Comparison of Feature selection for Monopolar and Bipolar EHG signal

Dima Alamadine, Mohamad Khalil, Catherine Marque

Comparison of Feature selection for Monopolar and Bipolar EHG signal

Dima ALAMEDINE[1,2*], Mohamad KHALIL[2], Catherine MARQUE[1]

[1] Université de Technologie de Compiègne, CNRS UMR 7338, Biomechanics and Bio-Engineering, Compiègne, France.
[2] Lebanese University, Azm platform for Research in Biotechnology and its Applications, LASTRE laboratory, Tripoli, Lebanon.
[*]dima.alamedine@gmail.com

Abstract — Different categories of features have been extracted from the uterine Electrohysterography Signal (EHG), in order to find specific useful information that can be used to classify between labor and pregnancy contractions. The selection of pertinent features is very important to solve classification problem. In our study we selected 26 features (linear, nonlinear and features related to the EHG propagation) extracted from the literature. In this paper we present two feature selection methods. The first one is a filter method named F-score. The second one is a wrapper method named Genetic algorithm. The aim of this paper is to compare the results of feature selection methods and classification obtained by using monopolar and then bipolar EHG signals.

Index terms - Biomedical sensors, feature selection, Signal Processing.

I. INTRODUCTION

The World Health Organisation (WHO) posted that the perinatal mortality rate is around 7 per 1,000 births in the most developed country [1]. This prematurity remains a major problem in obstetrics. The medical, physiological and socioeconomic consequences of Preterm birth are important. A promising method of preventing preterm birth started since the 80s based on the study of uterine electromyographic signal (electrohysterogram, EHG). This signal represents the electrical activity triggering the mechanical contraction of the myometrium. A large number of features have been extracted from EHG in order to find specific information leading to the detection of preterm birth, but based on very different populations and protocols. Therefore, we need a method to select the most significant subset, among features extracted from the bibliography, in order to discriminate pregnancy and labor contractions, tested on a same signal database. After an extensive literature review which allowed us to define a set of features (linear, nonlinear and features related to the EHG propagation), two feature selection methods are applied in order to select the most pertinent features. The first one is a filter method named F-score. The second one is a wrapper method named Genetic algorithm. In this study, we present the results of feature selection by using bipolar EHG signals on the one hand and monopolar EHG signals on the other hand. Additionally, a validation part was developed to test the performance of the selected subsets obtained by using either bipolar or monopolar EHG signals to test their diagnostic power.

II. MATERIALS AND METHODS

II.1. **Experimental protocol**

We collected EHG signals from different populations of pregnant women (normal pregnancies or at risk), by placing an array of 16 electrodes attached to the woman's abdomen. The bursts of uterine electrical activity that correspond to contractions were manually segmented, based on the tocodynamometer signal recorded simultaneously. After this manual segmentation of EHG bursts, we obtained a database containing 290 Monopolar pregnancy contractions and 189 monopolar labor contractions. These monopolar burst were denoised by using the powerful EHG filtering methods developed by [2]. Then, from these denoised monopolar signals we calculated the vertical bipolar EHG signals (vbi), for the 290 pregnancy and 189 labor contractions.

II.2. **Feature extraction**

26 features from different categories have been extracted from our EHGs. The linear features are: mean frequency (MPF), Peak Frequency (PF), deciles (D1…D9), parameters extracted from wavelet decomposition (W1…W5). The nonlinear features are: Time reversibility (Tr), Lyapunov exponent (LE), Sample Entropy (SE), Detrended fluctuation analysis (DFA) and Variance entropy (VarEn). The features related to the EHG propagation are: linear correlation coefficient (R2), nonlinear correlation Coefficient (H2), phase synchronization (y), nonlinear correlation coefficient with time varying (H2_TV) and general synchronization with time varying (H_TV) [3].

II.3. **Feature selection Methods**

To select the pertinent features, two feature selection methods were applied in our study. The first one is a filter method named F-score; this method calculates the discriminative ability of each feature [4]. Feature with higher F-score value are selected (value of individual Fscore > mean value of all Fscores) [4]. The second one is a wrapper method named Genetic algorithm (GA) [5]. This method evaluates a subset of features by its classification performance, using a learning algorithm.

We used here the classical K-nearest neighbors classifier (KNN). We then evaluated the performances of these selected subsets by calculating the percentages of correct classification obtained when the subset is used as input of a classifier (Validation part).

III. RESULTS

III.1. Feature selection Part

Table 1 presents the results obtained after applying F-score and GA-KNN-KFOLD (classifier KNN, data split used in learning algorithm is KFOLD validation with K=10) to our bipolar EHG signals on the one hand and monopolar EHG signals on the other hand. We used, for this learning step, a part of the database explained in section II.1 (139 labor contractions and 240 pregnancy contractions). Each feature subset selected by GA-KNN-KFOLD, corresponds to the one giving the maximum percentage of correct classification (88.12% for the subset obtained by using bipolar contractions and 88.14% for monopolar contractions).

Selection method	**Selected feature subset (Bipolar contractions)**	**Selected feature subset (Monopolar contractions)**
F-score	[DFA, VarEn, W2, MPF, H2_TV, H_TV]	[SE, DFA, VarEn, W2, H2_TV, H_TV]
GA-KNN-KFOLD	[Tr, SE, DFA, W1, W4, D1, D3, D4, D5, D8, MPF, R2, H2, y]	[DFA, W1, W2, W3, W4, D5, D6, MPF, R2, H_TV]

Table 1: Results of F-score and GA on bipolar and monopolar contractions.

III.2. Validation Part

A validation part was applied on the remaining contractions not used during the feature selection step (50 labor contractions and 50 pregnancy contractions), in order to evaluate the performances of the selected subsets obtained in section III.1. The results of evaluation, which are presented in table 2, are based on the calculation of mean ± standard deviation of the percentages of correct classification of 500 repetitions. We used there the same classifiers and data split as used in the selection step.

IV. DISCUSSION – CONCLUSION

We notice from table 1 that the filter method F-score gives smaller sets than the wrapper method GA. From the validation part presented above (table2), we see that the features selected by using both F-score and GA methods on bipolar EHG correspond to the higher percentage of correct classification (68.39 ± 2.16, 85.06 ± 1.63), compared to those obtained in monopolar contractions. Additionally, we notice that the features selected by using GA on bipolar contractions correspond to the higher percentage (85.06 ± 1.63). We conclude that the wrapper method GA gives better results when using bipolar signals. Only one feature is always selected, whatever the method (DFA). This nonlinear feature is thus confirmed to be of great interest for pregnancy/labor comparison [3]. Some other features, selected 3 over 4 times are also of possible great interest for diagnosis.

As perspective, we will test other methods of feature selection in the same bipolar and monopolar signal database and compare them to the results obtained in this paper. We will also apply a method of dimension reduction in order to compare the results with those obtain by using these feature selection methods.

Selection method	**Mean ± STD of the percentage of correct classification (Bipolar)**	**Mean ± STD of the percentage of correct classification (Monopolar)**
F-score-KNN-KFOLD	68.39 ± 2.16	61.29 ± 1.97
GA-KNN-KFOLD	85.06 ± 1.63	66.69 ± 1.99

Table 2: Mean ± STD of the percentage of correct classification using KNN of 500 repetitions for the selected features subset.

ACKNOWLEDGMENTS

We are indebted to French Regional Council of Picardy and FEDER funds for funding this work.

REFERENCES

[1] Neonatal and perinatal mortality : country, regional and global estimates; Report WHO 2006, http://whqlibdoc.who.int/publications/2006/9241563206_eng.pdf.

[2] M. Hassan, S. Boudaoud, J. Terrien, B. Karlsson, and C. Marque, "Combination of Canonical Correlation Analysis and Empirical Mode Decomposition applied to denoise the labor electrohysterogram," *IEEE Trans. Biomed. Eng.*, Sep. 2011, Vol. 85, no. 9, pp. 2441-2447.

[3] A.DIAB, "Study of The Nonlinear Properties And Propagation Characteristics Of The Uterine Electrical Activity During Pregnancy And Labor", Ph.D. dissertation, Thèse de l'université de Technologie de Compiègne, 2014.

[4] S. Ding, "Feature Selection Based F-Score and ACO Algorithm in Support Vector Machine,"*in Second International Symposium on Knowledge Acquisition and Modeling, 2009. KAM* '09, 2009,Vol. 1, pp.19-23.

[5] B. Oluleye, L. Armstrong, L. Jinsong, and D. Diepeveen, "Zernike moments and genetic algorithm: Tutorial and application," *Br. J. Math. Comput. Sci.*, 2014.Vol. 4, no. 15, pp. 2217–2236.