Least Squares Support Vector Machines for FHR Classification and Assessing the pH Based Categorization

C.D. Stylios[1], G. Georgoulas[1], P. Karvelis[1], J. Spilka[2], V. Chudáček[2], and L. Lhotska[2]

[1] Laboratory of Knowledge and Intelligent Computing, Dept. of Computer Engineering,
Technological Educational Institute of Epirus, Arta, Greece
[2] CIIRC, Czech Technical University in Prague, Czech Republic

Abstract— **Cardiotocography (CTG) is the major monitoring tool for fetal well-being surveillance during labor. It consists of two distinctive signals: the Fetal Heart Rate (FHR) and the Uterine Contractions signal. The CTG interpretation is classically performed by obstetricians with visual inspection for reassuring or ominous patterns, which are associated with fetus' condition. Deviations of the CTG and especially of the (FHR) from normality can be an indication of oxygen deprivation during the stressful labor process, which can lead to major neurological damage to the fetus or even death. This compromise is usually reflected at the pH level of newborn's blood. Therefore pH levels are usually used for the discrimination between healthy and compromised fetuses. In this work we present our preliminary results of the application of a machine learning approach, using least squares support vector machines, to FHR classification using the largest CTG open-access database so far.**

Keywords— **Fetal Heart Rate (FHR), Least squares support vector machines (LS-SVMs), Classification.**

I. INTRODUCTION

Cardiotocography (CTG) refers to the simultaneous acquisition of the fetal heart rate (FHR) and uterine contractions and their display on a single paper strip, cf. Figure 1. Experienced obstetricians read the CTG traces and assess fetal wellbeing following the guidelines of the International Federation of Gynecology and Obstetrics (FIGO) [1] or similar guidelines provided by other national boards [2]. However even though these specific guidelines are quite clear it has been reported that their use in clinical practice is a challenging endeavor. In fact CTG interpretation is accompanied by persistent high inter and intra-observer variability [3], [4] and also blamed for an increase of the number of caesarean sections. The CTG evaluation and following clinical decisions still remain subjective and are scrutinized in malpractice lawsuits [5]. Therefore in order to reduce the variability in CTG interpretation and also minimize the subjectivity of the assessment a number of computerized systems have been proposed based primarily on the analysis of the FHR signal that is the most informative component of the CTG.

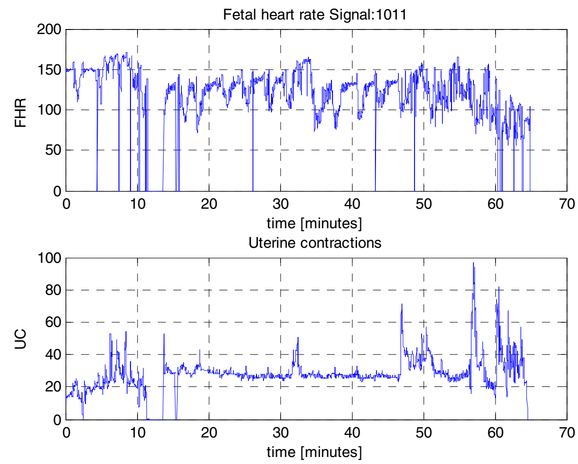

Fig. 1 FHR and Uterine Contractions (UC) signal

These computerized systems are usually built around a pattern recognition pipeline: after a stage of signal prerocessing/denoising, features are engineered and extracted from the FHR signal with some or all of them used to train a classifier that learns to discriminate different labor outcomes.

For the feature engineering/extraction stage a wealth of methods/attributes have been proposed, ranging from features that capture the essence of the FIGO guidelines (baseline level, number of accelerations, number of deceleration, decelerations depth, shape, and frequency etc.) [6],to features coming from the time domain [7], [8], the frequency domain [9], the time-frequency domain [10], the domain of nonlinear dynamics [11], [12], [13] or even features created by the application of genetic programming methods to basic features [14].

For the classification/decision stage also a plethora of methods have been proposed such as Support Vector Machines (SVMs) [9], [10], [12], [15], Artificial Neural Networks (ANNs) [16], [17], Hidden Markov Models (HMMs) [18], fuzzy systems [19], and other approaches such as ordinal classification approaches [20] and one class classifiers [21].

All the aforementioned classifiers share in common the need of a labeled training set where the labels correspond to

E. Kyriacou et al. (eds.), *XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016*,
IFMBE Proceedings 57,
DOI: 10.1007/978-3-319-32703-7_233

the classes that the newborn was assigned after the delivery. This assignment can be performed using various criteria [22]. Among them, probably the most straightforward is the one that is based on the pH value coming from the umbilical artery cord blood analysis. In that case, a threshold is used that discriminates between healthy fetuses and fetuses that had suffered somewhat from the excess stress during labor. The most commonly accepted threshold is that of a pH value equal to 7.05 [23].

In this work we investigate a method to classify FHR, using a set of features extracted from different domains combined with a variant of SVMs method and tested using the largest open-access CTG database [24], achieving competitive results compared to other similar approaches.

Section 2 describes briefly the involved methods. Section 3 presents the results and Section 4 concludes the paper offering also directions for possible future work.

II. Fetal Heart Rate Classification

A. Preprocessing

Labor is a very stressful period and this also reflects upon the recordings of the FHR, especially when recorded using ultrasound devices that is typically contaminated by spiky artifacts and also includes periods where the signal drops to zero due to mother movement and electrode displacement. Therefore, before any application of feature extraction algorithm, some kind of mitigation is needed. Even though more elaborate approaches have been proposed for noise removal from FHR recordings [25], in this work a more conventional and straightforward approach is adopted that relies on Hermite spline interpolation to fill short missing gaps (shorter than 15 seconds) of the FHR recordings. The results of the application of the preprocessing stage for a single FHR recording are depicted in Figure 2.

B. Feature Extraction

Feature engineering is one of the most crucial stages in the pattern classification process [26] with the FHR classification not being an exception. As it was mentioned in the introduction a variety of features have been proposed and tested for FHR classification. In this work trying to capture as much information as possible a feature set from different domains is compiled. In total 54 features coming from the time domain (measures of variability), frequency domain (energy at different frequency bands [7], [27]), non-linear domain (Approximate Entropy [28], Sample Entropy [29], Lempel Ziv Complexity [30], Fractal Dimension [31], Detrend Fluctuations Analysis [32] etc.) and morphological features trying to quantify the FIGO rational (number of accelerations, number of decelerations, baseline value etc.). A detailed description of the whole set can be found in our previous publications [10], [12], [20].

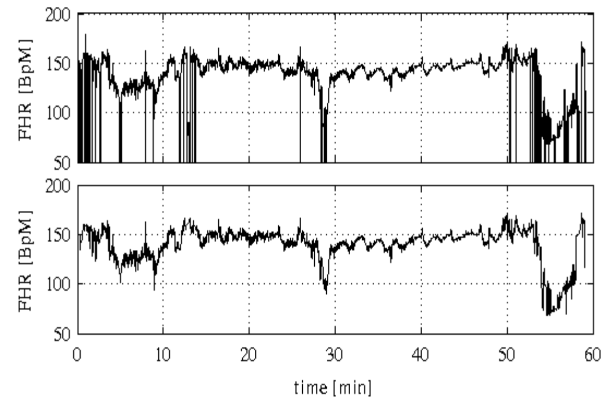

Fig. 2 FHR signal preprocessing. Top: the original FHR signal with missing data (unnatural zeroing). Bottom: denoised signal

C. Feature Selection

Feature extraction is sometimes more of an art than science [33]. Therefore usually one extracts more features than actually needed or even features that might be irrelevant for the problem at hand. This however can harm the classification performance of the system. Thus a feature selection stage usually follows feature extraction. The feature selection stage decreases the dimensionality of the input stage which apart from the apparent computational benefits often increases the generalization performance of the system [33].

Feature selection approaches can be divided into three basic categories: filters, wrappers, and embedded methods [34]. Wrappers require the use of a classification algorithm while filters do not. Embedded methods integrate the feature selection process with the building of the classifier.

In this work filter selection is used, which is very fast as it does not require the training of a classifier at each step. Instead it sorts the features based on some intrinsic characteristic

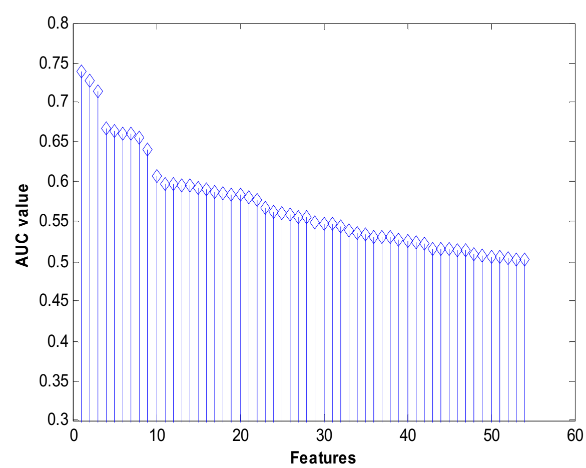

Fig. 3 The AUC values for all 54 features in a decreasing order.

(i.e. p-values, the Area Under the Receiver Operating Characteristic Curve (AUC) etc.). Due to the imbalance nature of the problem (see Section 3.A) the AUC is selected because it is not affected by class imbalance.

The initial plot of features' AUC values for a number of random samples of the data set, clearly showed, cf. Figure 3, that there were three distinct groups of features: three features provide the best AUC value and stand out from the rest, followed by other six features with an AUC value a bit higher than the rest 45 features. This observation prompted for direct analysis based on the three different sets of features: a) the top three features, b) the top nine features (top three and the next six features) and c) all 54 features.

D. Classification Using Least Squares Support Vector Machines (LS-SVM)

In this work the final classification stage, which assigns a FHR to one of the two predefined classes, is performed using the least-squares version of SVMs [36]. SMVs are among the state of the art methods for binary classification problems. Moreover SVMs are partly insensitive to the presence of correlated inputs [37], an issue which it not taken into consideration by the filtering process of the previous stage. On top of that, the LSSVM paradigm trains much quicker for moderate size problems compared to the conventional SVMs. Apart from the possible correlations between the inputs, the LS-SVM also needs to cope with the imbalanced nature of the problem since the pathological class accounts for around 8% of the total number of cases. Unequal penalty costs are selected to provide a balanced performance between the two classes [39]. The popular RBF kernel is used in this work to provide nonlinear capabilities to the classifier.

III. RESULTS

A. Data

The open-access CTG database is used in this work, consisting of 552 records, which is a subset of 9164 intrapartum CTG recordings [24] acquired between the years 2009 and 2012 at the obstetrics ward of the University Hospital in Brno, Czech Republic. The last 30 minutes of the 1st stage of labor are selected for subsequent processing. From the 552 recordings, 44 of them have pH value lower or equal to 7.05 which is selected as the limit to normality. Obviously, this creates a highly skewed class distribution.

B. Evaluation of the Classification Performance

For the classification performance evaluation a 44-fold stratified cross validation procedure is involved (at each fold one of the abnormal and 11 or 12 from the normal cases were left out for performance evaluation) repeated 15 times. During each fold an tuning process also takes place which does not involve in any way the cases left out for estimating the performance of the algorithm. This way the tuning process is completely decoupled from the performance estimation process [40]. As it was noted in the previous section there is high imbalance between normal and pathological cases. To take that into further consideration, during the tuning process (selection of hyper-parameters σ, *C*, and the imbalance factor of the LSSVM [41] formulation) we avoid the use of the traditional total accuracy as a measure of performance but we use the Matthews Correlation Coefficient (MCC) which is immune to class imbalance [42].

The results of the aforementioned procedures for the three different feature sets are presented in Table 1 and graphically illustrated in Figure 4. As it can be seen the best performance is achieved using only the three individually best features, for more features included the classification performance degrades. Figure 4 also shows that the inclusion of more features affects negatively the sensitivity of the method. However no clear pattern can be inferred for the specificity.

Table 1 Sensitivity and specificity measures based on tests with different number of features and MCC optimization criteria

#Features	Sensitivity	Specificity
3	**0.685**	**0.777**
9	0.672	0.724
54	0.558	0.769

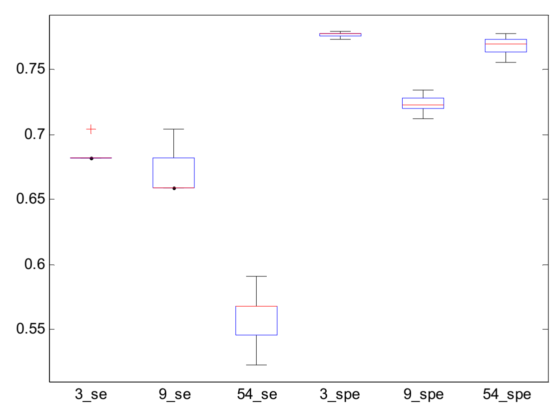

Fig. 4 Sensitivity and specificity for different feature sets. The x axis displays the number of features chosen together with a metric (SE: sensitivity, SPE: specificity).

IV. DISCUSSION-CONCLUSIONS

In this work, an integrated approach for FHR classification is presented and tested using the largest available open-access CTG database. The achieved classification performance is better compared to our previous work and is in full agreement with the literature. Our results further suggest that a sensitivity/specificity ratio of 70/70 seems to be reasonable for large dataset as it was summarized in [22] and further reported in [43]. This observation brings forward the question of whether pH labeling and FHR based features can reach higher performance values. Since pH is still predominantly used in clinical practice to determine adverse labor outcomes and will not be likely abandoned in a future, perhaps its use in combination with other criteria along with other clinical information would be beneficial [22]. Latent Class Analysis (LCA) model for aggregating information coming from different sources has given promising results towards that direction.

ACKNOWLEDGMENT

This work was partially supported by the research project "Intelligent System for Automatic CardioTocoGraphic Data Analysis and Evaluation using State of the Art Computational Intelligence Techniques" within the "Greece-Czech Joint Research and Technology projects 2011-2013" program of the General Secretariat for Research & Technology, Greek Ministry of Education and Religious Affairs, cofinanced by Greece, National Strategic Reference Framework (NSRF) and by Czech Grant Agency project number: 14-28462P Statistical methods of intrapartum CTG signal processing in the context of clinical information.

REFERENCES

1. FIGO (1986) Guidelines for the Use of Fetal Monitoring, Int. J. Gynaecol. Obstet. 25:159–167.
2. ACOG (2009) American College of Obstetricians and Gynecologists Practice Bulletin No. 106: Intrapartum fetal heart rate monitoring: nomenclature, interpretation, and general management principles. Obstet. Gynecol. 114(1):192-202.
3. Bernardes J, Costa-Pereira A, Ayres-de-Campos D, et al. (1997) Evaluation of interobserver agreement of cardiotocograms. Int. J. Gynecol. Obset. 57(1):33–37.
4. Hruban L, Spilka J, Chudáček V, et al. (2015) Agreement on intrapartum cardiotocogram recordings between expert obstetricians. Journal of Evaluation in Clinical Practice 21(4): 694-702.
5. Sartwelle TP, Johnston JC (2014) Cerebral palsy litigation: change course or abandon ship. J Child Neurol. 30(7):828-41. DOI 10.1177/0883073814543306.
6. Ayres-de-Campos D, Bernardes J, Garrido A, et al. (2000) SisPorto 2.0: a program for automated analysis of cardiotocograms. J. Matern. Fetal Med. 9:311–318.
7. Task-Force (1996) Heart rate variability. Standards of measurement, 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. 17(3):354–381.
8. Magenes G, Signorini MG., Arduini D (2000) Classification of cardiotocographic records by neural networks, IJCNN Proc. IEEE-INNS-ENNS, International Joint Conference on Neural Networks, IJCNN 2000, vol 3 pp. 637-641.
9. Georgoulas G, Stylios C, Groumpos P (2006) Predicting the risk of metabolic acidosis for newborns based on fetal heart rate signal classification using support vector machines. IEEE Trans Biomed Eng, 53(5):875-884.
10. Georgoulas G, Stylios C, Groumpos P (2006) Feature extraction and classification of fetal heart rate using wavelet analysis and support vector machines. International Journal on Artificial Intelligence Tools, 15(03):411-432.
11. Goncalves H., Rocha AP, Campos DA, et al. (2006) Linear and non linear fetal heart rate analysis of normal and acidemic fetuses in the minutes preceding delivery. Med. Biol. Eng. Comput. 44(10):847-855.
12. Spilka J, Chudáček V, Koucký M, et al. (2012) Using nonlinear features for fetal heart rate classification. Biomedical Signal Processing and Control. 7(4):350–357.
13. Chudáček V, Anden J, Mallat S, et al. (2014) Scattering Transform for Intrapartum Fetal Heart Rate Variability Fractal Analysis: A Case-Control Study. IEEE Trans Biomed Eng. 61:1100-1108.
14. Georgoulas G, Gavrilis D, Tsoulos IG, et al. (2007). Novel approach for fetal heart rate classification introducing grammatical evolution. Biomedical Signal Processing and Control 2(2), 69-79.
15. Warrick PA, Hamilton EF, Precup D, et al. (2010) Classification of normal and hypoxic fetuses from systems modeling of intrapartum cardiotocography. IEEE Trans Biomed Eng 57(4):771-779.
16. Czabanski R, Jezewski J, Wrobel J, et al. (2010) Predicting the risk of low-fetal birth weight from cardiotocographic signals using ANBLIR system with deterministic annealing and-insensitive learning. IEEE Transactions on Information Technology in Biomedicine 14(4):1062-1074.
17. Georgieva A, Payne SJ, Moulden M, et al. (2013) Artificial neural networks applied to fetal monitoring in labour. Neural Computing and Applications 22(1):85-93.
18. Georgoulas G, Stylios CD, Nokas G, et al (2004) Classification of fetal heart rate during labour using hidden Markov models. Proc. IEEE International Joint Conference on Neural Networks. vol 3, pp. 2471-2475.
19. Czabański R, Jezewski J, Wrobel J, et al. (2013) Application of fuzzy inference systems for classification of fetal heart rate tracings in relation to neonatal outcome. Ginekologia polska 84(1):38-43.
20. Karvelis P, Spilka J, Georgoulas G, et al. (2015) Combining latent class analysis labeling with multiclass approach for fetal heart rate categorization. Physiological measurement 36(5):1001 DOI http://dx.doi.org/10.1088/0967-3334/36/5/1001
21. Gavrilis D, Nikolakopoulos G, Georgoulas G (2015) A one-class approach to cardiotocogram assessment. EMC Proc. 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 518-521.
22. Spilka J (2013) Complex approach to fetal heart rate analysis: A hierarchical classification model, PhD Thesis, at Czech Technical University in Prague Department of Cybernetics.
23. Amer-Wåhlin I, Maršál K, (2011) ST analysis of fetal electrocardiography in labor. In Seminars in Fetal and Neonatal Medicine 16(1):29-35.
24. Chudáček V, Spilka J, Burša M, et al. (2014) Open access intrapartum CTG database, BMC Pregnancy and Childbirth, 14:16 DOI 10.1186/1471-2393-14-16.
25. Oikonomou VP, Spilka J, Stylios CD et al. (2013) An adaptive method for the recovery of missing samples from FHR time series. CBMS Proc. pp. 337-342.
26. Domingos P. (2012). A few useful things to know about machine learning. Communications of the ACM, 55(10):78-87.
27. Signorini MG, Magenes G, Cerutti S, et al. (2003) Linear and nonlinear parameters for the analysis of fetal heart rate signal from cardiotocographic recordings. IEEE Trans Biomed Eng 50(3):365–374.
28. Pincus S (1995) Approximate entropy (ApEn) as a complexity measure. Chaos 5(1):110–117.
29. Richman JS, Moorman JR (2000) Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol 278(6):2039–2049.

30. Lempel A, Ziv J (1976) On the complexity of finite sequences. IEEE Transactions on Information Theory 22(1):75–81.
31. Kinsner W (1994) Batch and real-time computation of a fractal dimension based on variance of a time series. Technical report, Department of Electrical & Computer Engineering, University of Manitoba, Winnipeg, Canada.
32. Peng C.K., Havlin, S. Stanley, H.E. et al. (1995) Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. Chaos 5 (1):82–87.
33. De Sa JM (2001) Pattern recognition: concepts, methods, and applications. Springer.
34. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. Journal of Machine Learning Research 3:1157–1182.
35. Wasikowski M, Xue-wen C (2010) Combating the small sample class imbalance problem using feature selection. IEEE Transactions on Knowledge and Data Engineering 22(10):1388-1400.
36. Suykens J, Vandewalle J (1999) Least Squares Support Vector Machine Classifiers, Neural Processing Letters. 9:293-300.
37. Joachims T (1998) Text categorization with support vector machines: Learning with many relevant features. ECML-98, Lecture Notes in Computer Science, 1398:137-142
38. Luts J, Ojeda F, Van de Plas R, et al. (2010) A tutorial on support vector machine-based methods for classification problems in chemometrics. Analytica Chimica Acta 665(2):129-145.
39. Salzberg S L (1997) On comparing classifiers: Pitfalls to avoid and a recommended approach. Data mining and knowledge discovery 1(3):317-328.
40. LS-SVMlab toolbox at http://www.esat.kuleuven.be/sista/lssvmlab/
41. Matthews BW (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et Biophysica Acta (BBA) - Protein Structure 405(2):442–451.
42. Spilka J, Frecon J, Leonarduzzi R, et al. (2016) Sparse Support Vector Machine for Intrapartum Fetal Heart Rate Classification. IEEE J Biomed Health Inform Submitted

Corresponding author:

Chrysostomos D. Stylios
Technological Educational Institute of Epirus
47100, Arta
Greece
stylios@teiep.gr