Classification of Cardiotocograms based on Independent Component Analysis and Support Vector Machines

George Georgoulas[1], Chrysostomos Stylios[2] and Peter P. Groumpos[1]

[1]LAR, Electrical & Computer Engineering Dept., University of Patras
26100 Patras, Hellas
[2]Computer Science Depart., University of Ioannina,
P.O. Box 1186, 455100, Ioannina, Hellas

e-mail:{georgoul@ee.upatras.gr and stylios@cs.uoi.gr}

Abstract. In this paper we present a novel method for classification of Fetal Heard Rate (FHR) the subcomponent of cardiotocogram (CTGs) based on a novel approach for feature extraction and classification. The feature extraction is implemented by means of Independent Component Analysis (ICA), and for the categorization, Support Vector Machines (SVM) are employed. In this fir introductory study we achieved a classification performance of 70%, which is quite promising for the daunting task of FHR classification

1 Introduction

Electronic Fetal Monitoring (EMF) has been widely used for antepartum and intrapartum fetal surveillance. EMF refers to the continuous recording and monitoring of Fetal Heart Rate (FHR) and Uterine Activity (UA), also known as cardiotocogram (CTG). Figure 1 shows a typical CTG, with the FHR at the upper window and the UA at the lower. During the crucial period of labor, CTG is used as the main screening test on fetal acid base balance [1]. Actually, the avoidance of metabolic acidosis is the desired outcome on the observation, evaluation and justification of CTG. The instantaneous FHR (beats/min) is the subtlest component of the CTG. [2].

Although CTG has been used for the last 4 decades, there is controversy regarding its efficiency. Moreover, the Dublin randomized trial has revealed an increase in operative vaginal deliveries in patients monitored using CTG during the intrapartum period [3]. In addition, studies of CTG reliability have shown significant inter-observer and intra-observer variation in tracing interpretation [4].

The inconsistency in interpretation and the increase of false positive diagnosis and, additionally, the technological advances in computers along with new signal processing methods have prompted many researches to develop computer systems capable of

actually monitor and infer on the condition of the fetus in a reliable, effective and reproducible manner.

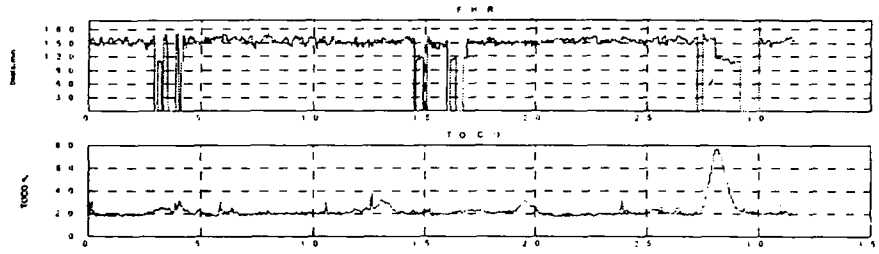

Fig. 1. A typical print out cardiotocogram.

Based on the belief that the CTG signal –and especially the FHR component- may convey much more information than what is usually interpreted by doctors, we propose a method to classify FHR and discriminate fetal acidosis, based on features extracted mathematically from the FHR signal (ignoring in this study the UA signal). The core of the proposed methodology is the use of Independent Component Analysis (ICA) for the extraction of a novel set of features and the use of Support Vector Machines (SVM) for the classification task.

ICA has found applications in many fields of multivariate signal processing. It has been used, among others, for speech separation [19], analysis of biomedical signals [20], and detection of trends in stock markets [21]. In this paper we use ICA for the extraction of independent "sources" which will be subsequently used as the representational basis for the FHR signal.

SVMs have gained great attention during the last decade and they have been used extensively in the field of pattern recognition [22]. Support vector machines can be reliable classifiers even when the sample population is small. Because our sample population is indeed small, SVMs were the classifiers selected to be used after the feature extraction stage.

This paper is structured as follows: section 2 presents a brief overview of ICA and how it is applied to estimate the independent components from a set of observed signals and, additionally, an introduction to SVMs. Section 3 describes analytically the proposed methodology. In section 4, the experimental results are presented, and section 5 outlines the conclusions of this study.

2 Background information

2.1 Independent Component Analysis

ICA has been mainly associated with Blind Signal Separation (BSS) [19] (the term blind indicates that neither the sources nor the way they were mixed are known). However, it has become clear that the principle underlying ICA has a lot of other

application areas of interest as well [23,24]. After all, ICA is a theoretical method and BSS is an application that can be solved with many methods, including ICA. Among the different applications of ICA, in this paper we will focus on the application of ICA as a mean for feature extraction from a set of signals.

In the basic ICA model, from a set of n observed signals, we seek for n statistically independent sources, which are linearly combined to produce the observed signal(s). Both the independent sources and their linear combination are unknown. Strictly speaking, we have the following generative model for the data

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{1}$$

where $\mathbf{x}$ is the observed (zero-mean) n-dimensional vector, $\mathbf{s}$ is a n-dimensional random vector, whose components are assumed mutually independent, and $\mathbf{A}$ is a nonsingular n-by-n constant matrix (for simplicity and without loss of generality we have assumed that the number of observations and the number of sources are equal).

The goal of ICA is to find a separating matrix $\mathbf{W}$ that, after multiplication with the observation vector $\mathbf{x}$, will retrieve the source vector $\mathbf{s}$. Though this is not feasible, however it is feasible to retrieve an arbitrarily scaled and permutated "version" of the original set of sources.

$$\mathbf{y} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s} \rightarrow \mathbf{D}\mathbf{P}\mathbf{s} \tag{2}$$

where D is a nonsingular diagonal matrix and P is a permutation matrix.

The problem of finding the separating matrix can be simplified by performing a preliminary sphering (or whitening) of the data [23]

There are currently many approaches to perform ICA [23]. In our experiments we used the Matlab toolbox "Fast ICA" [22]. FastICA, as implemented in this toolbox, uses algorithms close to Newton method rather than to a fixed-point iteration. For a more in depth explanation of the FastICA algorithm the reader can refer to [25]

2.2 Support Vector Machines

SVMs are a new family of learning machines. The main idea behind SVMs, when dealing with a pattern classification problem, is to preprocess the data in order to represent patterns in a high dimension – typically much higher than the original feature space via a nonlinear mapping $\varphi(\cdot)$. The training of a SVM consists of finding the optimal hyperplane, that is, the one with the maximum distance from the nearest training patterns. The support vectors are exactly those patterns. However, real life problems are rarely separable and, so, there always exists a number of misclassifications.

Formally speaking, given a training set $S = \{(\mathbf{x}_i, y_i)\}_{i=1}^{l}$, where each point $\mathbf{x}_i$ is a p-dimensional vector, the input pattern for the I-th example, and $y_i \in \{-1,1\}$ is a

label that specifies to which one of the classes the point $\mathbf{x}_i$ belongs to, the goal is to find a discriminating function of the form

$$f(\mathbf{x}) = sign(\mathbf{w} \cdot \boldsymbol{\varphi}(\mathbf{x}_i) + b) \tag{3}$$

where $\boldsymbol{\varphi}(\mathbf{x}) = (\phi_1(\mathbf{x}), ..., \phi_m(\mathbf{x}))$ corresponds to a mapping from $\Re^l$ to the higher dimensional space $\Re^m$ ($d > l$). The search for the "optimum" hyperplane and, thus, the best classifier leads to the following quadratic optimisation problem

$$\text{Minimize } \frac{1}{2}\mathbf{w}^T \cdot \mathbf{w} + C\sum_{i=1}^{l} \xi_i \tag{4}$$

$$\text{Subject to } y_i(\mathbf{w} \cdot \boldsymbol{\varphi}(\mathbf{x}_i) + b) \geq 1 - \xi_i \qquad \xi_i \geq 0,\ i = 1, 2, ..., l \tag{5}$$

The dual problem, which is in fact the one to be solved is

$$\text{Maximize } \sum_{i=1}^{l} a_i - \frac{1}{2}\sum_{i,j=1}^{l} a_i a_j y_i y_j \boldsymbol{\varphi}(\mathbf{x}_i)^T \boldsymbol{\varphi}(\mathbf{x}_j) \tag{6}$$

$$\text{Subject to } \sum_{i=1}^{N} y_i \alpha_i = 0 \quad (16) \tag{7}$$

$$0 \leq \alpha_i \leq C \ , \ i = 1, 2, ..., l$$

Parameter C controls the influence of training data points that will remain on the wrong side of a separating nonlinear hypersurfaces (hyperplanes) in the feature space. The discriminating function is finally given by

$$f(\mathbf{x}) = sign\left(\sum_{i=1}^{l} y_i \alpha_i \boldsymbol{\varphi}(\mathbf{x}_i) \cdot \boldsymbol{\varphi}(\mathbf{x}) + b\right) \tag{8}$$

The points for which a_i>0 are called Support Vectors. They are the most difficult patterns to classify (including the points on the boundary) and usually are a small portion of the training set.

If the nonlinear mapping function is chosen properly, the inner product in the feature space can be written in the following form

$$\boldsymbol{\varphi}^T(\mathbf{x}_i) \cdot \boldsymbol{\varphi}(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j) \tag{9}$$

where K is called the inner-product kernel. A kernel function is a function in input space and, therefore, we do not perform the nonlinear mapping $\boldsymbol{\varphi}(\cdot)$. Instead of performing the inner products in a feature space $\boldsymbol{\varphi}^T(\mathbf{x}_i) \cdot \boldsymbol{\varphi}(\mathbf{x}_j)$, one can directly calculate it using the kernel function $\boldsymbol{\varphi}^T(\mathbf{x}_i) \cdot \boldsymbol{\varphi}(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j)$. Therefore by

selecting an appropriate symmetric kernel function one does not even have to know what the actual mapping is [22].

3 The proposed methodology

3.1 Data set

Data have been acquired with a Hewlett Packard fetal monitor that performs sampling at 4 Hz (4 samples/s). All cardiotocographic records had been acquired during the final stage of the labor and, in fact, as close as possible to delivery. This means that the data sets are time-biased free and a direct association can be made between the segment of the signal used and the fetal outcome.

The main goal of the proposed method is to classify FHR. The experimental data have been categorized according to the umbilical arterial blood pH. The umbilical arterial blood pH and/or the apgar score, measured after delivery, are objective indexes [27] that assess fetal well being. Even though it could be argued that those indexes are affected by the specific handling of the delivery, this is still the best method towards an unbiased characterization of the newborn's condition.

The data set consisted of 40 signals and was divided into 2 subsets. In the first subset we classified those signals that belonged to fetuses with umbilical arterial blood pH less than 7.1 and in the second those that belonged to fetuses with umbilical arterial blood pH more than 7.2. It is known that babies with severe acidosis (pH 7.0 or less) will subsequently be normal in a percentage of 90%. However, these can be considered immediate outcomes that one would prefer to avoid [28].

3.2 Pre-processing

FHR is a very noisy signal with a lot of spiky artifacts and even periods of missing data due to the movement of the baby and the stress induced during the labor, leading to the displacement of the transducer used to acquire it. This kind of noise cannot be eliminated and is always present in cardiotocographic records (Figure 2).

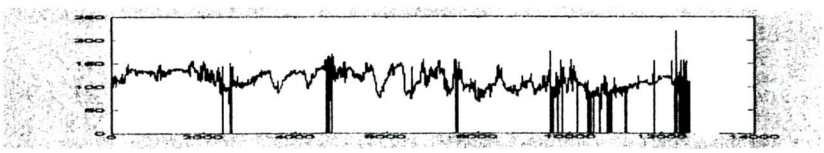

Fig. 2. Original FHR signal (the presence of noise id obvious).

Those spiky or missing segments must be removed before subsequent processing takes place. For the removal of the artifacts we used the algorithm firstly introduced in [14]. After the removal of artifacts, we proceeded to the selection of a 20-minute segment from each of the FHR signals. The segments were chosen to be as close to the final stage of the recording as possible. By doing so, we try to eliminate time bias

and also to ensure, in a way, the direct correspondence of the segment to fetal condition – which was qualified by the arterial blood pH value. The duration of the 20 minutes was selected because some of the recordings lasted no more than 20 minutes. In the final stage we subtracted the mean value of each signal (so as to have zero-mean observation signals) and we came up with a segment like the one depicted in Figure 3.

After this pre-processing step, we are ready to proceed to the main task of the categorization. This task can be divided into 3 steps:(a) Dimensionality reduction employing PCA, (b) Feature extraction using the signal from the stage of PCA and applying ICA on them and, finally, (c) Classification based on the features (extracted from the previous step) using SVMs.

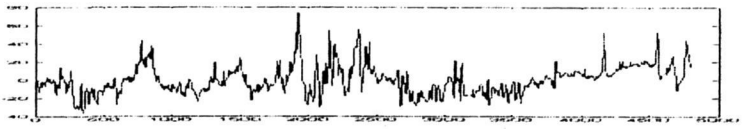

Fig. 3. Signal after pre-processing

3.3 Dimensionality reduction

The number of features is exactly the number of the Independent Components. Therefore if we had retained all the Independent Components for each one of the n signals, we would have associated (as it will be shown bellow) a feature vector of dimension equal to n, making the training task extremely difficult. Thus, dimensionality reduction is performed in order to reduce the final feature vector.

To perform the dimensionality reduction, we arranged the signals in a $n \times N$ matrix $\mathbf{x}_0$ (each row having zero mean). Then we calculated the $n \times n$ covariance matrix $\mathrm{E}\{\mathbf{x}_0 \mathbf{x}_0^T\}$ and stored the eigenvectors corresponding to the largest m eigenvalues as columns of a $n \times m$ matrix $\mathbf{E}_{\mathbf{PCA}}$ in descending order. By projecting the original signals on the subspace spanned by those eigenvectors $\mathbf{E}_{\mathbf{PCA}}^T \mathbf{x}_0$, we had a dimensionality reduction from n to m. This new set of signals, the $m \times N$ matrix, was then whitened and the "white" or "sphered" data were used for the extraction of the Independent Components.

3.4 Feature extraction

The main idea behind the application of ICA was to use the independent sources as a representational basis for the FHR that we had at our disposal. This means that we look for the linear combination of the sources that best (in a mean square sense) reconstruct the original FHR signal. The linear coefficients, corresponding to each signal, are derived from the solution of the following over-determined algebraic problem.

$$Coef_i \mathbf{y} = x_{0_i} \tag{10}$$

The solution was obtained using the pseudo-inverse method

$$Coef_i = x_{0_i} \mathbf{y}^{\#} \text{ where, } \mathbf{y}^{\#} = \mathbf{y}^T \left(\mathbf{y}\mathbf{y}^T \right)^{-1} \tag{11}$$

3.5 SVM classifier

Depending on how the inner-product kernel is generated, different learning machines can be constructed with quite different non-linear decision surfaces. In particular, the support vector-learning algorithm can be employed to construct the following (among others) types of learning machines: polynomial learning machines, radial basis function networks, and two layer perceptrons. In this study we used only polynomial learning machines.

As mentioned, the parameter C is a user defined variable. It can be determined either experimentally, using a training-validation procedure, or analytically, via the estimation of the Vapnick-Chervonenkis (VC) dimension [18]. In our case we determined parameter C using the experimental procedure. Thus, we tested various configurations and combinations of learning machines and input vectors.

The data set consisted of 40 cases. 20 cases with pH less than 7.1 (the risk group) and 20 cases with pH greater than 7.2 (the normal group). Because the available set of labeled data is restricted to 40 cases, in order to test the performance of our classification scheme, we used multifold cross-validation [18]. We divided the 40 cases into 5 (non-overlapping) subsets, each one with 4 examples from the "normal" and 4 from the "risk" group. The SVM classifier was trained on all subsets except for one, and the validation error was measured by testing it on the subset left out. We repeated this procedure 5 times, each time using a different subset for testing.

4 Experimental results

We experimented using 2 to 9 independent components for various values of the parameter $C \in (0,1]$ and for polynomials with degree ranking from 2 to 6. We achieved an overall classification performance of 70% (80% for the normal cases and 60% for the risk cases) (Figure 4) for 5^{th} order polynomial kernel. Individually, we managed to reach a classification performance of 90% for the normal cases, but with very bad performance for the risk group (using 2^{nd} order polynomials) (Figure 5). On the other hand, we achieved a classification rate for the risk cases of 70% and 60% for the normal group (meaning an overall classification of 65%), using polynomial kernels of degree 3 (Figure 6) or even 75% for the risk cases, but with very low per-

formance for the normal cases 55% (overall 65%) (Figure 7). In the following figures, classification performance is presented for different values of C.

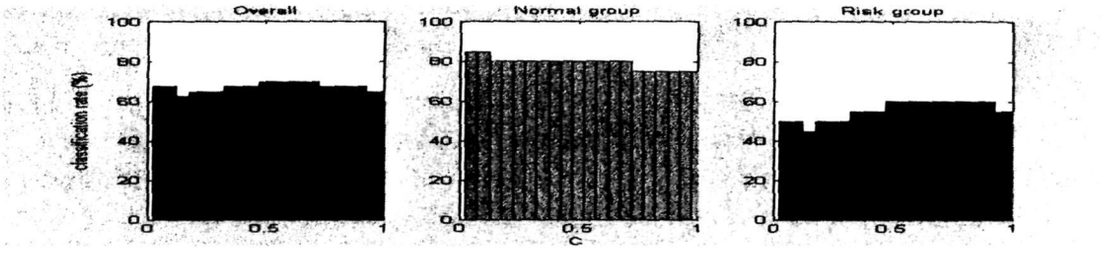

Fig. 4. 5th degree polynomial Kernels and 2 independent components

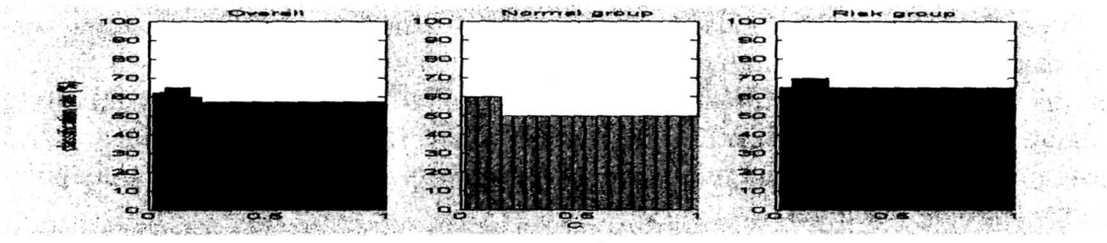

Fig. 5. 3rd degree polynomial Kernels and 2 independent components

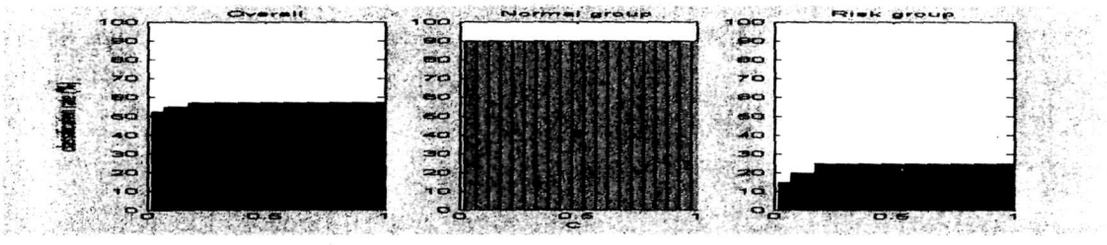

Fig. 6. 2nd degree polynomial Kernels and 2 independent components

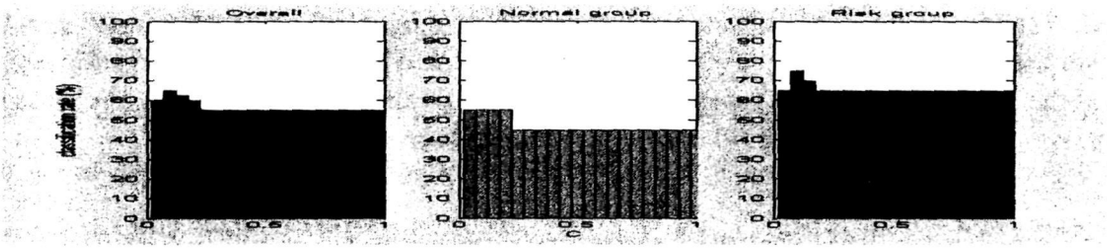

Fig. 7. 2nd degree polynomial Kernels and 2 independent components

4 CONCLUSIONS

In this paper we introduce a novel method based on the application of SVMs and ICA in order to test whether we can find a way to discriminate between fetuses with "normal" pH values and those who have a decreased pH and are suspicious of developing acidosis.

The overall procedure showed that, even though the classification of the normal cases is quite high, the classification rate of the risk cases is not quite satisfactory. The same result was derived in other studies, where a neural network instead of a SVM had done the classification task. This possibly indicates that the choice of the threshold for the pH value should be lower for the risk case. A more justified threshold could be the value of pH at 7, but this could compromise more the classification performance since only 2 cases would fulfill that criterion, leaving 38 to the normal set

In a similar attempt [10] where the cut-off point for the umbilical arterial blood pH was set to 7.15 and in a population of 73 fetuses, (8 fetuses with pH less than 7.15 and 65 fetuses with pH more than 7.15) the developed system managed to classify 7 of the abnormal cases to the right class (87.5% classification rate) and 50 of the normal cases to the right class (76.92% classification rate) respectively. (Overall classification performance 78.08%). Nevertheless, a number of problems have been reported, concerning the evaluation of these results, especially due to the high value of the selected cut-off. [29]. The selection of a lower threshold would lead to much worse results, and thus, no direct comparison can be made

In view of all the above, we will test the proposed method to a larger set of data to further validate it and we will also use the fetal apgar score as another index component for the formation of the classes, something which was not used in this present study, since all apgar scores but one were higher than 8.

References

1. Geijn, H. P. V.: Developments in CTG analysis. Bailliers Clin. Obstet. Gynaecol., Vol 10 no 2, (1996) 185-209
2. Carter, M. C.:Present-day performance qualities of cardiotocographs. Br. J. Obstet, Gynecol., vol. 100, Supplement 9,(1993) 10-14
3. MacDonald, D., Grant, A., Sheridan-Pereira, M., Boylan, M. P. and Chalmers I.:The Dublin randomized controlled trial of intrapartum fetal heart rate monitoring. Am. J. Obstet Gynecol. 152 (1985) 524-39
4. Donker, D. K. Geijn, H. P. V. and Hasman A.:Interobserver variation in the assessment of fetal heart rate recordings. European Journal of Obstetrics, Gynaecology, and Reproductive Gynecology, vol. 52 (1993) 21-28
5. Mantel, R. Geijn, H. P. V, Caron, F.J.M., Swartjes, J. M., Woerden, E. E. V. and Jongsma, H. W.:Computer analysis of antepartum fetal heart rate: 1. Baseline determination. Int. J. Biomed. Comput., vol. 25, no. 2, (1990) 261-272
6. Mantel, R. Geijn, H. P. V, Caron, F.J.M., Swartjes, J. M., Woerden, E. E. V. and Jongsma, H. W.: Computer analysis of antepartum fetal heart rate: 2. Detection of accelerations and decelerations. Int. J. Biomed. Comput., vol. 25, no. 2, (1990) 273-286
7. Maeda, K.:Computerized analysis of cardiotocograms and fetal movements. Bailliers Clin. Obstet. Gynaecol. vol. 4, no. 4, (1990) 1797-813
8. Jezewski, J. and Wrobel, J.: Foetal monitoring with automated analysis of cardiotocogram: The KOMPOR system. *15th Ann. Card. IEEE/EMBS.* San Diego (1993) 638-639
9. Berdinas, B. G., Betanzos, A. A., and O. F. Romero, O. F.:Intelligent analysis and pattern recognition in cardiotocographic signals using a tightly coupled hybrid system. Artificial Intelligence. vol. 136, (2002) 1-27

10.Chung, T. K. H., Mohajer, M. P. , Yang, X. J. , Chang A. M. Z. and Sahota, D.S.:The prediction of fetal acidosis at birth by computerized analysis of intrapartum cardiotocography. Br. J. Obstet. Gynaecol., vol. 102. (1995) 454-460
11.Krause, W.: A computer aided monitoring system for supervision of labour. Computers in Perinatal Medicine, Maeda, K. (eds) Elsevier Science Publishers. (1990). 103-111
12. Arduini, D., Rizzo, G., Piana, G., Bonalumi, A., Brambilla, P. and Romanini, C.: Computerized Analysis of Fetal Heart Rate: I. Description of the System (2CTG). J. Maternal Fetal Invest. vol. 3. (1993). 159-163
13.Taylor, G. M., Mires, G. J., Abel, E. W., Tsantis, S., Farrell, T., Chien, P. F. W. and Liu, Y.:The development and validation of an algorithm for real time computerized fetal heart rate monitoring in labour. Br. J. Obstet. Gynaecol, vol, 107. (2000). 1130-1137
14. Bernardes, J., Moura, C., Sa, J.P.M. and Leita, L.P.:The Porto system for automated cardiotocographic signal analysis. J. Perinat. Med. vol. 19. (1991) 61-65
15. Dawes, G. S., Moulden, M. and Redman, C. W.:Computerized analysis of antepartum fetal heart rate. *Amer. J. Obstet. Gynecol.*, vol. 173. no. 4. (1995). 1353-1354
16. Salamalekis, E., Thomopoulos, P., Giannaris, D., Salloum, I., Vasios, G., Prentza, A. and Koutsouris D.: Computerised intrapartum diagnosis of fetal hypoxia based on fetal heart rate monitoring and fetal pulse oximetry recordings utilising wavelet analysis and neural networks. British Journal of Obstetrics and Gynaecology. vol. 109. no. 10. (2002). 1137-1142
17. Skinner, J. F., Garibaldi, J. M. and Ifeachor E. C.: A Fuzzy System for Fetal Heart Rate Assessment. 6th Fuzzy Days Conference. Dortmund. (1999). 20-29
18. Haykin, S.: Neural Networks: A Comprehensive Foundation. Englewoo Cliffs, NJ: Prentice Hall, (1999)
19. Bell, A.J. and Sejnowski, T.J.: An information-maximization approach to blind separation and blind deconvolution. Neural Computation. Vol. 7. (1995). 1129-1159
20. Makeig, S., Jung, T.P., Bell, A. J., Ghahramani, D. and Sejnowski, T. J.: Blind separation of auditory event-related brain responses into independent components. National Academy of Science (USA), (1997). 10979-10984
21. Back, A. and Weigend, A.: A first application of independent component analysis to extracting structure from stock returns. Int. Journal of Neural Systems. vol. 8. no. 4. (1997). 473-484
22. Burges, C. J. C.: A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery. vol.2, (1998). 121–167 (1998)
23. Hyvarinen, A.:Survey on Independent Component Analysis. Neural Computing Surveys, vol. 2. (1999). 94-128
24. Hyvarinen, A. and Oja, E.: Independent component analysis: algorithms and applications. Neural Networks. vol. 13. (2000). 411-430
25. FastICA MATLAB Package [Online].: HTTP: http://www.cis.hut.fi/projects/ica/fastica
26. Hyvärinen, A.: Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. IEEE Transactions on Neural Networks vol. 10 (3) (1999) 626-634
27. Bernardes, J., Campos, D. A., Pereira, A. C., Leita L.P. and Garrido A.: Objective computerized fetal heart rate analysis. Int. J. Gynecol. Obst.vol. 62.(1998).141-147
28. Parer, J. T.: Handbook of fetal heart rate monitoring, Philadelphia, Pennsylvania: W. B. Saunders Company. (1997)29. Greene, K. R.: Intelligent fetal heart rate computer systems in intrapartum surveillance. Current Opinion in Obstetrics and Gynecology, vol 8, (1996) 123-127