Classification of Fetal Heart Rate using Grammatical Evolution

Dimitris Gavrilis
Electrical & Computer Engineering Dept.
University of Patras
Patras, Greece
gavrilis@upatras.gr

Ioannis G. Tsoulos
Computer Science Dept.
University of Ioannina
Ioannina, Greece
sheridan@cs.uoi.gr

Geroge Georgoulas
Electrical & Computer Engineering Dept.
University of Patras
Patras, Greece
georgoul@ee.upatras.gr

Euripidis Glavas
Communications, Informatic and Management Dept,
Technological Institute of Epirus
Arta, Greece
eglavas@teiep.gr

***Abstract*—There is an ongoing effort to develop advanced methods and computer-based systems to assist obstetricians in the difficult task of feature extraction and classification of the Cardiotocogram (CTG), which is the most widely used Electronic Fetal Monitoring (EFM) method worldwide. A novel method for feature construction is presented for efficient classification of CTG based on information extracted from Fetal Heart Rate (FHR) signal. The proposed method is based on grammatical evolution in order to construct new features from existing ones using nonlinear transformations. This method is tested on a data set of intrapartum cases achieving accuracy of 92.5%.**

I. INTRODUCTION

Electronic Fetal Monitoring (EFM) has been widely used for antepatrum (the period before labour) and intrapartum (the period during labour and delivery) fetal surveillance. The term EFM means the continuous recording and monitoring of Fetal Heart Rate (FHR) and Uterine Activity (UA), also known as cardiotocogram (CTG). Fig. 1 shows a typical CTG segment with the FHR at the upper part of the figure and UA at the lower part. The medical device that is used for acquiring, displaying and printing out the corresponding signals is called cardiotocograph. The instantaneous FHR (beats/min) can be obtained either 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 rapture of the membranes). The uterine activity is measured using an external tocodynamometer or with the use of an intra-uterine pressure catheter (mmHg) [1].

Before the baby delivery, the exchange of oxygen and carbon dioxide by the fetus takes place within the placenta. If a pathological change in either the maternal or fetal components of the placenta interferes with placental-fetal gas exchange, fetal asphyxia or oxygen deficiency can occur with major health problems for the baby. If prolonged and/or profound, asphyxia can lead to fetal hypoxia, a condition in which blood supply to tissues is reduced. The redistribution of blood flow to oxygen-sensitive organs, such as the brain and heart, enables the fetus to survive short periods of oxygen deficiency. However, neurological injury or fetal death may occur if fetal hypoxia is sustained for prolonged periods. Asphyxia can lead to an accumulation of carbon dioxide in the blood and tissues, resulting in fetal acidosis, a condition characterized by an increase in acidity reflected by a decrease in pH.

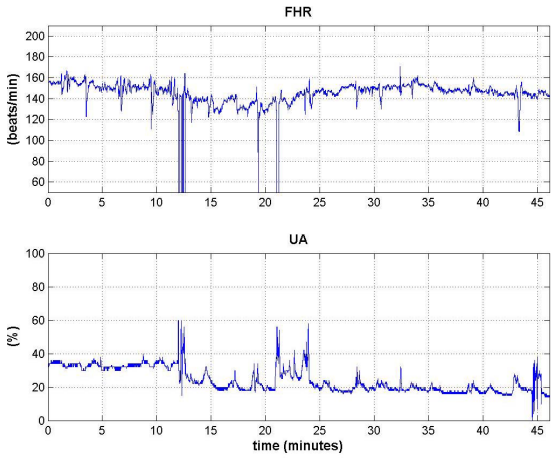

Figure 1. Typical Cardiotocogram

Therefore, during the crucial period of labor, FHR –the subtlest component of a CTG- is used as an indication of the fetus' status and, primarily, as a warning of possible fetal and neonatal compromise (metabolic acidosis) [2].

Although EFM has been used for the last 4 decades, there is still controversy regarding its efficiency. Moreover, the Dublin randomized trial has revealed an increase in operative vaginal deliveries in patients monitored with EFM during the intrapartum period [2]. Furthermore, studies of FHR reliability have shown significant inter-observer and intra-observer variation in tracing interpretation [3]. In part, the difficulty in distinguishing benign variant patterns from patterns associated with significant fetal acidemia, arose because FHR monitoring was introduced into clinical practice before the cause of FHR patterns was well understood.

The inconsistency in interpretation and increase of false positive diagnosis and, on the other hand, the technological advances in computers along with new signal processing methods, have prompted many researches to develop computer systems capable of analyzing the cardiotocogram CTG [5-19]. The employment of mathematical and algorithmic approaches has led to a reduction of inter- and intra-observer variability but there is still room for improvement regarding the identification of emergent situations which induce unacceptable stress on the fetus.

Driven by the belief that the FHR signal may convey information that it is not used by the standard interpretation, based on the guidelines given by the International Federation of Gynecology and Obstetrics (FIGO) [20], we propose a new method to discriminate fetuses that are suspicious of developing acidemia. This novel method proposes the creation of features extracted from the FHR signal in order to produce more descriptive features. The core of the proposed method is to combine a Backus Naur form (BNF) description and genetic programming to represent each constructed feature. This technique is known as grammatical evolution [21].

This paper is structured as follows; section 2 briefly presents the grammatical evolution scheme. Section 3 describes the processing steps in order to extract the original features from the FHR that will be used in the grammatical evolution stage; in section 4 the experimental results are presented and, finally, in section 5 some conclusions and ideas for future work are presented.

II. Grammatical Evolution

Grammatical evolution is a method that uses genetic programming and a BNF description to create programs in an arbitrary language. In grammatical evolution, chromosomes are a series of production rules of the appropriate BNF syntax. Each bit of the chromosome denotes a production rule from the BNF grammar. The algorithm starts from the start symbol of the grammar and gradually creates the program string by replacing non terminal symbols with the right hand of the production rule [21], [22]. The selection of the appropriate rule is performed by using the scheme:

$$\text{Rule} = B \bmod RN. \quad (1)$$

where B is the specific chromosome element and RN is the number of rules for the specific non-terminal symbol. This selection process is repeated until the end of the chromosome has been reached. After the newly constructed features have been created, a new train and test set is created from the original, according to the new features. The newly constructed feature is assigned a fitness value, which in this case is the test error of a supervised classifier. According to that fitness value, the newly constructed feature will be accepted or rejected in the successive generations.

The grammar used in the proposed method uses mathematical functions and operators as non-terminal symbols and the original features $(x_1, x_2, ..., x_n)$ and digits (0-9) as the terminal symbols. An example of this grammar can be seen below:

S ::= <expr>

<expr> ::= <expr> <op> <expr> | <func> (<expr>)

| <digit> | x_1 | x_2 | ... | x_n

<op> ::= + | - | * | /

<func> ::= sin | cos | exp | log

<digit> ::= 0 | 1 | 2 | ... | 9

This method also allows for the selection of the number of features that are to be constructed. This allows for the construction of a small number of features (e.g. 2 or 3) that can give good classification performance. By constructing a small number of features, the curse of dimensionality is alleviated. In the proposed method although the sin, cos, exp, log functions are used, any other function can be utilized. More details about the proposed method can be found in [22].

III. Preprocessing, Semgentaion and Feature Extraction

A. Preprocessing

The FHR signal is a noisy signal due to the method that is used for its acquisition and also due to extraneous factors that cannot be isolated. Although the missing or "spiky" data do not create problems to simple eye inspection, they may lead to wrong results when further digital processing is going to take place. Thus, in order to remove "spiky" segments or segments where the signal is zeroed, a pre-processing stage of the FHR signal has to take place. The pre-processing stage was introduced in [23].

B. Segmentation

In this work we used 40 FHR recordings. The recordings had various lengths, ranging from 20 minutes to more than 1 hour. Thirty of them were acquired using an HP 1350 fetal monitor and ten of them were acquired using a Toitu MT810B fetal monitor. In both cases, scalp electrodes were used for the acquisition. Due to the different durations of the recordings, we used segments of equal duration from each case and performed the subsequent analysis on these segments only. Therefore, we cropped, starting from the end of the recordings (or as close to the end as possible), segments lasting 20 minutes (maximum duration of some recordings). The segments were chosen as close to delivery time as possible so as to avoid time-bias. It must be mentioned that for some of the recordings, the very spiky 1-3 last minutes are not included in our data set.

C. Feature extraction

As in most classification problems, we extracted many features from different domains. For this particular work some of the features were extracted from the time and frequency domains and we also included "morphological" features, i.e. features that are more familiar to the doctors, such as baseline, number of accelerations, etc. [20].

1) Time domain

There are a number of methods that can evaluate variations in heart rate [24]. The features/indexes in time domain that we employed have already been used with reasonable success in the antepartum case [18] and, therefore, we decided to test them in the intrapartum case too. The features employed were the following:

Mean value of FHR signal (x_1),

Standard deviation of FHR signal (x_2),

$$\text{Delta}=\frac{\sum_{i=1}^{m}\left[\max_{i\in m}\left(FHR(i)\right)-\min_{i\in m}\left(FHR(i)\right)\right]}{m}=x_3$$

where max and min are computed within each minute of the signal and m is the number of minutes

$$STV=\frac{\sum_{i=1}^{24}\left|sFHR(i+1)-sFHR(i)\right|}{24}=x_4$$

where $sFHR(i)$ is the value of the signal $FHR(i)$ taken every 2.5 sec (i.e. once every 10 samples $sFHR(i)=FHR(10\times(i-1)+1)$), (Short Term Variability)

$$II=\frac{STV}{std\left[sFHR(i)\right]}=x_5$$, Interval Index

Long Term Irregularity (LTI) that is defined as the interquartile range $\left[\frac{1}{4},\frac{3}{4}\right]$ of the distribution $m(j)$ with

$$m(i)=\sqrt{FHR^2(i)+FHR^2(i+1)} \quad (x_6)$$

$$\text{Delta\_total}=\max_{i\in[1,N]}\left(\text{FHR(i)}\right)-\max_{i\in[1,N]}\left(\text{FHR(i)}\right)=x_7$$

2) Frequency domain

Various spectral methods have been used for the analysis of adults' heart rate [24]. However, in the case of FHR, there is no standardized use of frequency bands. For this work we divided the frequency range into 4 bands [18] and calculated the cumulative spectral power for each one of them:

Power at the range [0, 0.03) Hz (x_8)

Power at the range [0.03, 0.15) Hz (x_9)

Power at the range [0.15, 0.5) Hz (x_{10})

Power at the range [0.5, 1) Hz (x_{11})

And as a fifth feature (in frequency domain) we used the ratio

$$x_{12}=x_9/(x_{10}+x_{11})$$

3) Morphological Features

Conventional interpretation of FHR is based upon certain morphological characteristics, according to the guidelines given in [20] and [25]. In this work we used the following parameters:

Baseline (x_{13})

Number of accelerations (x_{14})

Number of Mild decelerations (x_{15})

Number of Prolonged decelerations (x_{16})

Number of Severe decelerations (x_{17})

The percentage of the time occupied by decelerations (x_{18})

From these 18 features, using grammatical evolution, we will construct an optimal set of new features for the classification of FHR.

It must be mentioned that the aforementioned features are not the only one that can be found or have been used for the classification of FHR. However, they construct a quite representative set of features from different domains in an attempt to incorporate as much knowledge as possible. The possible redundancies will be explored and, hopefully, eliminated by the subsequent stage that is the core of the proposed method and is responsible for the construction of a novel set of features.

IV. EXPERIMENTS- RESULTS

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 stratified cross-validation, an extension of regular cross-validation [26]. In *n*-fold stratified cross-validation, a data set *S* is partitioned into *n* folds such that each class is uniformly distributed among the *n* folds. Therefore we divided the 40 cases into 5 (non-overlapping) subsets each one with 4 samples from the "normal" group and 4 samples from the "risk" group.

Each time, using the grammatical evolution scheme, an expression is created and its fitness is evaluated using a Radial Basis Function (RBF) neural network [27] trained on all subsets except for one. The validation error was measured using the subset left out. We repeated this procedure 5 times, each time using a different subset for testing (Fig. 2) and we averaged the classification performances. The selection of the RBF neural network was performed on the basis of the reduced time requirements for training compared to other supervised classifiers (Multilayer Perceptrons (MLPs) trained with back-propagation, Support Vector Machines (SVMs) etc.). By doing so, we tried to compensate for the extremely time demanding procedure of the grammatical evolution.

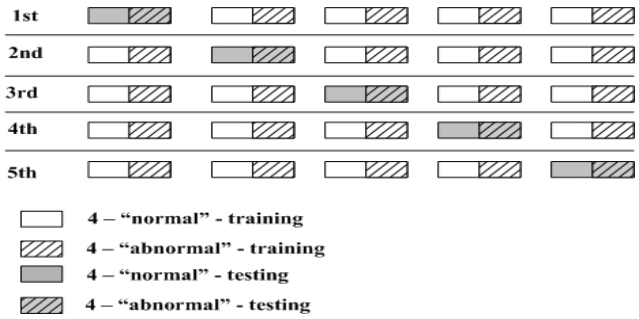

Figure 2. 5-fold cross validation

We used different topologies of RBF networks varying the number of the hidden neurons from 1 to 15. We also tested different number of constructed features (from 1 to 6). The best results were achieved using an RBF network with 9 hidden units and 5 created features. The accuracy achieved was 92.5%.

Some of the created features were the following:

$$f_1 = \sin(x_{13}) \cdot x_{17} + x_{18} \ . \quad (2)$$

$$f_2 = \sin(x_7) \cdot \exp(x_{17}) + x_{18} \ . \quad (3)$$

$$f_3 = x_{18} \ . \quad (4)$$

V. CONCLUSIONS

In this work we used for the first time, features constructed by a method employing grammatical evolution approach. The performance of the proposed method is promising but further evaluation is needed. The small amount of data prevents us from drawing safe conclusion. However, there are some interesting preliminary results that are worth mentioning.

First of all, the features generally used by the system were: $(x_2, x_5, x_7, x_{11}, x_{13}, x_{17}, x_{18})$. The other features were generally rejected, indicating that they do not give too much information to the classification process. On the other hand, in cases where a feature is by itself very useful to the classification process, the proposed system leaves it untouched (e.g. x_{18}).

Features (x_{13}, x_{17}, x_{18}) correspond respectively to number of accelerations, number of prolonged and severe decelerations. In particular, the number of severe decelerations is an indicator that the fetus is compromised and is used in everyday clinical practice by obstetricians, so it is probably not accidentally included as it is in the modified feature set. On the other hand, a trace with acceleration is regarded as a sign that the fetus is healthy [28]. Features (x_2, x_5, x_7) are different measures of the FHR variability, which is also regarded as a key factor in interpreting fetal wellbeing

Comparing the present work with our previous work involving features extracted from frequency and time domain along with Hidden Markov Models for the classification of FHR traces [29], where we had accuracy equal to 83%, the proposed method improved the performance by 9.5%.

Due to the special design of our experiment, no direct comparison can be made with other approaches found in the literature [16], [17]. However, if we attempt a comparison, we can state that our results are better than those reported in [17], where accuracy of 77% was achieved, and comparable to those reported in [16] where accuracy of 95% was achieved. In the latter case, however, an extra feature is used by the authors, which probably conveys information that is unavailable through the FHR.

In conclusion, even though the results are promising, we must be very careful before we definitely suggest the proposed methodology as the best choice for FHR classification. Therefore, even though the results indicate that certain criteria can be found to discriminate normal from acidemic outcome, something, which was questionable in the early 90s [30], the whole procedure has to be tested using more cases in order to allow "extrapolation" for any unseen case. Furthermore, at present we are also trying alternative means for classification (e.g. MLPs) something which was not done in this work due to time restrictions. In this work we assumed equal costs for both types of misclassification. In future work we will also test the performance of our method using the area under the receiver operating

characteristic (ROC) curve (AUC) as a general measure when the misclassification costs are unknown [31].

ACKNOWLEDGMENT

The authors would like to thank Prof. Joao Bernardes, Department of Gynecology and Obstetrics, Porto Faculty of Medicine, Porto, Portugal, and SisPorto Project, for providing us with the fetal heart data and with critical comments, in the context of the Research Project POSI/CPS/40153/2001, from Fundacao para a Ciencia e Tecnologia, Portugal.

REFERENCES

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

[2] H. P. V. Geijn, "Developments in CTG analysis," *Bailliers Clin. Obstet. Gynaecol.,* vol. 10, no 2, pp. 185-209, 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, pp. 524-39, 1985.

[4] J. Bernardes, A. C. Pereira D. A. de Campos, H. P. Van Geijn, and L.P. Leite, "Evaluation of interobserver agreement of cardiotocograms," *Int J Gynecol Obst.,* vol. 57**,** no. 1, pp.33-37, 1997.

[5] D. A. de Campos, J. Bernardes, A. Garrido, S. Marques, and L.P. Leite, "SisPorto 2.0 – a program for automated analysis of cardiotocograms," *J Matern Fetal Med* vol 9, pp. 311-318, 2000.

[6] 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. Maternal Fetal Invest.*, vol. 3, pp. 159-163, 1993.

[7] 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. 2, pp. 261-272, 1990.

[8] Mantel, R., 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: 2. Detection of accelerations and decelerations," *Int. J. Biomed. Comput*., vol. 25, no. 2, pp. 273-286, 1990.

[9] 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, 2000.

[10] 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, pp. 638-639, 1993.

[11] K. Maeda, "Computerized analysis of cardiotocograms and fetal movements," *Bailliers Clin. Obstet. Gynaecol.,* vol. 4, no. 4, pp. 1797-813, 1990.

[12] B. G. Berdinas, A. A. Betanzos and O. F. Romero, "Intelligent analysis and pattern recognition in cardiotocographic signals using a tightly coupled hybrid system," *Artificial Intelligence*, vol. 136, pp. 1-27, 2002.

[13] G. S. Dawes, M. Moulden, and C. W. Redman, "Computerized analysis of antepartum fetal heart rate," *Amer. J. Obstet. Gynecol.,* vol. 173, no. 4, pp. 1353-1354, 1995.

[14] J. F. Skinner, J. M. Garibaldi and E. C. Ifeachor, "A Fuzzy System for Fetal Heart Rate Assessment," I*n Proc of the 6th Fuzzy Days Conference*, Dortmund, Germany, pp. 20-29, 1999

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

[16] E. Salamalekis, P. Thomopoulos, D. Giannaris, I. Salloum, G. Vasios, 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. Gynaeco.,* vol. 109, no. 10, pp. 1137-1142, 2002.

[17] 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, 1995.

[18] 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. on Biomed. Engineering,* vol. 50, no 3. pp. 365-374, March 2003.

[19] B. Guijarro- Berdiñas, A. Alonso-Betanzos, and O. Fontenla-Romero, "Intelligent analysis and pattern recognition in cardiotocographic signals using a tightly coupled hybrid system," *Artificial Intelligence*, vol. 136, pp. 1-27, 2002.

[20] G. Rooth, A. Huch and H. Hough, "Guidelines for the use of fetal monitoring," *Int. J. Gynecol. Obstet,.*vol. 25, pp. 159-67, 1987.

[21] M. O'Neill and C. Ryan, Grammatical Evolution, IEEE Trans. Evolutionary Computation, Vol. 5, pp. 349-358, 2001.

[22] I. G. Tsoulos, D. Gavrilis, and E. Dermatas, Features Selection and Construction using Grammatical Evolution, submitted to IEEE Transactions on Evolutionary Computation.

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

[24] Heart rate variability. Standards of measurements, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Eur. Heart J.,* vol. 17 no. 3, pp. 354-381, 1996.

[25] 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, pp. 1385-90, 1997.

[26] L. Breiman, J. H. Friedman, R. A. Olsen and C. J. Olsen, *Classification and Regression Trees*, Wadsworth International Group, 1984.

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

[28] J. T. Parer, *Handbook of fetal heart rate monitoring*, Philadelphia, Pennsylvania: W. B. Saunders Company, 1997.

[29] G. Georgoulas, G. Nokas, C. Stylios, and P. P. Groumpos "Classification of Fetal Heart Rate during labour using Hidden Markov Models," in Proc*.of IJCNN 2004 International Joint Conference on Neural Networks & FUZZ IEEE 2004 IEEE International Conference on Fuzzy Systems pp* 2471-2476*,* Budapest, Hungary, July 25-29, 2004.

[30] G. S. Dawes, "Computerized fetal heart rate analysis" in A critical appraisal of fetal surveillance eds. E. P. van Geijn and F. J. A, Copray: Elsevier Science, 1994.

[31] A.P. Bradley. *The use of the area under the ROC curve in the evaluation of machine learning algorithms*. Pattern Recognition, vol. 30, no 7, pp. 1145-1159, 1997.