Integrated Approach for Classification of Cardiotocograms based on Independent Component Analysis and Neural Networks

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

***Abstract*--Nowadays fetal monitoring is based on the acquisition and interpretation of the Cardiotocogram (CTG). There is an ongoing effort to develop advanced methods and computer based systems to assist the obstetricians in the difficult task of feature extraction and the classification of the CTG. This paper describes an integrated methodology for CTG classification, introducing the reduction of the dimensionality of the input data space, using Independent Component Analysis (ICA). The milestone of the method is the utilization of the rows of the estimated mixing matrix after the implementation of ICA method as feature vectors, which are subsequently fed to a feed-forward Neural Network (NN) classifier, which categorizes the CTG.**

***Index terms*-- Cardiotocogram, Independent Component Analysis, Principle Component Analysis, Neural Networks**

I. INTRODUCTION

For the past 35 years Electronic Fetal Monitoring (EMF) has been used worldwide for antepatrum and intrapartum fetal surveillance. By the term EMF we almost exclusively mean the continuous recording of fetal heart rate (FHR) and uterine activity (UA), which is also referred as cardiotocogram (Figure 1). The major observation on CTG ensures that if the FHR signal is normal then the obstetrician is almost sure that the fetus is in a good condition.

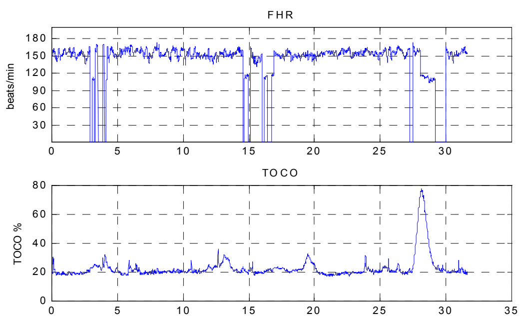

Figure 1. A printed sample of cardiotocogram

[1]Authors are with Lab. for Automation and Robotics, University of Patras, Patras,Greece, Email: {georgoul, groumpos} @ee.upatras.gr
[2]Author adjuncts to Computer Science Dept., University of Ioannina, 45510, Ioannina, Greece, Email: stylios@cs.uoi.gr.

The medical device that is used for acquiring and printing out the corresponding signal is called cardiotocograph. The instantaneous FHR (beats/min) can either be obtained by Doppler ultrasound or directly from the fetal electrocardiogram via scalp electrodes. The uterine activity is measured using an external tocodynamometer or with the use of an intra-uterine pressure catheter (mmHg) [1]. All of up-to-date cardiotocographs have a port that allows the communication of the medical tool with a personal computer and the storage of the cardiotocographic data in digitized form in the hard disk of the computer.

The subtlest component of a CTG is the Fetal Heart Rate signal (FHR). During the course of labor and in case of fetal stress (distress) it is highly likely that there will appear some warning alteration in the FHR indicating that probably the fetus cannot cope with the imposed stress. On the other hand a “normal” FHR signal is in most cases an indicator of a healthy fetus.

FHR signal is produced from the interaction of both linear and non-linear mechanisms, which are not yet completely understood and described [2]. The interpretation of FHR signal is based upon the basal heart rate, the presence of accelerations and decelerations and the baseline variability. Accelerations and decelerations are defined as deviations from the baseline, while the baseline itself is in fact an imaginary line. The baseline represents the heart rate average in the absence of accelerations and decelerations [3]. However the methodology for the evaluation of the CTGs has been accused of demonstrating a low predictive value for the fetal condition and a high value of false positive alarms. In view of these facts, researchers have tried to develop new methods to circumvent the drawbacks of the standard methodology involved in the interpretation of the CTG.

In recent years many attempts have been made to employ sophisticated techniques for the classification of the CTG based primarily on the FHR signal. NN have been successfully used in many medical applications areas including CTG analysis for classification, filtering, feature extraction, and estimation [4-10]. Unfortunately there has not yet been developed a generic accepted methodology for CTG classification based on NN. What makes NN appealing to use in medical applications is their ability to handle complex and even incomplete data sets, to learn and most important their intrinsic feature to generalize. On the other hand when artificial neural networks deal with large amounts of data they face the problem known as the "curse of the dimensionality". To "cure" this problem a feature selection step is required to reduce the dimensionality of the input space with the minimum loss of information. This step is very important because the performance of the neural classifier highly depends on the right set of input features [11].

In this paper, the ICA is applied for the first time for feature extraction from FHR signal. ICA is a signal processing technique that has aroused as a solution to the commonly referred as "the blind source separation" problem [12]. It can be viewed as an extension of the well-known Principle Component Analysis (PCA) [10,13,19]. However in the ICA case the axes of the new coordinate system, which are not constrained to be orthogonal, are determined not only by the data's first and second order statistics but also by higher order statistics. ICA has been used mostly for speech separation [14] but also for array antenna processing, for biomedical records [15,16], financial market data analysis [17], for face recognition [18] and other cases involving multivariate data sets.

The paper is organized in the following way: Section 2 introduces the ICA and how the corresponding algorithm is employed for the estimation of the independent sources of any signal. Section 3 describes the steps of the proposed methodology for the classification of CTG consisting of Pre-processing stage, PCA Dimensionality reduction, implementation of ICA, and then the NN classifier. Section 4 concludes the paper.

II. Independent Component Analysis

ICA is a method that, given a set of measurements, tries to find a linear combination of components –sources- that are statistically independent. The underlying assumption in ICA framework is that there exist a number of statistically independent sources, which are linearly combined to produce the observed signals. Both the independent sources and their linear combination are unknown. In mathematic notation, we assume that we have a source vector $\mathbf{U}(m)$ composed of m independent sources U_i.

$$\mathbf{U} = [U_1, U_2, ..., U_m]^T$$

The vector $\mathbf{U}$ undergoes linear transformation, by applying a nonsingular m-by-m matrix $\mathbf{A}$, the mixing matrix. The result of the multiplication $\mathbf{AU}$ is the observation vector $\mathbf{X}(\mathrm{m})$ $(\mathbf{X} = \mathbf{AU})$

$$\mathbf{X} = [X_1, X_2, ..., X_m]^T$$

Knowing only the observation vector $\mathbf{X}(\mathrm{m})$ (both the source $\mathbf{U}$ vector and the mixing matrix $\mathbf{A}$ are unknown), the goal of ICA is to find a separating matrix $\mathbf{W}$ that after the multiplication with the observation vector will produce a vector $\boldsymbol{Y}$ from which the original source vector $\mathbf{U}$ can be recovered.

$$\mathbf{Y} = \mathbf{WX}$$

where

$$\mathbf{Y} = [Y_1, Y_2, ..., Y_m]^T$$

Note:
The source signals are ordinarily assumed to have zero mean and thus the observation vectors must also have zero mean. When this assumption is not true the mean value of the rows are subtracted from the dataset.

The aforementioned processes are depicted in the following block diagram (Figure 2)

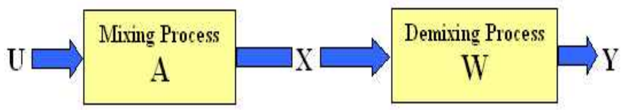

Figure 2. Block diagram of ICA

If a matrix $\mathbf{W} = \mathbf{A}^{-1}$ could be found then $\mathbf{Y=U}$ and the sources would be retrieved exactly. However what we can usually find is a separating matrix $\mathbf{W}$ whose individual rows are a rescaling and permutation of the mixing matrix $\mathbf{A}$.

To estimate the matrix $\mathbf{W}$, the unsupervised learning algorithm (Infomax algorithm), which has been proposed by Bell and Sejnowski, was implemented [12,14]. The idea behind this approach is to maximize the joint entropy of the output, which as a consequence forces the individual outputs to become statistically independent (or as statistical independent as possible).

The weight adaptation rule for $\mathbf{W}$ is:

$$\Delta\mathbf{W} = \eta\left[\mathbf{I} - f(\mathbf{U})\mathbf{U}^T\right]\mathbf{W},$$

with $f(\cdot) = \tanh(\cdot)$ and
η the learning rate parameter.

The non-linearity $f(U_i)$ is an estimate of the cumulative density function of the source estimate U_i. To what extent the density estimate must approximate the true source density is still an open issue [12]. Since we do not know the sources U_i, in the above formula we use the estimation of the sources $\mathbf{Y} = \mathbf{WX}$ (which is recalculated during each new iteration of the algorithm).

Note:
There are some prerequisites that an observation signal must fulfill in order to apply ICA and calculate the separating matrix **W**.

- No more than one source has a Gaussian distribution

→ experimentally we found that the FHR signal has super Gaussian distributions.

- The sources must indeed have no statistical dependencies

→ in our case this is an assumption that we made.

- The signals must be stationary

→ by taking quite short segments we can assume that the signal remains stationary.

III. Integrated Classification Methodology

The FHR is a very noisy, because of the method that is acquired and the FHRs that comprised the data set were irregularly sampled. Having this in mind and the requirements that a signal has to fulfill in order to apply ICA, we propose the following algorithm to modify our signal.

First of all a pre-processing stage is being employed which is divided into 4 steps:

a) Artifact removal
b) Irregular to "regular" sampling
c) 20 minute segment selection and
d) subtraction of the mean.

After the pre-processing stage the produced signal has characteristics that permit us to implement ICA. However there are some intrinsic limitations – which are explained in the following sections – that prevent us from applying ICA to the signals that come out from the pre-processing stage. Thus a stage, which performs dimensionality reduction based on PCA, precedes ICA. After this step, ICA is performed and the independent sources that are retrieved from this stage are used as basis for the feature extraction stage that follows. Finally for the classification stage a Multilayer Perceptron (MLP) NN is utilized.

The whole procedure that we propose and implement is:

- Preprocessing
- Dimensionality reduction using PCA
- ICA for the extraction of the independent sources
- Feature extraction based on the sources estimated from the previous step, and finally
- Classification using a MLP NN

and is illustrated in the following block (Figure 3) and we will briefly explain each one of these stages in the following sections.

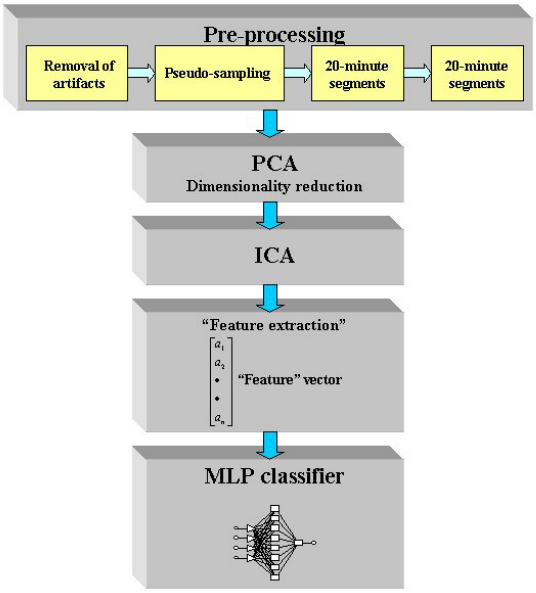

Figure 3. Experimental procedure

The proposed methodology was implemented for a data set consisting of 20 intrapartum cardiotocograms, which were collected in 1992 with a Toitu MT 810B fetal monitor, through a digital parallel port at the Dep. Obstetrics and Genecology University of Porto, Portugal. It must be mentioned that no data reduction or averaging had taken place.

During the recording time of this data set, doctors evaluated and categorized them according to their experience. Those signal were recorded during the labor**,** were evaluated without knowing the outcome of the labor and they were classified in two categories: distressed and non-distressed infants. (It must be noted that cases that had been validated as protracted descent cases were also classified as distressed).

A. Pre-processing

The fetal heart rate is a noisy signal due to the method that is used to acquire it and also do to extraneous factors that cannot be isolated. Although the missing or "spiky" data does not provoke problems to simple eye inspection, it may lead to wrong results when further digital processing is going to be implemented. In order to avoid misleading of the proposed algorithm that is used for classification, a preprocessing stage of the FHR signal is implemented. The preprocessing stage was introduced in [20] and firstly detects a stable FHR segment, which is defined as a segment where the difference (in beats/min) between five adjacent samples is less than 10 beats/min. Whenever a difference between adjacent beats higher than 25 beats/min is found, a linear interpolation will be applied between the first of those two signals and the first signal of a new stable FHR segment) (Figure 4).

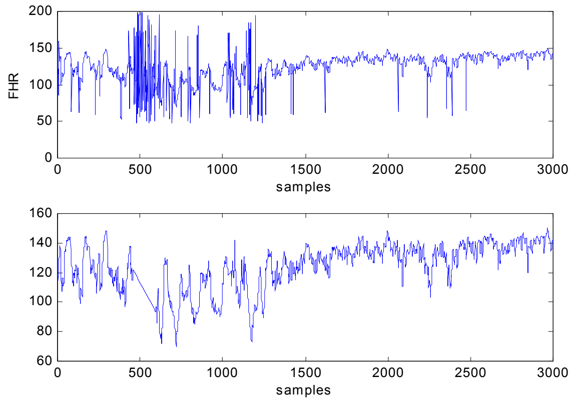

Figure 4. Data before and after the removal of artifacts

The FHR signal was an irregularly sampled signal due to the acquisition method. The irregularry sampled signal consists of samples that were recorded whenever available and not in regularly spaced time intervals. This means that whenever a new beat of the fetal heart rate was detected, the instantaneous FHR was calculated and was sent to the output of the cardiotocograph. Taking into account that FHR is calculated in beats/min and that there 2 reference time points are used, we can transform it into a pseudo-regular sampled signal exploiting the fact that no FHR signal is recorded until a new beat is detected:

$$FHR(n) = \frac{60}{t_n - t_{n-1}} \Leftrightarrow t_n = t_{n-1} + \frac{60}{FHR(n)}$$

t_n is the time that a new heart beat is detected

t_{n-1} is the time that the last beat was detected

$FHR(n)$ is the instantaneous FHR value calculated at time instant t_n

This step is employed so that there can be a direct correspondence between the length of the input vector and the elapsed time.

The next step after the two aforementioned steps is the selection of 20-minute segments that will be used for subsequent analysis. The maximum duration of most of recordings is 20 minutes and so for homogeneity reasons we performed the 20 minutes selection. . When there were very spiky segments they were excluded from subsequent processing. In order to avoid time-bias, the segments where chosen near the end of the recording, so that the selected segments are as close to delivery time as possible. The problem that we encountered was that because of the stress of baby delivery, the last minutes of the recordings were "contaminated" by artifacts. These very spiky last minutes of the recording were not included in our data set.

The final pre-processing step involves the subtraction of the average of each observation vector so as to have zero-mean observation vectors ready for further processing.

B. *Classification scheme*

For our experiments and in order to test the proposed classification methodology we had only 20 labeled signals. Due to this lack of "sufficient" number of labeled examples we decided to implement the "leave one out" method [13] (which is the extreme form of multifold cross-validation) in order to evaluate the classification performance of the proposed method.

The milestone of the proposed method is to find a feature vector that best describes the different classes of FHR signal with the minimum number of features. First of all because of the very few labeled signals, we introduced the reduction of the feature space in order to achieve better training results and increase the generalization capabilities of the classifier.

The "ICA representation" consists of the coefficients of the linear combination of the independent sources that comprise the FHR signal. As it is obvious, the dimensionality of the feature vector is determined completely by the number of the "assumed" sources ("number of sources" = "dimensionality of feature vector"). Thus, we want to reduce as much as possible the number of signals on which we implement ICA, whilst preserving at the same time as much relevant information as possible. For this reason we propose the implementation of the widely used technique of Principle Component Analysis before the ICA as it is illustrated in Figure 3.

1) *PCA-Dimensionality reduction*

The goal in dimensionality reduction is to find an invertible transformation $\mathbf{T}$ such that having a multivariate quantity $\mathbf{X}$, the truncation of $\mathbf{T(X)}$ will be optimum according to a predefined criterion. In our case we tried to find a linear transformation and the criterion was the mean-square error. In other words we sought a way to combine the m original signals to produce k new signals (k < m) representative of the original ones with the minimum loss of information. It has been proven that this can be implemented by orthogonally projecting the input data onto the subspace spanned by the eigenvectors belonging to the dominant eigenvalues of the covariance matrix of the input data (this kind of dimensionality reduction can sometimes improve the generalization capabilities of a classifier [11]). The algorithm for the linear dimensionality reduction was a batch implementation of PCA (also known as the Karhunen-Loeve transformation) [11,13].

In our experiments we used 19 signals-vectors out of the 20 signals that we had for the calculation of the principle components. Thus we had a 19x19-covariance matrix and 19 eigenvectors (19x1). By varying the number of the eigenvectors that we retained (and that corresponded to the largest eigenvalues) we also varied the number of the new signals-vectors at the output of this stage (the new vectors, which are the PC representation of the original

19 signals, are the ones to be used for the ICA in the next stage). We experimented using 3 to 6 eigenvectors in order to find an optimum number for the reduced representation. As it was mentioned earlier, the dimensionality of the feature vector equals the number of the assumed sources. Thus reducing the number of signals that are employed in the ICA scheme we reduce the number of "features" to be fed to the classifier.

2) Independent Component Analysis

Using the PCA we achieved the reduction of the dimensionality of the feature vector from m to k (where m equals 19 in our experiments and k varies from 3 to 6. It should be also mentioned that the PCA does not affect the ICA algorithm as it can only affect the second order statistic relationships of the data, leaving "untouched" the higher order relationships. Thus ICA can be performed on this new set of signals.

After the reduction of the input signals during the PCA stage, we performed ICA to the set of signals that were produced by the aforementioned stage. This way, an estimation of the assumed independent sources was produced i.e. a new set of signals $\mathbf{Y} = \mathbf{WX}$, as independent as possible.

3) Feature extraction

For each one of the original signals (both the 19 used in PCA and ICA involved and the 1 that has been excluded from these stages) we computed the coefficients that multiplied by the Independent Sources best describe that particular signal in a least square manner.
Thus, the coefficients were stored in the solution vector of the over-determined algebraic problem.

$$Coef_i \mathbf{Y} = X\_zero\_mean_i$$

Where:

$X\_zero\_mean_i$ is the input signal (after the removal of the mean) $1 \times N$ row vector (N is the number of samples 19 in our case)

$Coef_i$ is the row vector $1 \times k$ containing the coefficients that correspond to the input signal $X\_zero\_mean_i$ that will then be fed to the NN

$\mathbf{Y}$ is the $k \times N$ matrix whose rows are the estimated independent sources

The solution was obtained using the pseudoinverse method

$$Coef_i = X\_zero\_mean_i \mathbf{Y}^{\#}$$

where, $\mathbf{Y}^{\#} = \mathbf{Y}^T(\mathbf{YY}^T)^{-1}$ $(k \times k)$

4) MLP Classifier

From the set of extracted "feature" vectors we used the 19 of those, corresponding to the signals that had been employed in the ICA stage, to train a MLP to perform the classification task and we tested the classifier using the feature vector of the signal that was not used during the training process (the one that was not used during the ICA). We repeated this procedure 20 times and averaged the performance over the 20 trials.

We tried to keep the NN as simple as possible and thus we implemented architectures with only one hidden layer. We tested various numbers of hidden nodes (2-10) and various numbers of input nodes (3-6), corresponding to the number of PC retained.

For the implementation of the NN, the Neural Network Toolbox of Matlab was used, and the training algorithm employed was the Levenberg-Marquardt [9].

As we had suspected, the classifier performed better with fewer number of features. The best classification rate was achieved when 4 Principle Components were retained and the neural network had 3 hidden units. With this configuration we achieved a classification rate of 85%. The classification rates for 4 Independent sources are shown in the following diagram (Figure 5).

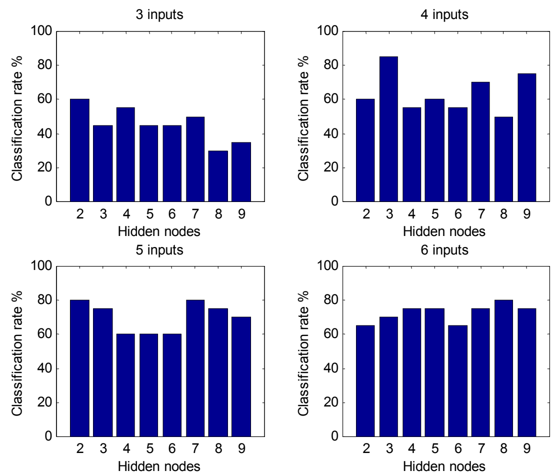

Figure 5. Classification rates for different number of inputs

IV. Conclusions

In this paper an integrated methodology for the classification of CTG has been proposed. There has been introduced the use of ICA for the estimation of independent signal sources of CTG that can be considered as an attempt to isolate the different physiological mechanisms in the fetal nervous system that are responsible for the generation of the fetal heart rate. For the dimensionality reduction there was implemented PCA and then the MLP classifier was used.

The results of the classification using MLP Neural Network are 85% correct comparing to the classification performed by experienced clinician. This result indicates that the proposed methodology may be a very powerful approach for the processing and evaluation FHR.

However, because it is well known that the doctor's classification cannot be regarded as a "golden standard", it is necessary to introduce more "objective" classifications methods [21] that will based on the apgar score and the umbilical cord's pH and other physiological characteristics.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. João Bernardes, Department of Gynecology and Obstetrics, Porto Faculty of Medicine, Porto, Portugal for providing us with the fetal heart data collected within the Research Project POSI/CPS/40153/2001, from Fundação para a Ciência e Tecnologia, Portugal.

REFERENCES

[1] M. C. Carter, Br J Obstet, "Present-day performance qualities of cardiotocographs", Gynecol 1993, Vol 100, Supplement 9, pp 10-14.

[2] Task Force of the Europ. Soc. Of Cardiol. & North Am. Soc. Of Pacing and Electrophys. Heart Rate Variability, standard of measurement, physiological interpretation and clinical use, Circulation, 1996, 93: 1043-1065

[3] R. Mantel, H.P. van Geijn, F.J.M Caron, et al., (1990), Computer analysis of antepartum fetal heart rate recordings, Int J Biomed Comput, 25, 261-272.

[4] S. Kol,, I. Thaler, N. Paz and O. Shmueli (1995). Interpretation of nonstress tests by an artificial neural network. *Am J Obstet Gynaecol* 172 pp. 1372-1379.

[5] A. Gelmetti., A. Marazzi, F. Germagnoli, P.M. Brambilla and G. Magenes (1994). Neural Network Filtering of Antepartum Fetal Heart Rate for the Baseline Determination In ICANN '94, M. Marino e P.G. Morasso Eds., Springer Verlag, pp. 1319-1322

[6] R.D.F Keith,., J. Westgate , E.C. Ifeachor and K.R. Greene (1994). Suitability of artificial neural networks for feature extraction from cardiotocogram during labour *Med & Biol Eng & Comput*. 32 pp. 51-57

[7] G. Magenes, M.G Signorini and D. Arduini (2000). Classification of cardiotocographic records by neural networks. *D.* Neural Networks, 2000. IJCNN 2000 Proceedings of the IEEE-INNS-ENNS International Joint Conference vol 3 pp. 637-641

[8] C. Ulbricht, G. Dorffner and A. Lee (1998). Neural networks for recognizing patterns in cardiotocograms. *Artificial intelligence in Medicine* 12 pp. 271-284

[9] A.A. Betanzos, O.F. Romero, B,G, Berdinas and J,C, Principe (1999). A Multi-Resolution Principal Componenet Analysis Neural Network for the Detection of Foetal Heart Rate Patterns, In Proc. 7th Europ Congres on Intelligent Techniques & Soft Computing pp. 13-16, Aachen Germany

[10] L.P. Reis, J.N. Lau, J.P.M. de Sa and J. Berrnardes (1994). Cardiotocographic Signals Parameter Estimation Using Artificial Neural Networks. RECPAD 94 pp. 177-184.

[11] C. M. Bishop 1995 Neural Networks for Pattern Recognition, Oxford University Press

[12] T.W. Lee, "Independent Component Analysis: Theory and Applications. Kluwer Academic Publishers (1998)

[13] S. Haykin, (1999). Neural Networks: A Comprehensive Foundation 2nd ed. Englewoo Cliffs, NJ: Prentice Hall

[14] A.J. Bell and T.J. Sejnowski 1995. An information maximisation approach to blind separation and blind deconvolution, *Neural Computation,* 7, 6, 1129-1159

[15] S. Makeig, A.J. Bell, T.P. Jung and T.J. Sejnowski 1995. Independent Component Analysis of Electroencephalographic Data, in Mozer M. et al (eds) *Advances in Neural Information Processing Systems 8*, MIT press

[16] A. Koutras, I. Christoyianni, E. Dermatas, G. Kokkinakis, Feature Extraction in Digital Mammography: An Independent Component Analysis Approach, Lecture Notes in Computer Science series from Springer-Verlag, June 2001

[17] A.D. Back, and A.S. Weigend, 1998. "A first application of independent component analysis to extracting structure from stock returns," International Journal of Neural Systems, vol. 8 (4), 473-484

[18] M. Bartlett, and T.J. Sejnowski, (1997).Viewpoint invariant face recognition using independent component analysis and attractor networks. Advances in Neural Information Processing Systems 9, pages 817-823, MIT Press

[19] K. Diamantaras, S.Y. Kung, Principal Component Neural Networks, John Wiley & Sons Inc

[20] J. Bernardes, C. Moura, J.P.M. de Sa, and L.P. Leita, "The Porto system for automated cardiotocographic signal analysis", J. Perinat. Med. 19 (1991) 61-65

[21] J. Bernardes, D. A. de Campos, A. C. Pereira, L.P. Leita and A. Garrido. "Objective computerized fetal heart rate analysis", Int. J. Gynecol. Obst. 62 (1998) 141-147.