Research Laboratory from

1981 through 1986 and a member of the Technical Committee of the Advanced Manufacturing Center from 1985 through 1987 at Cleveland State University. He participated on a Technology Transfer Program with the Ministry of Higher Education of Egypt from 1981 to 1984. He was an Associate Editor for Book Reviews for the *IEEE Control Systems* magazine from 1980 through 1985. For the academic year 1987–1988, he was a Fulbright visiting scholar at the University of Patras. He was the Greek National Representative to ESPRIT (1990–94). Presently he is the National Representative to the High-Level Group for EUREKA and to the IST program; consultant to a number of companies in the USA and Greece. He is the Greek NMO representative to IFAC. He has published over 150 journals and conference papers, book chapters, and technical reports. His main research interests are modeling of complex systems, intelligent manufacturing systems, process control, robotics, simulation methods, theories of hierarchical and large-scale systems, intelligent control, soft computing methods, and bioinformatics.

He is an Associate Editor for the international journals *Computers in Industry* and *Studies in Informatics and Control*. He is a member of the Honorary Societies Eta Kappa Nu and Tau Beta Pi. He was the Coordinator of the ESPRIT Network of Excellence in Intelligent Controls and Integrated Manufacturing Systems (ICIMS-NOE) and he was the Editor-in-Chief of ICIMS-NEW's for the period of 1994–2000. He was organizing every summer the Advanced Summer Institute (ASI) for the period 1992–2000 and has organized as a General Chairman more than ten International Conferences been held in Greece or on other parts of Europe.