# *Preterm Birth Prediction from EHG Signals using Linear, Non-linear and Statistical Features*

Dr. Shahzad Younis
SEECS NUST
Islamabad, Pakistan
Muhammad.shahzad@seecs.edu.pk

Muhammad Zohaib Arshad
SEECS NUST
Islamabad, Pakistan
13beemarshad@seecs.edu.pk

Hamza Abid
SEECS NUST
Islamabad, Pakistan
13beehabid@seecs.edu.pk

*Abstract*— **Premature Birth prediction is essential to prevent infant morbidity and death. Analysis and classification of EHG signals is an important tool for this purpose. In this paper, linear and non-linear features including FFT have been extracted from EHG signals and have been applied to an SVM classifier. The data is obtained from an open source platform consisting of a total of 300 records for Term and Preterm Birth. The obtained results indicate that non-linear features show better results than linear ones and FFT is an important tool for successful prediction of premature birth.**

***Keywords—EHG; Labor; Preterm Birth; Pregnancy; SVM***

## I. Introduction

Premature Birth is the delivery of babies that are born alive, before the 37 weeks of gestation. Premature birth is one of the most important factor contributing to infant indisposition and mortality. According to recent statistics about 7% of total babies born are premature [1] [2] and around 50% of all infant deaths are caused by preterm delivery of babies [3]. In time prediction and treatment of Premature Birth can save many infant lives and with proper treatment the consequences of premature birth like impairments to hearing, vision and non-communicable diseases can be avoided and treated properly. Around 40% of the survivors of premature birth develop chronic lung disease [4].

One of the major hurdles in the successful prediction of premature labor are the unpredictable uterine contractions. However, research has shown that the most accurate method yet established is the classification of EHG signals [5]. Electrohysterography (EHG) measures electrical activity in the uterus, and is a specific form of electromyography (EMG), the measurement of such activity in muscular tissue. EHG signal is recorded inexpensively and noninvasively using bio-potential electrodes from the abdominal wall of pregnant women [5]. Previous researches indicate that these signals can be helpful to separate uterine records of term and pre-term deliveries [6], [7], [8].

Some notable research has been done on the subject which includes the classification of Term and Preterm Deliveries using linear features, using non-linear features [6] and other features like wavelet transform [8].

It is known from similar work on the problem that non-linear features like Peak frequency and Sample Entropy yield much better results as compared to the linear ones [9].

Since the underlying physiological mechanisms of biological systems are non-linear processes [10] and the female uterine wall is composed of billions of interconnected cells whose responses are non-linear, it may be regarded as a complex, non-linear dynamic system. To analyze the outputs of such a system, non-linear signal processing techniques are applicable. Therefore, one can hypothesize that non-linear signal processing techniques may yield better results in analysis of the EHG than linear ones. Besides, previous studies on the use of some nonlinear signal processing techniques [11] have produced promising results.

Most of the recent research studies have utilized using EHG for detection of true labor, In contrast this paper centers around the classification of EHG using non-linear techniques and features like Fast Fourier Transform. A comparison of linear, non-linear and FFT is also discussed in this study. The ultimate classification is achieved by comparing various machine-learning classifiers against an open dataset, containing 300 records) [12].

The rest of the paper is classified as follows: Section II provides the information about the data acquisition, preprocessing, extracted features and SVM method. The results and discussion are presented in section III and a final conclusion is made in section IV.

## II. METHODS

### A. Records

The Term-Preterm EHG Database included in this research contains the EHG records performed from 1997 until 2006 at the Department of Obstetrics and Gynecology, Medical Centre Ljubljana, Ljubljana [6]. The records were obtained from the general population as well as the patients admitted to the hospital, including both term and pre-term cases. Each patient was recorded only once, with the record duration of 30 minutes and the sampling frequency ($f_s$) of 20 Hz. The scanning system had a resolution of 16 bits with the amplitude range of $\pm$ 2.5 mV. Each record was obtained using four $AgCl_2$ electrodes, which were placed on the abdominal surface, with their potential differences forming three channels [6]. The first electrode (E1) was placed 3.5 cm to the left and 3.5 cm above the navel; the second electrode (E2) was placed 3.5 cm to the right and 3.5 cm above the navel; the third electrode (E3) was placed 3.5 cm to the right and 3.5 cm below the navel; the fourth electrode (E4) was placed 3.5 cm to left and 3.5 cm below the navel. I.e. The electrodes were placed such that they are 7 cm apart, centered at the navel **(fig. 1)**. Three channels were produced by recording the differences in the potential electrodes: S1 = E2 – E1 (first channel); S2 = E2 – E3 (second channel); S3 = E4 – E3 (third channel) [5]. The signals were filtered before sampling using analog three-pole Butterworth filter having a bandwidth from 0 to 5 Hz.

A total of 300 EHG records were used in this research, which were divided in two groups:

1. Term Records: (pregnancy duration $\geq$ 37 weeks)
   A total of 262 records were term, of which 143 were *early* (obtained before the 26th week of pregnancy) and 119 were *later* (obtained after the 26th week of pregnancy)

2. Pre-Term Records: (pregnancy duration $\leq$ 37 weeks)
   A total of 38 records were term, of which 19 were *early* (obtained before the 26th week of pregnancy) and 19 were *later* (obtained after the 26th week of pregnancy)

   Among these, the *early* records showed a relatively low frequency of contraction [13]. So these records were given more importance, as they had lesser noise levels.

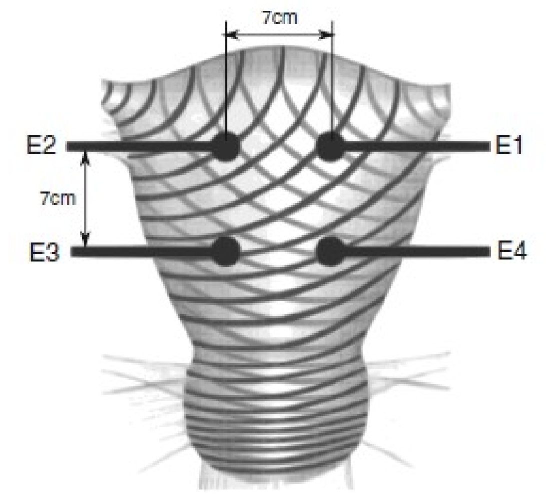


**Fig. 1** The placement of the electrodes on the abdomen, above the uterine surface. Signal 1: *E2–E1*, signal 2: *E2–E3*, signal 3: *E4–E3*

### B. Preprocessing

The lower frequencies of the EHG signals contain noise due to skin stretching and breathing. Thus, the recorded signals are first preprocessed using Butterworth digital filters of different frequency bands, i.e. 0.08 – 4 Hz [14], 0.05 – 4 Hz [16]. 0.2 – 4 Hz [15]. The beginning and ending part of 90 seconds was also removed from the records because of the presence of transient effects of the filters. The EHG signal records before preprocessing are shown in **fig. 2** and after preprocessing are shown in **fig. 3**. Afterwards, various linear, non-linear and statistical features were then used to classify to records.

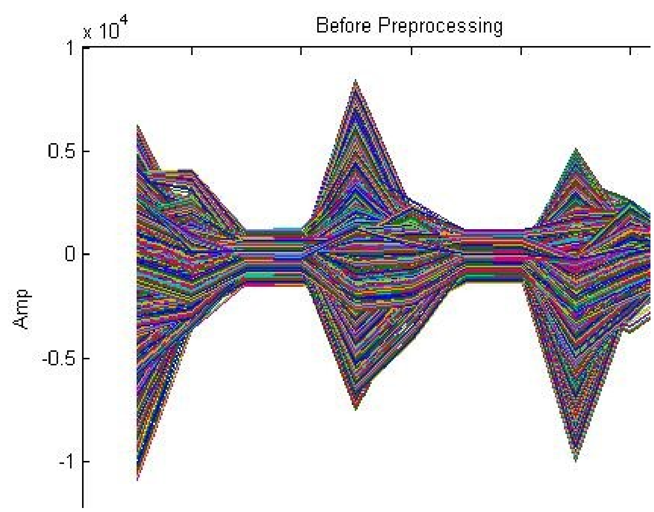


**Fig. 2** EHG Record (before preprocessing)

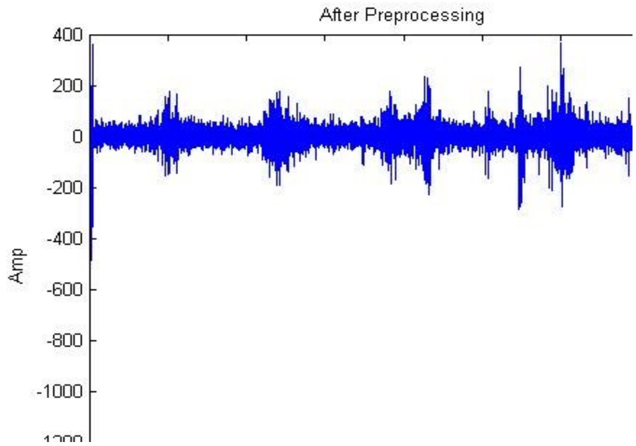

**Fig. 3** EHG Record (after preprocessing)

### *C. Feature Extraction*

This research includes 3 linear, 1 non-linear and 3 statistical features extracted from the EHG signals in order to differentiate between term and pre-term records.

1. Root Mean Square

**RMS** value of each signal was calculated as the root of the mean of the square if all samples in a signal, i.e.

$$RMS = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1} x(i)^2}$$

Where;
$x(t)$: A time-series Signal
t = 0, 1, 2 … N – 1.

2. Peak Frequency of the Power Spectrum

The fast discrete Fourier transform was used to calculate the power spectrum P for each signal $x(t)$. Then, the peak frequency $f_{max}$ was calculated as follows:

$$f_{max} = \arg(\frac{f_s}{N} max_{i=0}^{N-1} P(i))$$

3. Median Frequency

The median frequency $f_{med}$ was calculated using:

$$f_{max} = i_m \frac{f_s}{N}$$

$$\sum_{i=0}^{i=i_m} P(i) = \sum_{i=i_m}^{i-N-1} P(i)$$

4. Sample Entropy

The sample entropy **sampEn** is defined as:

$$sampEn_{m,r}(x) = \begin{cases} -\log\left(\frac{c_m}{c_{(m-1)}}\right) & ; \quad c_m \neq 0 \wedge c_{(m-1)} \neq 0 \\ -\log\left(\frac{N-m}{N-m-1}\right) & ; \quad c_m = 0 \vee c_{(m-1)} = 0 \end{cases}$$

5. Mean

The mean value $\bar{x}$ of a signal was calculated using:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x(i)$$

6. Variance

The variance $s^2$ of a signal was calculated as follows:

$$s^2 = \frac{\sum_{i=1}^{N}(x(i) - \bar{x})^2}{N-1}$$

7. Standard Deviation

The standard deviation $s$ of a signal was calculated as follows:

$$s = \sqrt{\frac{\sum_{i=1}^{N}(x(i) - \bar{x})^2}{N-1}}$$

### *D. Support Vector Machine*

The Machine Learning classifier used in this research was an SVM (Support Vector Machine) classifier. Given a set of training examples, each marked for belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a plausible classifier.

It is basically a two layer neural network, employing hidden layer of radial units and one output neuron. The procedure of creating this network and learning its parameters is organized in the way in which we deal only with kernel functions instead of direct processing of hidden unit signals. Basic SVM is linear but it can be used for nonlinear data by using kernel function to first indirectly map non-linear data into linear feature space. Basic SVM is also a two-class classifier however; with some modification, multiclass classifier can be obtained. The SVM works with a construction of optimal separating hyper-plane which can separate data from two classes which in our case is Term data and Preterm data. It seeks to find a hyper-plane by focusing on the training cases that are placed at the edge of the class descriptors [17].

The distance between the hyper-plane and the nearest data points is called the margin of the SVM classifier, which is mathematically stated as:

$$D(x) = (w.x) + b$$

Where x is a vector of the dataset, and w and b are parameters of the hyper-plane that the SVM has to estimate.

## III. Results and discussion

For the classification of Term and Preterm records, linear and non-linear features have been extracted and applied to binary classifier (SVM). To improve the accuracy of the results a combination of all the extracted features have been applied to the classifier. As already mentioned, 3 channels have been used to record EHG signals, thus features extraction and classification process have been performed based on all 3 channels. As evident from the graphs and tables, statistical features from channel 1 can be more effective for diagnosing term and pre-term labor. It is also clear that the prediction becomes relatively easily with the classification using non-linear features and FFT, however SVM being a linear classifier requires a kernel function for implementation and promising results.

These results indicate that with the correct selection of features, SVM can positively predict premature birth and with EHG measurement and classification being a non-invasive and inexpensive method, the study can practically be implemented in hospitals for successful prediction of premature birth as opposed to the conventional methods of detection.

## IV. conclusion

The main idea of this research was to classify pre-term and term birth using EHG signals. For the said purpose statistical and non-linear features had been extracted and had been applied for classification with SVM based algorithm. The obtained results show that features from channel 1 can be more effective for diagnosis of pre-term labor. Also FFT is an important feature which greatly improves the accuracy of the classifier when applied with a kernel function.


## Acknowledgments

1. Prediction of Preterm Labor from EHG signals using Statistical and Non-linear Features by Danial Taheri Farl, Matin Beiranvand and Mohammad Shahbakhti.
2. Physiobank ATM open source database (Term-Preterm EHG Data)
3. A comparison of various linear and non-linear signal processing techniques to separate uterine EMG records of term and pre-term delivery groups by G. Fele-Zˇ orzˇ Æ G. KavsˇekÆ Zˇ Novak-Antolic ˇ Æ F. Jager


## References


[1] B. Karlsson, J. Terrien, V. Gudmundsson, T. Steingrimsdottir, and C. Marque, "Abdominal ERG on a 4 by 4 grid: mapping and presenting the propagation of uterine contractions," in 11th Mediterranean Conference on Medical and Biomedical Engineering and Computing 2007, voL 16, 2007, pp. 139-143.

[2] Erdenik I, Pajntar M and Leskosek "Uterine electrical activity as predictor of preterm birth in women with preterm contractions" Eur. J. Obstet. Gyn. RB 95149_153.

[3] Baker PN, Kenny L (2011) Obstetrics by Ten Teachers. Hodder Arnold Press. 436 p.

[4] Greenough A (2012) Long Term Respiratory Outcomes of very Premature Birth (< 32 weeks). Semin Fetal Neonatal Med 17(2): 73–76.

[5] S.M.S.Baghamoradi, M.Naji and R.aryadoost, "Evaluation of Cepstral Analysis of ERG Signals to Prediction of Preterm Labor" Biomedical Engineering (ICBME), 2011 18th Iranian Conference, Dec. 2011, pp.81 – 83.

[6] G. Fele-ZDorzD, G. Kavsek, Z. Novak-Antolic and F. Jager, "A comparison of various linear and non-linear signal processing techniques to separate uterine EMG records of term and pre-term delivery groups," Med Bioi Eng Comput, Apr, 2008, pp. 911-922

[7] M. Hassan, J. Terrien, A. Alexandersson, C. Marque and B. Karlsson, "Nonlinearity of EHG signals used to distinguish active labor from normal pregnancy contractions," Conf Proc IEEE Eng Med Bioi Soc, 2010, pp.2387-90.

[8] M. Hassan, J. Terrien, B. Karlsson and C. Marque, "Interactions between Uterine EMG at Different Sites Investigated Using Wavelet Analysis: Comparison of Pregnancy and Labor Contractions," Eurasip Joumal on Advances in Signal Processing, vo1.201O, 201O,pp. 1-10.

[9] C. Marque and M. A. Khalil, "Modeling And Classification for Uterine EMG Signals Using Autoregressive Model," Intelligent Engineering systems through artificial neural networks, ASME press, Vo1.l6, 2006, pp.317-323.

[10] Akay M (2001) Nonlinear biomedical signal processing, vol II. Dynamic analysis and modeling. IEEE Inc., New York.

[11] Fele-Zˇ orzˇ G (2006) Characterization of EMG data. In: Rudel D, Oberzˇan D (Eds) First international workshop on pregnant uterine smooth muscle EMG activity. Institute of biomedical informatics, University of Ljubljana, pp 15–15.

[12] Fele-Zˇ orzˇ G, Kavsˇek G, A comparison of various linear and non-linear signal processing techniques Medical & biological engineering & computing 46(9): 911–22.

[13] Iams JD, Newman RB, Thom EA, Goldenberg RL, Mueller- Heubach E, Moawad A, Frequency of uterine contractions and the risk of spontaneous preterm delivery. N Engl J Med 346(4):250–255.

[14] Verdenik I, Pajntar M, Leskosek B (2001) uterine electrical activity as predictor of preterm birth in women with preterm contractions. Eur J Obstet Gynecol Reprod Biol 95(2):149–153.

[15] Maul H, Maner WL, Olson G, Saade GR, Garfield RE (2004) Non-invasive transabdominal uterine electromyography correlates with the strength of intrauterine pressure and is predictive of labor and delivery. J Matern Fetal Neonatal Med 15(5):297– 301.

[16] Maner WL, Garfield RE (2007) Identification of human term and preterm labor using artificial neural networks on uterine electromyography data. Ann Biomed Eng 35(3):465–473.

[17] A. Tzotsos, "A Support Vector Machine Approach for Object Based Image Analysis" In Proceedings of International Conference on Object-Based Image Analysis, 2006.